Some issues on the concept of Causality in Spatial Econometric models

Jesús Mur
University of Zaragoza
Department of Economic Analysis
and
Jean H.P. Paelinck
George Mason University
School of Public Policy

Abstract:

In general, cross-sectional econometric models are specified under the assumption of simultaneity, which means that the relations among the agents are solved at the same moment in time. This aspect does not facilitate the distinction between causes and effects. However, the notion of cause is of paramount importance in order to specify any model where it is assumed that a variable, called endogenous, is caused by the variables introduced in the right hand side of the equation, the regressors.

Our impression is that this problem has been treated very informally in the Spatial Econometrics literature, where the specification of the equation depends almost exclusively on theoretical considerations. In this sense, the content of the paper focuses on questions related to the specification process. We examine what may be called current traditional practice and discuss the position that the concepts of identification and causality should play in this context. Our objective is to produce useful econometric guidelines in order to help the user to improve the theoretical foundations of the equation.

Keywords: Causality; Identification; Spatial Econometric Models.

JEL Classification: C21; C50; R15

1.- Introduction

The concepts of causality and endogeneity are basic elements on the specification of an econometric model. In general, it is supposed that the variables that appear on the right-hand side of the equation cause the variable that appears on the left-hand side. The treatment of the relation is easier if, furthermore, the variables that appear in the right-hand part are either exogenous or predetermined. This discussion forms part of the habitual practice, although it has not been elaborated so much in a spatial context (Anselin, 1988).

One of the main difficulties is related with the nature of the data because, on many occasions, we only have a cross-section without a time perspective. Nevertheless, the literature on causality has, from its origins, insisted on the principle of *temporal succession*, under which the cause must precede the effect. If we loose the temporal perspective, the discussion about causality is more complicated, but not irrelevant. On the contrary, we think that is very important to be aware of the possible existence of relations of causality between the variables of the model. With this objective, we begin the discussion by revising a series of concepts such as identification, predeterminedness or exogeneity. These concepts are related and they have clear connections with the question of causality.

In the second section we present the notation that we are going to use, together with some fundamental concepts and definitions. The context to which we refer in this case is the habitual one of time series. In the third section we introduce the spatial dimension which results in different problems. In the fourth section we examine some proposals in relation to the question of causality whose behavior is check in the fifth section by means of a Monte Carlo experiment. In the sixth section we apply the proposed procedure to the case of the relation between income and activity of the agricultural sector in the Spanish provinces. We finish the paper with a section of conclusions.

2- Definitions and essential concepts.

The econometric literature has discussed the concepts of causality, exogeneity and identification in great depth, so there is a certain consensus about the interpretation and use of these terms. This section brings together a sample of these definitions (see Bresson and Pirotte, 1995, Davidson, 2000, or Greene, 2007, for a more general discussion).

An econometric model relates a set of variables with different objectives that can range between prediction and simulation. Generally, attention is focused on the endogenous variables, y_t, whose behavior is explained by the model. The other variables, z_t, may be also endogenous, predetermined or exogenous and are of interest because they explain us how the endogenous variable is formed. There are other variables, supposed not to be relevant and concentrated in vector w_t. Other elements needed to complete the structure of the model are the parameters and different deterministic terms (Charemza and Deadman, 1997).

The joint density function, $D\left(w_t; y_t; z_t; \Psi\right)$, where Ψ is a vector of parameters, encapsulates the idea of a data generating process (DGP from now on) and a model. As Davidson (2000, p. 74) indicates, '*The analysis is said to be conditional on z_t, whereas the model is marginalized with respect to w_t. The marginalization process is often taken for granted in empirical work (....). However, if a variable that should be in z_t is incorrectly assigned to w_t, its omission constitutes a misspecification*'. In other words:

$$\begin{aligned} D\left(w_t; y_t; z_t; \Psi\right) = {} & D_{w|y,z}\left(w_t \mid y_t; z_t; W_{t-1}; Y_{t-1}; Z_{t-1}; d_t; \Psi_2\right) \\ & D_{y|z}\left(y_t \mid z_t; Y_{t-1}; Z_{t-1}; d_t; \Psi_1\right) D_z\left(z_t \mid W_{t-1}; Y_{t-1}; Z_{t-1}; d_t; \Psi_2\right) \end{aligned} \quad (1)$$

where W_{t-1}, Y_{t-1} y Z_{t-1} refers to the history of the variables w, y and z until (t-1) and d_t is the set of deterministic elements that intervene in the DGP of the variables. Obviously, the factorization of (1) is arbitrary and depends on the objective of the analysis.

Ψ_1 are the *parameters of interest* while Ψ_2 are *nuisance parameters*. It is important that the density functions $D_{w|y,z}$ and D_z do not depend on vector Ψ_1 and that there are no crossed restrictions between vectors Ψ_1 and Ψ_2 (Hendry et al, 1983, use the term *sequential cut*). Similarly, it is important that the conditioned density function of y, $D_{y|z}$, does not depend on $w_{t-j}\ (\forall j > 0)$ which assures that the latter variables do not affect y. In fact, if both conditions are verified (there exists a sequential cut in Ψ and, also, W_{t-j} is not relevant in $D_{y|z}$), all the information we need to know about the behavior of y can be found in $D_{y|z}$. The last density function completely represents the stochastic mechanism generating the variable y.

A series of interesting properties are fulfilled in (1). For example, vector z is *weakly exogenous* for Ψ_1, which means that this vector intervenes in the conditional modeling of the variable y but not in the generation process of z. This is a relation between variables and parameters whose meaning and implications are always relative to the problem under investigation.

The concept of causality relates variables and, in line with Granger (1969), is constructed upon the idea of *predictability*: if the group of variables z_t cause y_t, the information about the former variables must improve our knowledge about the latter, which are our variables of interest.

This discussion revolves around the structure of $D_{y|z}$ in (1). For example, if $D_{y|z}\left(y_t \mid z_t; Y_{t-1}; Z_{t-1}; d_t; \Psi_1\right) = D_{y|z}\left(y_t \mid Y_{t-1}; Z_{t-1}; d_t; \Psi_1\right)$, the conclusion will be that *z* does not contemporaneously cause y; if the absence of relation also extends to the past: $D_{y|z}\left(y_t \mid z_t; Y_{t-1}; Z_{t-1}; d_t; \Psi_1\right) = D_{y|z}\left(y_t \mid Y_{t-1}; d_t; \Psi_1\right)$, the conclusion is that z does not cause y relative to y_{t-1}. The definition of Granger noncausality is purely operative, in the sense that the causal variables should help predict the caused variable better, whereas the definition of weak exogeneity is formal. However, and in spite of their apparent similarity, the two concepts (Granger noncausality and weak exogeneity) are not necessarily related. As is well known, weak exogeneity plus non-causality results in strong exogeneity for Ψ_1. In fact, if z is weakly exogenous with respect to Ψ_1 and, at the same time, y does not cause z, we can handle the conditional $D_{y|z}$ and the marginal D_z separately. For practical purposes, this means that the variables z act as if they were fixed in the conditioning model.

Another important concept is that of *invariance,* in reference to a parameter that remains constant under a certain type of interventions. In particular, if all the parameters of a conditional model are invariant for any change in the distribution function of the conditioning variables, we can speak of structural invariance. If, furthermore, we add the weak exogeneity property of the conditioning variables (the z's) in relation to the parameters of interest of the conditional model (vector Ψ_1), the result is super exogeneity with respect to the parameters of interest.

Engle et al. (1983, p.286) describe the role of these concepts: '*weak exogeneity validates conducting inference conditional on z_t while Granger noncausality validates forecasting z and then forecasting y conditional on the future z's (...) Obviously, if estimation is required before conditional predictions are made, then strong exogeneity which covers both Granger noncausality and weak exogeneity becomes the relevant concept*'.

Another part of the story deals with the concepts of predeterminedness and strict exogeneity, which refer to the relation between a variable on the right-hand side of the equation and the disturbance term of the model. If the variable in question is independent of future disturbances, we can speak of predeterminedness and, if this relation is maintained whatever the temporal direction considered, we obtain strict exogeneity. As Davidson, (2000, p. 79) indicates, the disturbances '*are fictional constructs that have no reality outside of the particular model and parametrization we have chosen*' which brings the discussion back to the terrain of the relationship between variables and parameters.

Weak exogeneity tends to be associated with predeterminedness and strict exogeneity with strong exogeneity. However, Engle et al. (1983, Theorem 4.3) demonstrate that there remain many situations where they differ given that both characterizations (weak exogeneity plus strict exogeneity vs predeterminednes plus strong exogeneity) are defined in different contexts. The first, weak exogeneity plus strict exogeneity, refers to the parameters of interest in the structural form whereas the second, predeterminednes plus strong exogeneity, operates on the reduced form of the model. It is important to remember that the relation between the two forms may be ambiguous, unless the supposition of *identification* is verified. Knowledge of the sample data moments between the endogenous and the regressors will suffice to determine the parameters of the reduced form, but there may exist different structural models compatible with the same reduced form equations, so '*the economic theory must fix some elements of these structural matrices in advance. When there are not sufficient prior restrictions to rule out observationally equivalent structures, the model is said to be underidentified, in whole or in part*' (Davidson, 2000, p.185). For a system of homogeneous linear equations, the structural and reduced forms are, respectively:

$$\begin{array}{ll} \text{Structural Form:} & \text{Reduced Form:} \\ \mathbf{B}y_t + \boldsymbol{\Gamma} z_t = u_t & y_t = \boldsymbol{\Pi} z_t + v_t \end{array} \tag{2}$$

y_t is a (Gx1) vector of endogenous variables, z_t an (Mx1) vector of predetermined variables, $u_t = y_t - E\left[y_t \middle| I_t\right]$, and I_t is the informative base for the period t which comprises the previous history of y_t and z_t as well as contemporaneous data for the exogenous variables. $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are structural matrices of order (GxG) and (GxM), respectively, and $\boldsymbol{\Pi}$ is a (GxM) matrix of reduced form parameters which are always identified by the sampling information. Finally, v_t is the (Gx1) vector of reduced form error terms ($v_t = \mathbf{B}^{-1} u_t$). As is well-known, the identification condition for this case is:

$$\operatorname{rank}\left(\mathbf{A\Phi}\right) = G(G-1) \rightarrow \begin{cases} \mathbf{A} = \begin{bmatrix} \mathbf{B} & -\mathbf{\Gamma} \end{bmatrix}_{(Gx(G+M))} \\ \mathbf{\Phi}_{((G+M)xR)} \end{cases} \qquad (3)$$

where $\mathbf{\Phi}$ is the matrix of restrictions on the structural parameters. That is, we need to complete the structural form of (2) with G(G-1) restrictions on the parameters. The clause of (3) defines necessary and sufficient conditions to assure identification, whereas the *order condition,* $R \geq G-1$, is necessary but not sufficient. Extending the discussion for each of the equations of the system, the identification condition of the i-th equation, the *rank condition,* after normalizing, appears to be:

$$\operatorname{rank}(\mathbf{A\Phi}_i) = G - 1 \qquad (4)$$

where $\mathbf{\Phi}_i$ is the matrix of restrictions of the equation. A necessary condition for the identification of the equation is $M - m_i \geq g_i - 1$; that is, the number of excluded predetermined variables, M-m_i, should be greater than or equal to the number of included endogenous minus one, G-1. It is not necessary to remember the position of Sims (1980, p.14) with respect to the indiscriminate use of restrictions on the structural form: '*I have argued earlier that most of the restrictions on existing models are false, and the models are nominally overidentified'*.

3- A general spatial model.

In this section we are going to specify a general spatial model in order to discuss the different concepts introduced in the previous section. We assume that we have sample information for a set of R individuals taken over T periods; that is, initially we have a panel with which we specify a model of simultaneous equations like the following:

$$\begin{array}{l} y_{rt} = \alpha_{r1} y_{1t} + \cdots \alpha_{rR} y_{Rt} + x'_{1t}\beta_{r1} + \cdots x'_{Rt}\beta_{rR} + u_{rt} \\ r = 1,2,\ldots,R \quad t = 1,2,\ldots,T \quad \alpha_{rr} = 1 \end{array} \qquad (5)$$

In the r-th equation we explain the endogenous variable, y_{rt}, based on what happens in its neighborhood, in which we include the variable itself observed in points of space other than r, and a vector of predetermined and exogenous variables located both at point r and on other different places; x_{mt} is a vector of order (kx1) of observations taken at point m and β_{rm} the corresponding vector of parameters that intervene in equation r, also of order (kx1); u_{rt} is an error term. Equation (5) has been normalized so that α_{rr}=1.

Using a more compact notation:

$$Y=\begin{bmatrix} y_{11} & y_{12} & y_{13} & \cdots & y_{1T} \\ y_{21} & y_{22} & y_{23} & \cdots & y_{2T} \\ y_{31} & y_{32} & y_{33} & \cdots & y_{3T} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ y_{R1} & y_{R2} & y_{R3} & \cdots & y_{RT} \end{bmatrix}_{(RxT)} \qquad X=\begin{bmatrix} x_{11} & x_{21} & x_{31} & \cdots & x_{R1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{R2} \\ x_{13} & x_{23} & x_{33} & \cdots & x_{R3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{1T} & x_{2T} & x_{3T} & \cdots & x_{RT} \end{bmatrix}_{(TxRk)}$$

$$x_{mt}=\begin{bmatrix} x_{m1t} & x_{m2t} & x_{m3t} & \cdots & x_{mkt} \end{bmatrix}$$

$$B=\begin{bmatrix} 1 & -\alpha_{12} & -\alpha_{13} & \cdots & -\alpha_{1R} \\ -\alpha_{22} & 1 & -\alpha_{23} & \cdots & -\alpha_{2R} \\ -\alpha_{31} & -\alpha_{32} & 1 & \cdots & -\alpha_{3R} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\alpha_{R1} & -\alpha_{R2} & -\alpha_{R3} & \cdots & 1 \end{bmatrix}_{(RxR)} \qquad \Gamma=\begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} & \cdots & \beta_{R1} \\ \beta_{12} & \beta_{22} & \beta_{32} & \cdots & \beta_{R2} \\ \beta_{13} & \beta_{23} & \beta_{33} & \cdots & \beta_{R3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \beta_{1R} & \beta_{2R} & \beta_{3R} & \cdots & \beta_{RR} \end{bmatrix}_{(RxRk)} \tag{6}$$

$$\beta_{mr}=\begin{bmatrix} \beta_{m1r} & \beta_{m2r} & \beta_{m3r} & \cdots & \beta_{mkr} \end{bmatrix}_{(kx1)}$$

$$U=\begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1T} \\ u_{21} & u_{22} & u_{23} & \cdots & u_{2T} \\ u_{31} & u_{32} & u_{33} & \cdots & u_{3T} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ u_{R1} & u_{R2} & u_{R3} & \cdots & u_{RT} \end{bmatrix}_{(RxT)} \qquad \Rightarrow BY=\Gamma X'+U$$

From the structural form of (6), the reduced form can be easily obtained:

$$\begin{aligned} BY=\Gamma X'+U \Rightarrow Y=\Pi X'+V \\ \Pi_{(RxRk)}=B^{-1}\Gamma \\ V=B^{-1}U \end{aligned} \tag{7}$$

Without loss of generality, we complete the specification with the following propositions:

$$\begin{aligned} & z_t=\begin{bmatrix} Y_t \\ X_t \end{bmatrix}_{(R(k+1)x1)} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix};\begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}_{R(k+1)xR(k+1)}\right] \\ & \to \quad \mu_{t(Rx1)}=E\left[Y_t \mid X_t=x_t\right]=D'x_t \\ & \qquad D=\Sigma_{XX}^{-1}\Sigma_{XY} \\ & \to \quad v_{t(Rx1)}=Y_t-E\left[Y_t \mid X_t=x_t\right]=Y_t-\mu_t \\ & \qquad E\left[v_t\right]=E\left[E\left[v_t \mid X_t=x_t\right]\right]=0 \\ & \qquad E\left[v_t v'_s\right]=E\left[E\left[v_t v'_s \mid X_t=x_t\right]\right]=\begin{cases} \Omega; & t=s \\ 0; & t\neq s \end{cases} \\ & \qquad \Omega_{(RxR)}=\Sigma_{YY}-\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \\ & \qquad E\left[\mu_t v'_t\right]=E\left[E\left[\mu_t v'_t \mid X_t=x_t\right]\right]E\left[\mu_t E\left[v'_t \mid X_t=x_t\right]\right]=0;\forall t \end{aligned} \tag{8}$$

There are R(R-1) parameters in matrix B and R^2k in matrix Γ, making a total of R[R(k+1)-1] structural parameters of interest. Furthermore, in the reduced form of (7), there are R^2k statistical parameters in matrix Π. In each of the matrices of covariances, Δ and Ω, we find another R^2 parameters. It is obvious that the model is *underidentified*. To achieve identification, it will be necessary to introduce, at least, R(R-1) restrictions on the structural parameters.

In a spatial context, these restrictions can be obtained in various ways, for example, through the matrix B of (5). If we assume, as usual, that there exists a weighting matrix, of binary type for simplicity's sake, which correctly reflects the structure of spatial dependencies, we can write:

$$\begin{aligned} &B = I - \rho W \\ &w_{rs} = \begin{cases} 1 & \text{if } s \in N_r \\ 0 & \text{if } s \in N_r \end{cases} \end{aligned} \tag{9}$$

where N_r is the set of neighbors which are related to point r and ρ a parameter of spatial autocorrelation. In (9) we obtain [R(R-1)-1] restrictions. To achieve identification we need, at least, one more restriction that can now be obtained from matrix Γ. It appears reasonable to restrict the capacity of interaction between the endogenous, located at a given point, and the predetermined variables located elsewhere in space. This means that some of the β's located outside the main diagonal of matrix Γ will be zero. If, furthermore, we introduce the restriction of homogeneity between these vectors of parameters, we can write:

$$\begin{cases} \beta_{sr} = 0 & \text{if } s \neq r \\ \beta_{ss} = \beta_{rr} = \beta & \text{if } s{=}r \end{cases} \tag{10}$$

We obtain $(R^2$-1)k restrictions on the parameters of matrix Γ, which, added to those already introduced in the composition of matrix B, give us a total of $[R^2(2k+1)-(R+k)]$ restrictions. The number of parameters of the reduced form is still R^2k while in the structural form only intervene (k+1) parameters:

$$\underbrace{B^{-1} = [I_R - \rho W]^{-1} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1R} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2R} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3R} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{R1} & b_{R2} & b_{R3} & \cdots & b_{RR} \end{bmatrix} \quad \Gamma = \begin{bmatrix} \beta & 0 & 0 & \cdots & 0 \\ 0 & \beta & 0 & \cdots & 0 \\ 0 & 0 & \beta & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \beta \end{bmatrix}}$$

$$\Pi = B^{-1}\Gamma \Rightarrow \begin{bmatrix} \pi_{11} & \pi_{12} & \pi_{13} & \cdots & \pi_{1R} \\ \pi_{21} & \pi_{22} & \pi_{23} & \cdots & \pi_{2R} \\ \pi_{31} & \pi_{32} & \pi_{33} & \cdots & \pi_{3R} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \pi_{Rk1} & \pi_{Rk2} & \pi_{Rk3} & \cdots & \pi_{RkR} \end{bmatrix} = \begin{bmatrix} b_{11}\beta & b_{12}\beta & b_{13}\beta & \cdots & b_{1R}\beta \\ b_{21}\beta & b_{22}\beta & b_{23}\beta & \cdots & b_{2R}\beta \\ b_{31}\beta & b_{32}\beta & b_{33}\beta & \cdots & b_{3R}\beta \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{R1}\beta & b_{R2}\beta & b_{R3}\beta & \cdots & b_{RR}\beta \end{bmatrix} \tag{11}$$

$$\Pi{=}B^{-1} \otimes \beta \Leftrightarrow B\Pi = I_R \otimes \beta = \Gamma$$

The model, now, is *overidentified*; we have R(R-1) restrictions of overidentification. The matrices of variance and covariances would allow us to derive new restrictions. For example, if we assume incorrelation between the error terms of the cross-sections of the structural form of (6):

$$\begin{aligned} u_{t(Rx1)} &= B\left[Y_t - E\left(Y_t \middle| X_t = x_t\right)\right] \\ E\left[u_t\right] &= BE\left[E\left[v_t \middle| X_t = x_t\right]\right] = 0 \\ E\left[u_t u'_s\right] &= BE\left[E\left[v_t v'_s \middle| X_t = x_t\right]\right]B' = \begin{cases} \Lambda \rightarrow \Omega = B^{-1}\Lambda B'^{-1}; & t = s \\ 0; \quad t \neq s \end{cases} \end{aligned} \tag{12}$$

where Λ is a diagonal matrix. In this way we have R(R-1) new restrictions or (R^2-1) if we also assume homoskedasticity. In the first case, we have a total of 2R[R(k+1)-1]-k restrictions of overidentification and, in the second case, $2R^2$(k-1)-(R+k+1). That is to say, as long as we have a sufficient number of cross-sections (it must hold that T>Rk), the simultaneous equations model of (5) can be estimated in the usual way.

Problems arise when the temporal size is reduced to only one cross-section (T=1). In order to assure the identification of the model we must introduce, at least, all the restrictions mentioned previously:

$$Y = \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ \cdots \\ y_{Rt} \end{bmatrix}_{(Rx1)} \qquad X = \begin{bmatrix} x_{1t} & x_{2t} & x_{3t} & \cdots & x_{Rt} \end{bmatrix}_{(1xRk)} \qquad U = \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ \cdots \\ u_{Rt} \end{bmatrix}_{(RxT)}$$

$$x_{rt} = \begin{bmatrix} x_{r1t} & x_{r2t} & x_{r3t} & \cdots & x_{rkt} \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & -\rho w_{21} & -\rho w_{31} & \cdots & -\rho w_{R1} \\ -\rho w_{12} & 1 & -\rho w_{32} & \cdots & -\rho w_{R2} \\ -\rho w_{13} & -\rho w_{23} & 1 & \cdots & -\alpha_{R3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\rho w_{1R} & -\rho w_{2R} & -\rho w_{3R} & \cdots & 1 \end{bmatrix}_{(RxR)} = I - \rho W \tag{13}$$

$$\Gamma = \begin{bmatrix} \beta & 0 & 0 & \cdots & 0 \\ 0 & \beta & 0 & \cdots & 0 \\ 0 & 0 & \beta & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \beta_{RR} \end{bmatrix}_{(RxRk)} = I \otimes \beta$$

$$\beta = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_k \end{bmatrix}_{(kx1)}$$

However, under this setting it is easier to use the structural form of the model:

$$y = \rho W y + x\beta + u \Leftrightarrow$$

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ \cdots \\ y_{Rt} \end{bmatrix} = \rho \begin{bmatrix} 0 & w_{21} & w_{31} & \cdots & w_{R1} \\ w_{12} & 0 & w_{32} & \cdots & w_{R2} \\ w_{13} & w_{23} & 0 & \cdots & w_{R3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ w_{1R} & w_{2R} & w_{3R} & \cdots & 0 \end{bmatrix} \begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ \cdots \\ y_{Rt} \end{bmatrix} + \begin{bmatrix} x_{11t} & x_{21t} & x_{31t} & \cdots & x_{k1t} \\ x_{12t} & x_{22t} & x_{32t} & \cdots & x_{k2t} \\ x_{13t} & x_{23t} & x_{33t} & \cdots & x_{k3t} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{1Rt} & x_{2Rt} & x_{3Rt} & \cdots & x_{kRt} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \cdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ \cdots \\ u_{Rt} \end{bmatrix} \tag{14}$$

The equation of (14) contains (k+1) parameters, the same as the reduced form:

$$y = \left(I - \rho W\right)^{-1} x\beta + \left(I - \rho W\right)^{-1} u \tag{15}$$

The model is overidentified, giving that the number of cross-sectional observations, R, is greater than the number of parameters, k+1, as it is usual. The reduced form of (15) is nonlinear in parameters although it can be '*linearized'* by using the expansion of the inverse matrix in term of powers of the weighting matrix:

$$
\begin{aligned}
&(I-\rho W)^{-1} \approx I+\rho_1 W+\rho_2 W^2+\rho_3 W^3+\cdots \\
&\Rightarrow y = x\beta + x^{(1)}\beta_1 + x^{(2)}\beta_2 + x^{(3)}\beta_3 + \cdots + v \\
&\qquad x^{(j)} = W^j x \\
&\qquad \beta_j = \beta\rho_j \\
&\qquad v = (I-\rho W)^{-1} u
\end{aligned}
\tag{16}
$$

being v the error term of the reduced form, spatially autocorrelated and heteroskedastic. As said, it is simpler to work with the structural form of (14) than with the reduced form of (15); see Paelinck and Nijkamp, 1978, and Paelinck and Klaassen, 1979, for more details. In any case, the point that we would like to stress in this moment is that there exists a long list of restrictions underlying the specifications of (14) and (15). They must be assumed in order to assure the identification of the model and we must be aware of them.

Another point to note in relation to the problem of identification is that the topology of the space does not specially matter; that is, the composition of the weighting matrix (which it is supposed to reflect the structure of the space) does not have any incidence on that question. The minimum requirement is that, at least, two regions should be connected.. The shape of the spatial system (in terms of number and distribution of connections) will affect the robustness and the variance of the corresponding estimations.

4-. A first look into the topic of spatial Granger causality.

The concept of Granger causality relies, among other things, on the principle of '*temporal succession*' which implies that the cause must precede, in a temporal sense, the effect. This principle implies, for example, that the past of the variables should be used in order to check for the existence of causality relationships between two variables. However, a cross-section usually contains observations of the variables that are coincident in time or, at least, dated at the same moment of time. This seems to preclude the use of time dynamic specifications. Another well-established principle of this humenian strand of literature (Pearl, 2000) refers to the '*contiguity relation*' between the cause and the effect: both elements must coincide in a specific time and location. However, the '*allotopy'* is one of the main features of spatial econometrics models which, as explained, by Ancot et al (1990, p.141) implies that '*very often, the factors* *that explain a given economic fact in a region of space are located in distinct places*'. The principle of allotropy relaxes the restriction of physical contiguity between the agents that intervene in a given.

In short, the simultaneity of the data, which is a characteristic of cross sections, affects the applicability of the 'temporal succession' principle whereas the '*contiguity relation*' should be relaxed because of the type of models in which we are interested. The question then is if the concept of spatial dynamics may replace, at least partially, the role played bye the concept of temporal dynamics in the analysis of causality.

Briefly stated, Granger causality test (Granger, 1969) develops around the idea of predictability in the sense that, if variable x causes variable y, the past of the first variable must help to improve the forecasting performance of the second: $\sigma^2\left(y_{t+1} \mid I_t\right) < \sigma^2\left(y_{t+1} \mid I_t - x_t^*\right)$, where $\sigma^2(-)$ denotes uncertainty (the variance, in general), I_t the informative base up to period t and x_t^* the information of variable x up to period t. The test is very simple (see also Heyde, 1957, or Holly, 1984). The null hypothesis implies noncausality and results in a set of zero restrictions:

$$y_t = \alpha_0 + \sum_{k=1}^{k_1} \alpha_{1k} y_{t-k} + \sum_{k=1}^{k_2} \alpha_{2k} x_{t-k} + u_{1t}$$

$$\left.\begin{array}{l} H_0 : \alpha_{21} = \ldots = \alpha_{2k_2} = 0 \\ H_A : \text{No } H_0 \end{array}\right\} \rightarrow F = \frac{SR_0 - SR_A}{SR_A}\frac{T-(k_1+k_2)}{k_2} \underset{as}{\sim} F(T-(k_1+k_2);k_2) \tag{17}$$

being SR_0 and SR_A the sum of squared residuals of the model of the null and alternative hypothesis, respectively. If we change the terms past/future, which form the basis of the equation of (17), by proximity/remoteness, typical of spatial econometric models, we obtain:

$$y = \alpha_0 + \sum_{k=1}^{k_1} \alpha_{1k} W_k y + \sum_{k=0}^{k_2} \alpha_{2k} W_k x + u$$

$$\left.\begin{array}{l} H_0 : \alpha_{20} = \alpha_{21} = \ldots = \alpha_{2k_2} = 0 \\ H_A : \text{No } H_0 \end{array}\right\} \rightarrow F = 2\left(l_{H_A} - l_{H_0}\right) \underset{as}{\sim} \chi^2(k_2+1) \tag{18}$$

where $\{W_1, W_2, \ldots. W_{kj}\}$ is a succession of weighting matrices of order 1, 2, etc., with W_0 =I; y and x are (Rx1) vectors of observations of the two variables, in period t, and u a vector of error terms assumed, for simplicity, white noise. The F statistic is a likelihood ratio, where l_{H_A} and l_{H_0} refer to the estimated log-likelihood of the model of the alternative and null hypothesis respectively. The null hypothesis implies that the information about the spatial distribution of the x variable does not help to improve our understanding of the spatial distribution of the y variable at time t. Obviously, in continuation we have to invert the order of the variables to test the hypothesis that y does not cause x.

The two-step procedure described can be combined in a direct formulation specifying a spatial VAR (Di Giacinto, 2003, 2006, Beenstock and Felsenstein, 2007, 2008):

$$\left.\begin{array}{l} y = \alpha_0 + \sum_{k=1}^{k_{11}} \alpha_{1k} W_k y + \sum_{k=0}^{k_{12}} \alpha_{2k} W_k x + u_y \\ x = \beta_0 + \sum_{k=1}^{k_{21}} \beta_{1k} W_k x + \sum_{k=0}^{k_{22}} \beta_{2k} W_k y + u_x \end{array}\right\}$$

$$\begin{array}{lcl} \text{x does not cause y} & \rightarrow & H_0 : \alpha_{20} = \alpha_{21} = \ldots = \alpha_{2k_{12}} = 0 \\ \text{y does not cause x} & \rightarrow & H_0 : \beta_{20} = \beta_{21} = \ldots = \beta_{2k_{22}} = 0 \\ \begin{array}{c} \text{x does not cause y} \\ \text{AND} \\ \text{y does not cause x} \end{array} & \rightarrow & \begin{array}{l} H_0 : \alpha_{20} = \alpha_{21} = \ldots = \alpha_{2k_{12}} = \\ \qquad \beta_{20} = \beta_{21} = \ldots = \beta_{2k_{22}} = 0 \end{array} \end{array} \tag{19}$$

$$\rightarrow F = 2\left(l_{H_A} - l_{H_0}\right) \underset{as}{\sim} \chi^2(k_J)$$

being k_J the number of restrictions corresponding to each case. The bivariate system of (19) can be estimated by maximum-likelihood methods.

As a kind of exploratory, preliminary analysis of the possible causality relationships present in given set of variables, it can be interesting to combine the Simultaneous Dynamic Least Squares (SDLS from now on) estimators of Paelinck (1990) with a very simple and popular statistic in applied econometrics as the partial correlation coefficient. Let us introduce the case assuming a spatial model, like the following:

$$y = \rho Wy + x\beta_0 + Wx\beta_1 + u \tag{20}$$

where x is a matrix of explanatory (possibly exogenous) variables, if necessary including (partial) unit column vectors to take bare of region-specific constants; the x matrix can contain on-localized explanatory variables.

If the estimators are obtained by SDLS in the case of (20), they are equivalent to the estimators of the reduced form (Paelinck, 2007). This implies that, if x is a matrix of strongly exogenous variables, the endogenously generated $\hat{y}$ vector has the same property, and they can both be combined in a matrix z=[x, Wx, $\hat{y}$]. Then the incremental contributions of all variables can be computed according to Theil (1971, pp. 168-169). Recall that R_i^2 is defined as the multiple correlation coefficient from the multiple regression of y (observed) on z minus its i[th] column; it is further known that:

$$R^2 \approx \sum_{i=1}^{k} R_i^2 \tag{21}$$

being R^2 the global multiple correlation coefficient of the equation. Obviously, if the x variables are not important in explaining the spatial distribution of y, the marginal determination coefficient associated to these variables should be very small in comparison with that assured from the lag structure of the y variable.

5- Some Monte Carlo evidence.

In this section we are going to present the main results obtained from a small Monte Carlo experiment in which we have simulated the likelihood ratios of (18). The data have been obtained dynamically, using a dynamic spatial panel data model which allows us to respect the principle of '*temporal succession*'. However, we have used only the last cross-section to discuss the existence of causality relations between the variables. Specifically, the data generating process is the following:

$$\left.\begin{array}{c} y_t = \alpha_{10} y_{t-1} + \alpha_{01} \mathbf{W} y_t + \alpha_{11} \mathbf{W} y_{t-1} + \beta_{00} x_t + \beta_{10} x_{t-1} + \beta_{01} \mathbf{W} x_t + \beta_{11} \mathbf{W} x_{t-1} + \Gamma_t + \varepsilon_t \\ y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Rt} \end{bmatrix}; x_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{Rt} \end{bmatrix}; \Gamma_t = \begin{bmatrix} \lambda_{1t} \\ \lambda_{2t} \\ \vdots \\ \lambda_{Rt} \end{bmatrix}; \varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{Rt} \end{bmatrix} \end{array}\right\} \tag{22}$$

$$\varepsilon_t \sim N(0; \sigma_\varepsilon^2 I); Cov(\varepsilon_t; \varepsilon_{t \pm m}) = 0 \quad \forall m \neq 0$$

where W is the usual weighting matrix, α_{10}, α_{01}, α_{11}, β_{00}, β_{10}, β_{01} and β_{11} are parameters that measure the dependence (temporal, spatial or mixed) that is present in the panel; Γ_t is a vector of unobservable (fixed or random) individual effects; finally, ε_t is a white noise random vector (assumed to be normal and homoskedastic). The DGP of (22) is a simplified version of the general first-order serial and spatial autoregressive distributed lag model of Elhorst (2001), which includes a set of exogenous variables as well as their time and spatial lags.

In each of the T cross-section the data has been simulated on a ($\sqrt{R}x\sqrt{R}$) regular lattice. We have used four sample sizes as (T;R) = (5; 49), (5; 100); (10;49) and (10; 100). The random terms have been obtained from a gaussian distribution with zero mean and unit variance, $\varepsilon_t \sim N(0;I)$, whereas the vector x comes from a uniform distribution on the (0,1) interval, $x_t \sim U(0;1)$. W matrix has been specified always as a first order contiguity matrix, row-standardized. Finally, in these experiments and for the sake of simplicity, we have assumed that the Γ_t vector is composed by only a constant term, common to all individuals and cross-section, $\Gamma_t=\Gamma=\tau 1$, being l a (Rx1) vector of ones and τ a parameter. In the future we will relax this restriction introducing (time, spatial) fixed or random effects following Elhorst (2003).

The estimated size of the test, under different configurations, appears in Table 1. In case A we have simulated, strictly, the model of the null hypothesis; that is, in the DGP of the variable y only intervene its past and/or contemporaneous values plus the error term and the intercept: it is a pure spatiotemporal autoregressive model. The case B also belongs to the null hypothesis, in the sense that variable x does not appear in the DGP of variable y. Indeed, we have simulated the equation of (22) but using a variable z, in place of x. However, the testing equation has been specified using the variable x. In case B.1 both variables, x and z, are uncorrelated whereas in Case B.2 they have a correlation of about 0.5.

Table 1: Causality likelihood ratio. Estimated size.

Case A	$\beta_{00}=0; \beta_{01}=0; \beta_{10}=0; \beta_{11}=0$					
	α_{01}		α_{10}		α_{11}	
	0.1	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**
(5; 49)	0.036	0.043	0.060	0.066	0.009	0.072
(5; 100)	0.041	0.050	0.062	0.056	0.018	0.088
(10;49)	0.042	0.058	0.055	0.078	0.069	0.054
(10; 100)	0.061	0.059	0.072	0.045	0.059	0.066
Case B.1	$\beta_{00}=0.4; \beta_{01}=0.4; \beta_{10}=0; \beta_{11}=0$					
	α_{01}		α_{10}		α_{11}	
	0.1	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**
(5; 49)	0.075	0.047	0.064	0.035	0.001	0.051
(5; 100)	0.081	0.033	0.055	0.072	0.002	0.039
(10;49)	0.049	0.061	0.027	0.072	0.042	0.061
(10; 100)	0.034	0.081	0.059	0.032	0.035	0.073
Case B.2	$\beta_{00}=0.4; \beta_{01}=0.4; \beta_{10}=0; \beta_{11}=0$					
	α_{01}		α_{10}		α_{11}	
	0.1	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**
(5; 49)	0.016	0.005	0.019	0.002	0.004	0.018
(5; 100)	0.008	0.032	0.008	0.022	0.007	0.001
(10;49)	0.014	0.031	0.019	0.030	0.017	0.020
(10; 100)	0.029	0.003	0.031	0.032	0.033	0.019

In general terms, we can say that the likelihood ratio test of (18) is well sized, especially when the DGP is correctly specified (Case A). If the DGP is wrongly specified, the test suffers some size distortions, which are more acute when the omitted variable (z) presents some correlation with the variable of interest (x), as in Case B.2.

Table 2 summarizes the results corresponding to the estimated power of the test. We have used two different configurations of the DGP. Case C is a purely static model, in the sense that there is any spatial or temporal lag of variable y in the DGP. Case D introduces some elements of spatial or temporal dynamics of variable y, in combination with the structure associated to variable x. In the configuration D.1 it appears the temporal lag of y, the spatial lag in the configuration D.2 and the temporal lag of the spatial lag in D.3. Moreover, for simplicity's sake, we have introduced only one element associated to x: the contemporaneous and spatially coincident values, x_t (column β_{00}); the temporal lag of the spatially coincident values, x_{t-1} (column β_{10}); the spatial lag of the contemporaneous values, Wx_t (column β_{01}) or the spatial lag of the temporal lag of the values, Wx_{t-1} (column β_{11}).

Table 2: Causality likelihood ratio. Estimated power.

Case C	α_{10}=0; α_{01}=0; α_{11}=0							
	β_{00}		β_{10}		β_{01}		β_{11}	
	0.1	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**
(5; 49)	0.663	0.975	0.596	0.780	0.617	0.735	0.633	0.759
(5; 100)	0.786	0.979	0.673	0.886	0.758	0.822	0.744	0.890
(10;49)	0.727	0.983	0.699	0.887	0.657	0.855	0.651	0.881
(10; 100)	0.790	0.983	0.793	0.929	0.856	0.955	0.755	0.940
Case D.1	α_{10}=0.4; α_{01}=0; α_{11}=0							
	β_{00}		β_{10}		β_{01}		β_{11}	
	0.1	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**
(5; 49)	0.587	0.946	0.577	0.664	0.549	0.728	0.517	0.659
(5; 100)	0.780	0.964	0.617	0.778	0.700	0.805	0.631	0.884
(10;49)	0.678	0.921	0.686	0.831	0.545	0.762	0.632	0.851
(10; 100)	0.740	0.936	0.680	0.854	0.830	0.851	0.648	0.866
Case D.2	α_{10}=0; α_{01}=0.4; α_{11}=0							
	β_{00}		β_{10}		β_{01}		β_{11}	
	0.1	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**
(5; 49)	0.591	0.931	0.591	0.729	0.601	0.644	0.557	0.685
(5; 100)	0.691	0.951	0.665	0.809	0.734	0.752	0.651	0.829
(10;49)	0.660	0.945	0.611	0.846	0.587	0.813	0.593	0.834
(10; 100)	0.739	0.884	0.771	0.867	0.829	0.926	0.706	0.849
Case D.3	α_{10}=0; α_{01}=0; α_{11}=0.4							
	β_{00}		β_{10}		β_{01}		β_{11}	
	0.1	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**	**0.1**	**0.8**
(5; 49)	0.624	0.949	0.568	0.714	0.585	0.706	0.590	0.707
(5; 100)	0.763	0.973	0.662	0.852	0.703	0.776	0.715	0.828
(10;49)	0.703	0.923	0.633	0.857	0.628	0.819	0.646	0.822
(10; 100)	0.742	0.936	0.789	0.895	0.827	0.917	0.742	0.924

It is evident that the test performs better the simpler is the DGP. The best results correspond to Case C where a (spatially and temporally) static model intervenes. Overall, the percentage of rejection seems reasonable and it increases with the sample size (with R) and with the symptoms of causality, as reflected by the corresponding parameter β. Interestingly, the procedure works better when the cross-section analyzed is more distant from the origin.

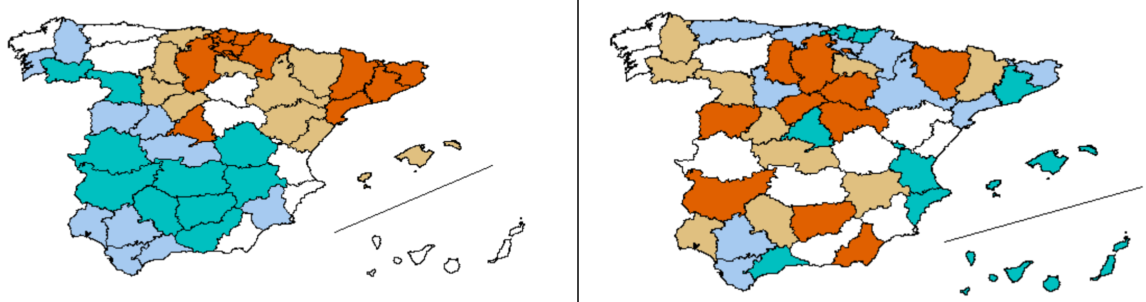
6- An application to the Spanish case: Personal income vs agriculture.

As an example, we present the case of the spatial distribution of the income per capita and the presence of the agricultural sector in the Spanish provinces in the year 2006. The two variables are represented in Figure1. The first (ipc) is measured as an index, with value 100 for the national average, and the second (ag) as the percentage that the agricultural sector represents on the gross value added of each province in 2006.

The first question to note is that the spatial distribution of both variables is very different. The two are positively spatially correlated, but the structure of the second is diffused (with a Moran's I of 0.15 and pvalue of 0.06) whereas the income per capita shows a strong Northeast-Southwest tendency (the Moran statistic is 0.66 with a pvalue of 0.00).

Figure 1: Spatial distribution of income per capita and weight of agriculture. 2006.

Income per capita (Spain=100) | **Weight of agriculture on PIB**



First (lowest) **Second** **Third** **Fourth** **Fifth (highest)**

To begin with the discussion, let us introduce in the first place the equation specified to relate both variables:

$$y = \alpha_0 + \sum_{k=1}^{2} \alpha_{1k} W_k y + \sum_{k=0}^{2} \alpha_{2k} W_k x + u \qquad (23)$$

It is the same equation that appears in (18) but restricting the lag structure to a second order (k_1=k_2=2), where y and x may correspond to variables ipc or ag. If we associate y with ipc and x with ag, we will test for causality relations from agriculture (cause) to income (effect), and on the contrary if we identify y with ag and ipc with x. In each of these two cases, we have obtained the SDLS estimate of the corresponding y variable, $\hat{y}$, using a reduced version of (23), namely:

$$y = \alpha_0 + \alpha_{11} W_1 y + \sum_{k=0}^{2} \alpha_{2k} W_k x + u \qquad (24)$$

This allows to complete the matrix z=[$W_1\hat{y}$, $W_2\hat{y}$, x, W_1x, W_2x] and proceed as indicated in section 4. The results are shown in the upper part of Table 3. Under the heading of ipc or ag, there appear the cumulative percentage obtained from the corresponding marginal coefficients. For example, the LS regression for the ipc variable produces a multiple correlation coefficient of 0.7667. The spatial structure of the variable ipc accounts for the 69.6% of this value, whereas the information of the ag variable only amounts to the 30.4%. In the case of the regression for the ag variable, the multiple coefficient is smaller, 0.5309, and depends mainly on the information associated to the variable ipc, 75.1%, whereas the spatial distribution of the agricultural sector only accounts for to the 24,9%. In short, it appears that income may have some effect in the distribution of the agricultural sector but the maintenance of the contrary relation it is, at least, dubious.

At the bottom of the Table, under the heading of ML estimation, we present the results of the likelihood ratio of (18). The pvalue of the first relation (agriculture does not cause the distribution of income) does not allows to reject the null hypothesis at the usual 5% significance level. In the second relation (income does not cause the distribution of agriculture) we observe a vey low pvalue for the F statistics which allows to reject the null hypothesis at the same significance level.

Table 3: Income pc vs agriculture. Causality results

(1) SDLS+ (2) LS Estimation				
Explained	**ipc**		**Explained**	**ag**
W1ipc	46.3%		**W1ag**	19.4%
W2ipc	69.6%		**W2ag**	24.9%
ag	15.3%		**ipc**	38.3%
W1ag	15.4%		**W1ipc**	74.6%
W2ag	30.4%		**W2ipc**	75.1%
Coeff. Corr	0.7667		**Coeff. Corr**	0.5309
ML Estimation				
Explained	**ipc**		**Explained**	**ag**
Log. ample	-202.193		**Log. ample**	-135.068
Log. restri.	-195.785		**Log. restri.**	-120.595
F statistic	6.408		**F statistic**	14.473
p-value	0.0934		**p-value**	0.0023

7- Conclusions.

This paper is a first approach to the analysis of causality in a spatial context. We are convinced that this is a very important topic that must be checked in order to assure the consistency of any spatial econometric model. There are obvious difficulties in tackling the question and the characteristic of the information used in this type of models is not a minor aspect. In this sense, the intention of our paper is to motivate the discussion.

We present some preliminary results in terms of an exploratory technique, based on the decomposition of the multiple correlation coefficient of an general regression, as well as a test which may be seen as an adjusted version of the popular Granger causality test.

References

Ancot, J., J. Kuiper and J Paelinck (1990): Five Principles of Spatial Econometric Illustrated. In Chatterjy, M. and R. Kuenne: ‘Dynamics and Conflict in Regional Structural Change’ (pp. 141-155). McMillan: Londres.

Anselin L. (1988): Spatial Econometrics: Methods and Models’. Kluwer: Dordrecht.

Beenstock, M. and D. Felsenstein (2007): Spatial Vector Autoregressions. Spatial Economic Analysis 2, 167-196.

Beenstock, M. and D. Felsenstein (2008): Regional Heterogeneity, Conditional Convergence and Regional Inequality. Regional Studies 42, 475-488.

Bresson, G. and Pirotte, A., 1995, Econométrie des séries temporelles, Théorie et applications. Presses Universitaires de France, Paris.

Charemza, W, and D. Deadman, 1997. New Directions in Econometric Practice. Cheltenham: Edward Elgar.

Davidson, J. (2000). *Econometric Theory*. Oxford: Blackwell.

Di Giacinto, V. (2003): Differential Regional Effects of Monetary Policy: A Geographical SVAR Approach. International Regional Science Review 26, 313-341.

Di Giacinto, V. (2006). A Generalized Space-Time ARMA model with an Application to Regional Unemployment Analysis in Italy. International Regional Science Review 29: 159-198.

Elhorst, J.P., 2001. Dynamic models in space and time. Geographical Analysis 33, 119–140.

Elhorst JP (2003): Specification and Estimation of Spatial Panel Data Models. International Regional Science Review 26: 244-268

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

Granger, C. (1969): Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. Econometrica, 37, 424-438.

Greene, W.H., 2007, Econometric Analysis, Pearson/Prentice Hall, Upper Saddle River, New Jersey.

Griffith, D. (2003): Spatial Autocorrelation and Spatial Filtering. Springer: Berlin

Hendry, D. and J. Richard (1983): The Econometric Analysis of Economic Time Series (with discussion). International Statistical Review, 51, 111-163.

Heyde, J.E.(1957): Entwertung der Kausalität. Kohjlhammer: Stuttgart.

Holly, A. (1984): Tests d’exogénéité. I.M.E. Document de travail No 71, Université de Dijon.

Paelinck, J.H.P. (1990), Some new estimators in spatial econometrics, in D.A. Griffith (ed.) *Spatial Statistics: Past, Present and Future*, Syracuse University, Institute of Mathematical Geography, Monograph No 12, 163-181

Paelinck, J.H.P., (2007): A mixed specification for Lotka-Volterra models, with endogenously generated SDLS variables, Working Paper.

Paelinck J. and P. Nijkamp (1978). Operational Theory and Method in Regional Economics. Saxon House.

Paelinck J and L. Klaassen (1979): Spatial Econometrics. Saxon House: Farnborough.

Pearl, J. (2000): Causality: Models, Reasoning and Inference. Cambridge: Cambridge University Press.

Sims, C. (1980): Macroeconomics and Reality. Econometrica, 48, 1-48.

Theil, H. (1971): Principles of Econometrics. North-Holland Publishing Company: Amsterdam-London