

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/30463862>

# An Introduction to Spatial Econometrics

Article in *Revue D Économie Industrielle* · September 2008

DOI: 10.4000/rei.3887 · Source: OAI

---

CITATIONS

674

READS

4,063

1 author:



[James P. Lesage](#)

University of Toledo

222 PUBLICATIONS 13,902 CITATIONS

[SEE PROFILE](#)

# AN INTRODUCTION TO SPATIAL ECONOMETRICS

**Mots-clés :** Économétrie spatiale, dépendance spatiale, processus spatial autorégressif.

**Key words :** Spatial Econometrics, Spatial Dependence, Spatial Autoregressive Processes.

## I. — INTRODUCTION

Spatial regression methods allow us to account for dependence between observations, which often arises when observations are collected from points or regions located in space. The observations could represent income, employment or population levels, tax rates, and so on, for European Union regions delineated into NUTS regions, countries, postal or census regions (1). We might also have individual firm establishment point locations referenced by latitude-longitude coordinates that can be found by applying geo-coding software to the postal address. It is commonly observed that sample data collected for regions or points in space are not *independent*, but rather *spatially dependent*, which means that observations from one location tend to exhibit values similar to those from nearby locations.

There are a number of theoretical motivations for the observed dependence between nearby observations. For example, Ertur and Koch (2007) use a theoretical model that posits physical and human capital externalities as well as

(\*) McCoy College of Business Administration - Department of Finance and Economics - Texas State University - San Marcos.

(1) NUTS is a French acronym for Nomenclature of Territorial Units for Statistics used by Eurostat. In this nomenclature NUTS1 refers to European Community Regions and NUTS2 to Basic Administrative Units. For the United Kingdom NUTS1 is typically used because there is no official counterpart to NUTS2 units which are drawn up only for the European Commission use as groups of counties.

technological interdependence between regions. They show that this leads to a reduced form growth regression that should include an average of growth rates from neighboring regions. In time series, time dependence is often justified by theoretical models that include costly adjustment or other behavioral frictions which give rise quite naturally to *time lags* of the dependent variable. The theoretical work of Ertur and Koch (2007) is similar in spirit, using the notion of « spatial diffusion with friction » to provide a motivation for a *spatial lag*, which takes the form of an average of neighboring regions.

Another justification is that observed variation in the dependent variable may arise from unobserved or latent influences. Latent unobservable influences related to culture, infrastructure, recreational amenities and a host of other factors for which we have no available sample data can be accounted for by relying on neighboring values taken by the dependent variable. This works when the latent influences change slowly as we moved across regions.

Conventional regression models commonly used to analyze cross-section and panel data assume that observations/regions are independent of one another. As an example, a conventional regression model that relates commuting times to work for region  $i$  to the number of persons in region  $i$  utilizing different commuting modes and the density of commuters in region  $i$ , assumes that mode choice and density of a neighboring region, say  $j$  does not have an influence on commuting time for region  $i$ . Since it seems unlikely that region  $i$ 's network of vehicle and public transport infrastructure is independent from that of region  $j$ , we would expect this assumption to be unrealistic. Ignoring this violation of independence between observations will produce estimates that are biased and inconsistent.

Spatial econometrics is a field whose analytical techniques are designed to incorporate dependence among observations (regions or points in space) that are in close geographical proximity. Extending the standard linear regression model, spatial methods identify cohorts of « nearest neighbors » and allow for dependence between these regions/observations (Anselin, 1988 ; LeSage, 2005). Note that even with observational units such as firms operating in world markets where the notion of spatial proximity is not appropriate, we might still see dependence in behavior of « peer institutions », those that are most similar to each other. The spatial regression methods described here can be applied to these situations by relying on the analogy that the set of say «  $m$ -nearest neighbors » from the case of spatial regression can be construed as a group of «  $m$  peer institutions ». This is a generalization of neighbors based on distance that could be used to structure dependence in behavior, leading to a model that is formally analogous to the geographical « nearest-neighbors » considered here.

The next section introduces a spatial regression methodology that accommodates spatial dependence. Spatial autoregressive processes are introduced since they are a key component of these models. Use of these models for cases involved non-spatial structured dependence between cross-sectional observations is also discussed.

In section 3 we discuss methods for estimating these models and ways to compare models based on different specifications and spatial connectivity structures. Examples are provided using our applied example based on a commuting time regression relationship.

Interpretation of the parameter estimates from these models is the subject of section 4.

Section 5 illustrates spatial regression estimates and inferences along with analysis of spatial feedback impacts using an applied illustration that relates commuting times and explanatory variables based on a Census sample of 3,110 US counties in the lower 48 states and District of Columbia.

## II. — SPATIAL REGRESSION MODELS

Relaxing the conventional assumption of independent observations in a cross-sectional setting requires that we provide a parsimonious way to specify structure for the dependence between the  $n$  observational units that make up our size  $n$  data sample. In our application, we wish to structure the dependence to reflect the relationship between commuting times from one observation (county) and neighboring counties.

### 2.1. Spatial autoregressive processes

The spatial autoregressive process shown in (1) and the implied data generating process in (2) provide a parsimonious approach to representing the dependence structure.

$$y = \alpha u_n + \rho W y + \varepsilon \quad (1)$$

$$(I_n - \rho W)y = \alpha u_n + \varepsilon$$

$$y = (I_n - \rho W)^{-1} \alpha u_n + (I_n - \rho W)^{-1} \varepsilon \quad (2)$$

$$\varepsilon \sim N(0_{n \times 1}, \sigma^2 I_n)$$

We introduce a constant term vector  $u_n$  and associated parameter  $\alpha$  to accommodate situations where the vector  $y$  does not have a mean value of zero. The  $n$  by 1 vector  $y$  contains our dependent variable and  $\rho$  is a scalar parameter, with  $W$  representing an  $n$  by  $n$  spatial weight matrix. We assume that  $\varepsilon$  follows a multivariate normal distribution, with zero mean and a constant scalar diagonal variance-covariance matrix  $\sigma^2 I_n$ .

The matrix  $W$  quantifies the connections between regions, which we illustrate by first considering a 5 by 5 matrix  $W$  associated with 5 regions. For sim-

plicity, consider that connections only exist between each region and  $m = 2$  neighboring regions.

Given these assumptions, we can form an  $n$  by  $n$  binary indicator matrix  $P$  shown in (3), where the rows of the matrix correspond to observations/regions 1 to 5. Values of 1 are used in each column to indicate « neighboring » observations associated with each row. For example,  $P(1,2) = 1$  and  $P(1,3) = 1$  indicates that the second and third observations/regions represent the two nearest (measured using distance from the center of each region). This reflects that regions #2 and #3 are the nearest neighbors to region #1, which meets our definition of  $m = 2$  neighbors. Similarly, in row 2 we have  $P(2,1) = 1$  and  $P(2,3) = 1$ , indicating that regions 1 and 3 are the  $m = 2$  nearest neighbors to region #2. Similarly, region #5 is a neighbor to regions #3 and #4.

$$P = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix} \quad (3)$$

The main diagonal elements of  $P$  are zero to prevent an observation from being defined as a neighbor to itself. We can normalize the matrix  $P$  to have row-sums of unity by dividing all elements of the matrix  $P$  by the number of neighbors, in this case  $m = 2$ . This leads to a matrix we label  $W$ . This *row-stochastic* form of the spatial weight matrix will be useful for expressing our spatial regression model.

$$W = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{pmatrix} \quad (4)$$

Consider the product of the matrix  $W$  and a vector of observations  $y$  on commuting times for the five regions shown in (5). This matrix product known as a *spatial lag* produces an  $n$  by 1 vector containing an average of commuting times from regions defined as neighbors by the matrix  $P$ .

$$Wy = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix}$$

$$= \begin{pmatrix} 1/2y_2 + 1/2y_3 \\ 1/2y_1 + 1/2y_3 \\ 1/2y_2 + 1/2y_4 \\ 1/2y_3 + 1/2y_5 \\ 1/2y_3 + 1/2y_4 \end{pmatrix} \quad (5)$$

## 2.2. Spatial regression models

We can use the spatial autoregressive process in (3) to construct an extension of the conventional regression model shown in (6), along with the associated *data generating process* in (7). The model has been labeled the spatial autoregressive (SAR) model. The dependent variable vector  $y$  is of dimension  $n$  by 1, containing (logged) commuting times for each region/observation. The  $n$  by  $k$  matrix  $X$  contains exogenous explanatory variables possibly including a constant term vector, and the  $k$  by 1 vector  $\beta$  are associated regression parameters. The  $n$  by 1 spatial lag vector  $Wy$  reflects an average of (log) commuting times from neighboring regions specified by the matrix  $W$ , and the associated scalar parameter  $\rho$  reflects the strength of spatial dependence. When the scalar parameter  $\rho$  takes on a value of zero, the model in (6) simplifies to the conventional linear regression model. Finally, we assume the  $n$  by 1 disturbance vector  $\varepsilon$  contains independent, normally distributed terms with a vector mean zero ( $0_{n \times 1}$ ), constant variance, ( $\sigma^2$ ).

$$y = \rho Wy + X\beta + \varepsilon \quad (6)$$

$$y = (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1} \varepsilon \quad (7)$$

$$\varepsilon \sim N(0_{n \times 1}, \sigma^2 I_n)$$

The spatial regression model adds a spatial lag vector reflecting the average commuting times from neighboring regions to help explain variation in commuting times across the regions. Intuitively, the model states that commuting times in each region are related to the average commuting times from neighboring regions. The average strength of this relationship across the sample of regions will be determined during estimation by the scalar parameter  $\rho$ .

Note that  $\rho$  is not a conventional correlation coefficient between the vector  $y$  and the spatial lag vector  $Wy$ , since this parameter is not generally restricted to the range  $-1$  to  $1$  (see LeSage and Pace (2004b) for details regarding the bounds on the parameter  $\rho$ ). From (7) we can see that the variance-covariance for the spatial regression is :  $E[(I_n - \rho W)^{-1} \varepsilon \varepsilon' (I_n - \rho W)^{-1}]$ . We need to ensure that  $(I - \rho W)$  is non-singular and that the product  $(I - \rho W)^{-1} (I - \rho W)^{-1}$  which equals the variance-covariance matrix is positive-definite. For row-stochastic weight matrices  $W$  where the row elements sum to unity, a sufficient condition for positive-definite variance covariance matrices is that  $-1 < \rho < 1$ . The eigenvalues of non-symmetric weight matrices can be complex, but this overly strong restriction ensures invertibility and positive definiteness.

### 2.3. Simultaneous feedback

Simultaneous feedback is a feature of the spatial regression model that arises from dependence relations. These lead to feedback effects from changes in commuting times in neighboring regions  $j$  that arise from a change originating in region  $i$ . To see this, consider the data generating process associated with the spatial regression model, shown in (8).

$$y = (I_n - \rho W)^{-1} X \beta + (I_n - \rho W)^{-1} \varepsilon \quad (8)$$

The model statement in (8) can be interpreted as indicating that the expected value of each observation  $y_i$  will depend on the mean value  $X \beta$  plus a linear combination of values taken by neighboring observations scaled by the dependence parameter  $\rho$ . The *data generating process* statement in (8) expresses the *simultaneous* nature of the spatial autoregressive process. To further explore the nature of this, we use the following well-known expansion to express the inverse as an infinite series Debreu and Herstein (1953) :

$$(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 + \dots \quad (9)$$

Which leads to a re-expression of spatial autoregressive data generating process for a variable vector  $y$  :

$$\begin{aligned} y &= (I_n - \rho W)^{-1} X \beta + (I_n - \rho W)^{-1} \varepsilon \\ y &= X \beta + \rho W X \beta + \rho^2 W^2 X \beta + \dots \\ &+ \varepsilon + \rho W \varepsilon + \rho^2 W^2 \varepsilon + \rho^3 W^3 \varepsilon + \dots \end{aligned} \quad (10)$$

Consider powers of the row-stochastic spatial weight matrices  $W^2, W^3, \dots$  that appear in (10), where we assume that rows of the weight matrix  $W$  are constructed to represent *first-order* contiguous neighbors. The matrix  $W^2$  will reflect *second-order* contiguous neighbors, those that are neighbors to the first-order neighbors. Since the neighbor of the neighbor (second-order neighbor) to an observation  $i$  includes observation  $i$  itself,  $W^2$  has positive elements on the diagonal. That is, higher-order spatial lags can lead to a connectivity relation for an observation  $i$  such that  $W^2 \varepsilon$  will extract observations from the vector  $\varepsilon$  that point back to the observation  $i$  itself. This is in stark contrast with the conventional independence relation in ordinary least-squares regression where the Gauss-Markov assumptions rule out dependence of  $\varepsilon_i$  on other observations  $j$ , by assuming zero covariance between observations  $i$  and  $j$  in the data generating process.

### 2.4. Steady-state equilibrium interpretation

One might suppose that feedback effects would take time, but there is no explicit role for passage of time in our cross-sectional model. Instead, we view

the cross-sectional sample of regions as reflecting an equilibrium outcome or steady state of the commuting time generation process where the regional characteristics change slowly relative to commuting times. To elaborate this point, consider a relationship where travel time to work at time  $t$  is denoted by  $y_t$ , and this depends on current period own-region characteristics  $X_t\beta$  plus the spatial dependence on observed travel times from an average of neighboring regions from a single past period,  $t-1$ . This is represented by a *space-time lag* variable  $Wy_{t-1}$ , leading to the model in (11), where the regional characteristics in the matrix of explanatory variables  $X_t$  are assumed to not change over time.

$$y_t = \rho Wy_{t-1} + X\beta + \varepsilon_t \quad (11)$$

Intuitively, ignoring the time subscript for regional characteristics in (11) seems reasonable, since choice of transportation mode by population living in a region, regional transportation infrastructure, commuting density, etc., are not likely to exhibit a great deal of variation over time (2).

Note that we can replace  $y_{t-1}$  on the right-hand-side of (11) with:  $y_{t-1} = \rho Wy_{t-2} + X\beta + \varepsilon_{t-1}$  and continue this for  $q$  past periods leading to (12) and (13).

$$y_t = (I_n + \rho W + \rho^2 W^2 + \dots + \rho^q W^q) X\beta + \rho^q W^q y_{t-q} + u \quad (12)$$

$$u = \varepsilon_t + \rho W\varepsilon_{t-1} + \rho^2 W^2\varepsilon_{t-2} + \dots + \rho^{q-1} W^{q-1}\varepsilon_{t-(q-1)} \quad (13)$$

In (13),  $E(u) = 0$  since  $E(\varepsilon_{t-i}) = 0, i = 0, \dots, q$ . The magnitude of  $\rho^q W^q$  becomes small for large  $q$ , since  $\rho < 1$  and the maximum eigenvalue for a row-stochastic matrix  $W$  is one. In the limit, as  $q \rightarrow \infty$ ,  $\rho^q W^q y_{t-q}$  approaches a vector of zeros.

Therefore, the long-run expectation, which can be interpreted as the steady-state equilibrium, takes a form consistent with the data generating process for our cross-sectional spatial regression model :

$$\begin{aligned} \lim_{q \rightarrow \infty} E(y_t) &= (I_n + \rho W + \rho^2 W^2 + \dots + \rho^q W^q) X\beta + \rho^q W^q y_{t-q} + u \\ &= (I - \rho W)^{-1} X\beta \end{aligned} \quad (14)$$

Simultaneous feedback is a feature of the equilibrium steady-state for the spatial regression model. Feedback effects can arise from changes in commuting times of one region that will potentially exert impacts on all other regions. However, the fact that the parameter  $-1 < \rho < 1$  leads to a decay of influence as we move to higher-order neighbors.

- (2) It is possible to produce a similar result as that shown here if the explanatory variables  $X_t$  evolve over time in a number of ways. For example, a stochastic trend, stochastic geometric, or exponential growth in  $X_t$  with mean zero randomness, a stationary autoregressive time or spatial process governing  $X_t$ . Therefore, the assumption of fixed  $X$  is for simplicity.



Intuitively, if changes arise in neighboring regions travel times and region  $i$ 's travel time is dependent on neighboring region commuting times, these changes will exert an impact on travel times of region  $i$ . Similarly, any changes in region  $i$ 's commuting times arising from these impacts will in turn influence regions that are neighbors to  $i$ , which will in turn influence neighbors of the neighbors, and so on. In the context of our static cross-sectional model where we treat the observed sample as reflecting a steady state equilibrium outcome, these feedback effects appear as instantaneous, but they should be interpreted as reflecting a movement to the next steady state.

## 2.5. Non-spatial structured dependence

Social science regression models commonly applied to cross-section and panel data assume observations on decision-making units are independent of one another. This assumption is important to contemplate since violation results in regression estimates that are biased and inconsistent. If there is a pattern of dependence between the observations that quantified using the matrix  $W$  presented in section 1, then spatial regression methods can be employed in these non-spatial contexts.

For example, in the case of a cross-sectional sample of firms, one might replace geographical distance with measures of similarity such as firm size or industry similarity. Blankmeyer *et al.* (2007) point out that salary benchmarking practices used in determining US management compensation violate the assumption of independence. Salary benchmarking is a US practice that adjusts top-level management compensation to reflect salaries offered by « peer institutions », those against which firms compete when hiring high-level managers. They point out that spatial regression models based on « nearest neighbors » that allow dependence between observations can be extended to the case of management compensation. In this setting, the observational units are institutions, not regions or points located in space. However, they rely on the analogy that the set of say «  $m$ -nearest neighbors » from the case of spatial regression can be construed as a group of «  $m$  peer institutions » that are most similar to each observation/institution in the sample. They point out that use of generalized notions of distance allows geographical « nearest-neighbors » to be treated formally as analogous to institutional peer groups.

A key step is to replace measures of geographical proximity based on distances by a criterion that reflects the degree of similarity among institutions for the case of a cross-section of firms. The context of similarity would be specific to each problem, so this might reflect similarity in size of the firms, production processes, resource or product markets in which the firms operate, etc. Blankmeyer *et al.* (2007) illustrate these ideas applying spatial regression models to management compensation in Texas nursing facilities. Similar or « peer institutions » were defined using alternative measures of size such as the square foot area of the facilities, the size of the nursing staff and expenditures related to the nursing functions. Given univariate or multivariate criteria of ins-

titudinal similarity, conventional measures of univariate or multivariate distance (e.g. Euclidean or Mahalanobis distance) can be calculated. Peer institutions to each observation were identified as those that are most similar, that is, those that exhibited smaller distances constructed using the similarity criterion.

Another structured dependence context in which spatial regression models might be used would be what is known as network autocorrelation. Black (1992) distinguished between network autocorrelation and spatial autocorrelation affecting bilateral flows, and suggested these would lead to bias in classical estimation procedures typically used for spatial interaction models. He suggested that « *spatial autocorrelation usually concerns itself with variable values at given locations being influenced by variables values at nearby or (contiguous) locations in a spatial context. Network autocorrelation concerns the dependence of variable values on given links to such values on other links to which it is connected in a network context* ». Then, he suggested that « *autocorrelation may also exist among random variables associated with the links of a network, although this has not been examined previously* ». This suggests uses for the spatial regression models in the context of various types of flows between firms operating in a connected network context.

It is interesting to note that the social networking literature has relied on a similar approach to defining connectivity between individuals located at nodes in a network of social relationships. (Katz, 1953 ; Bonacich, 1987) interpret the vector  $b = (In - \rho P)^{-1} \iota_n$  as a measure of *centrality* of individuals in a social network, where the matrix  $P$  is a binary matrix like our (3). In this case, the vector  $b$  reflects row sums of the matrix inverse and has been referred to as Katz-Bonacich Centrality in social networking. This measures the number of direct and indirect connections that an individual in a social network has. For example, if the matrix  $P$  identifies friends, then  $P^2$  points to friends of friends,  $P^3$  to friends of friends of friends, and so on. In social networking, individuals are considered located at nodes in a network, and the parameter  $\rho$  reflects a discount factor that creates decay of influence for friends/peers that are located at more distant nodes.

Also interesting is that Ballester Calvó-Armengol and Zenou (2006) use this social networking context to considering players in a noncooperative network game with linear quadratic payoffs as a return to effort inputs, where the game exhibits local complementarity with efforts of other players. They show that in the case of a simultaneous move  $n$  player game, there is a unique (interior) Nash equilibrium for effort exerted (Ballester Calvó-Armengol and Zenou, 2006, Theorem 1, Remark 1). The equilibrium effort exerted by each player is proportional to a heterogenous variant of the Katz-Bonacich centrality of the player's node in the network that replaces the vector  $\iota_n$  in the measure  $b$  with a heterogeneous vector. Their results apply to both symmetric and asymmetric structures of complementarity across the players represented by the  $n$  by  $n$  matrix  $P$ .

A final motivation arises from the interregional trade and input-output literature, where obvious connections exist between the Bonacich index and

« feedback loops analysis » developed in the input-output literature (Sonis, Hewings and Okuyama, 2001). Their goal was identifying hierarchies of regions based on their interregional sectoral linkages in an interregional framework, where the chain of bilateral interregional/interindustry influences was based on sectoral linkages. The economic interpretation of a feedback loop indicates how strongly each region is tied to all other regions included in that loop. Focusing on feedback loops, they evaluated the position of each region vis-à-vis all other regions. Of course, usually the rank of these bilateral relations between regions is conditioned by the distance between them (contiguity), the presence of common factor endowments, similar or complementary sectoral structures (clusters) and strong networking linkages.

Note that both the Bonacich index and feedback loop approach focus on measuring centrality of an element within a system in space using their linkages with other elements in the network, but this could also be done for the cases of industry sectoral structure or size similarity as noted earlier. In all cases, the measure of inter-linkages is based on identifying the number of possible complete loops that start and end at each node based on connections with the remaining elements in the system.

We do not explore these non-spatial uses of spatial regression, since our focus is on exposition of statistical modeling implications of spatial dependence. However, in the conclusion of this article a number of issues that may arise in extending spatial regression methods to the case of structured dependence relations between firms are discussed.

## 2.6. Alternative spatial regression specifications

We will employ a variation of model (6) in our application. Our model specification will allow characteristics that determine commuting times (variables contained in the matrix  $X$ ) from neighboring regions to exert an influence on commuting times of region  $i$ . This is accomplished by entering an average of the explanatory variables from neighboring regions, created using the matrix product  $W X$ . The resulting model is shown in (15), where we have eliminated the constant term vector  $\iota_n$  from the explanatory variables matrix  $X$  (3).

$$y = \rho W y + \alpha \iota + X \beta + W X \theta + \varepsilon \quad (15)$$

$$\varepsilon \sim N(0_{n \times 1}, \sigma^2 I_n)$$

This variant of our original model often labeled the spatial Durbin model (SDM) allows commuting times for each region to depend on own-region fac-

(3) This is necessary because  $W \iota_n = \iota_n$  (for row-stochastic matrices  $W$ ), which would result in the explanatory variables set ( $X \ W X$ ) containing a perfect linear combination.

tors from the matrix  $X$  that influence commuting times, plus the same factors averaged over the  $m$  neighboring regions,  $W X$ .

Other specifications that rely on the spatial autoregressive process can be used to produce spatial regression models that exhibit: spatial dependence in the disturbance process  $\varepsilon$  which leads to a spatial error model (SEM) (shown in (16)); and a general model that exhibits spatial dependence in both the dependent variable  $y$  and the disturbances leading to a model we label SAC (shown in (17)). The SAC model in (17) can be implemented with a single spatial weight matrix,  $W_1 = W_2 = W$ .

$$y = X \beta + u, \quad u = \lambda W u + \varepsilon \quad (16)$$

$$y = \rho W_1 y + X \beta + u, \quad u = \lambda W_2 u + \varepsilon \quad (17)$$

There are some other more exotic spatial regression models that have been implemented in the literature. For example, Lacombe (2004) uses the model shown in (18) to analyze policies that varied across states.

$$y = \rho_1 W_1 y + \rho_2 W_2 y + X \beta + \varepsilon \quad (18)$$

The model involves a sample of counties that lie on the borders of US states, and the spatial weights  $W_1$  represent an average of the variable  $y$  based on neighboring counties within the state, while the weights  $W_2$  reflect an average of the dependent variable from neighboring counties in the bordering state. This model separates the influence of within- and between-state neighbors on the dependent variable  $y$ .

There is also the SARMA model shown in (20), where we could implement the model with  $W = W_1 = W_2$ .

$$y = \rho W_1 y + X \beta + u, \quad u = (I_n - \theta W_2) \varepsilon \quad (19)$$

For the case of a single matrix  $W$ , the spatial process applied to the disturbances in this model can be written in expanded form as :

$$\begin{aligned} (I_n - \rho W)^{-1}(I_n - \theta W) &= I_n - \theta W + \rho W(I_n - \theta W) + \rho^2 W^2(I_n - \theta W) + \dots \\ &= I_n - \theta W - \rho \theta W^2 - \rho^2 \theta W^3 + \dots \\ &\quad + \rho W + \rho^2 W^2 + \dots \end{aligned} \quad (20)$$

This suggests a relatively rapid decay of influence for the terms in (20) contributed by the spatial moving average, those involving the parameter  $\theta$ . To see this, consider parameter values of  $\rho = \theta = 0.5$ , which would result in the terms  $-\rho\theta$ ,  $-\rho^2\theta$ ,  $\rho^3\theta$  taking on values of 0.25, 0.125, 0.0625. Further, these decay factors are being applied to smaller and smaller weight elements contained in the higher powers of the matrix  $W$ . These models have a « local inter-

pretation » in contrast to the SAR, SAC, SDM models that are interpreted as modeling « global influences ». They have received much less attention in the spatial econometrics literature.

### III. — ESTIMATION AND MODEL COMPARISON

Ordinary least-squares cannot be used to produce consistent estimates for spatial regression models. A number of different approaches have been proposed for estimating the parameters of spatial regression models, including: maximum likelihood estimation (Ord, 1975), an instrumental variables generalized moments (IV/GM) approach suggested by (Kelejian and Prucha, 1998, 1999), spatial filtering (Griffith, 2003), Bayesian Markov Chain Monte Carlo (LeSage, 1997), generalized maximum entropy (Marsh and Mittelhammer, 2004), and use of matrix exponential transformations (LeSage and Pace, 2007).

There have also been methods proposed for extensions of spatial regression models that deal with: binary dependent variables (LeSage, 2000 ; Smith and LeSage, 2004), polychotomous dependent variables (Autant-Bernard, LeSage and Parent, 2007), poisson distributed dependent variables (LeSage, Fischer and Scherngell, 2007), censored and missing values of the dependent variables (LeSage and Pace, 2007), and space-time panel data settings that involve continuous (Elhorst, 2003), and binary dependent variables (Kazuhiko, Polasek, Wago, 2007).

Details are beyond the scope of this work, but a great deal of public domain software allows these models to be estimated in a relatively straightforward manner (Anselin, 2006 ; LeSage, 1999 ; Pace, 2003).

The log likelihood function for the SAR and SDM models takes the form in (22) (Anselin, 1988, p. 63), where :  $Z = (\iota_n X)$  for the SAR model, and  $Z = (\iota_n X W X)$  for the SDM model. In applied practice, this is usually concentrated with respect to the coefficient vector  $\delta$  and the noise variance parameter  $\sigma^2$ , (see Pace and Barry (1997) for details).

$$\ln L = - (n/2)\ln(\pi\sigma^2) + \ln|I_n - \rho W| - (1/2\sigma^2)(e'e) \quad (21)$$

$$e = y - \rho W y - Z \delta$$

$$\delta = (Z'Z)^{-1} Z'(I_n - \rho W)y$$

$$\sigma^2 = e'e/(n - k)$$

The most challenging part of maximizing the log-likelihood involves computing the term :  $\ln|I_n - \rho W|$  which involves an  $n$  by  $n$  matrix, and this same issue arises for Bayesian MCMC estimation. A number of suggestions have been made for dealing with this in a computationally efficient manner (Pace

and Barry, 1997 ; Barry and Pace, 1999), which are discussed in detail by LeSage and Pace (2004b).

### 3.1. Model comparison

An issue that arises in applied practice is the need to compare models based on : 1) alternative spatial weight matrix specifications (*e.g.*, five versus six nearest neighbors, or contiguity-based  $W$  versus distance or nearest neighbors structures), 2) alternative sets of explanatory variables for the matrix  $X$ , and 3) varying spatial regression model specifications (*e.g.*, the SAR versus SDM or other model specifications).

For models estimated using maximum likelihood methods, likelihood ratio test statistics can be used to address some of these model comparison issues such as alternative specifications for the spatial weight matrix, or perhaps different sets of explanatory variables. Comparison of different model specifications using likelihood-based testing is set forth in (Florax and Folmer, 1992 ; Florax, Folmer and Rey, 2003).

There is a great deal of literature on Bayesian model comparison for regression models, where alternative models consist of those based on differing matrices of explanatory variables. For example, Koop (2003) sets forth the basic Bayesian theory behind model comparison for the case where a *finite, discrete* set of  $m$  alternative models  $M = M_1, M_2, \dots, M_m$  is under consideration. The approach involves specifying prior probabilities for each model as well as prior distributions for the parameters  $\pi(\eta)$ , where  $\eta = (\rho, \delta, \sigma)$  [*e.g.*, Koop (2003)]. Posterior model probabilities are then calculated and used for inferences regarding the alternative models based on different sets of explanatory variables.

We illustrate model comparison based on differing numbers of neighbors  $m$  used to define the spatial weight matrix in our applied illustration. If the sample data are to determine the posterior model probabilities, the prior probabilities should be equal, making each model equally likely *a priori*. These are combined with the likelihood for  $y$  conditional on  $\eta$  as well as the set of models  $M$ , which we denote  $p(y|\eta, M)$ . The joint probability for  $M, \eta$ , and  $y$  takes the form :

$$p(M, \eta, y) = \pi(M)\pi(\eta|M)p(y|\eta, M). \quad (22)$$

Application of Bayes rule produces the joint posterior for both models and parameters as :

$$p(M, \eta, y) = \frac{\pi(M)\pi(\eta|M)p(y|\eta, M)}{p(y)}. \quad (23)$$

The posterior probabilities for the models take the form :

$$p(M|y) = \int p(M, \eta|y) d\eta, \quad (24)$$

which requires integration over the parameter vector  $\eta$ .

An issue in Bayesian model comparison is that posterior model probabilities can be sensitive to alternative specifications for the prior information. In our applied illustration, we avoid this problem by relying on diffuse priors for the model parameters. Since the explanatory variables are fixed in alternative models with changes in only the spatial weight matrix, the log-marginal likelihood is well-defined for models based on uninformative priors [see (Koop, 2003)]. For other types of model comparisons, see LeSage and Parent (2007) who describe assignment of strategic priors that produce inferences that are robust with respect to the prior setting.

LeSage and Parent (2007) derive explicit expressions for the log-marginal likelihood needed for our comparison of models based on differing numbers of neighbors. The resulting expression requires univariate numerical integration of the parameter  $\rho$  over the  $(-1,1)$  interval. In conclusion, we can compare alternative model specifications based on varying numbers of neighbors used in constructing the spatial weight matrix. It is also possible to compare models with weight matrices based on different distance measures used to define neighboring regions, for example « travel time distances » versus « map distances ». This would involve application of the same methods to an additional set of  $m$  models,  $m = 1, \dots, M$  based on an alternative weight matrix, say  $\tilde{W}$ . Posterior model probabilities could then be calculated for the set of  $2M$  models to determine both the number of neighbors as well as the appropriate distance criterion.

#### IV. — INTERPRETING THE PARAMETER ESTIMATES

If  $\rho \neq 0$ , then the interpretation of the parameter vectors  $\beta$  (and  $\theta$ ) in the spatial Durbin model is different from a conventional least squares interpretation, (Pace and LeSage, 2006). In least-squares the  $r$ th parameter,  $\beta_r$ , from the vector  $\beta$ , is interpreted as representing the partial derivative of  $y$  with respect to a change in the  $r$ th explanatory variable from the matrix  $X$ , which we write as  $x_r$ . Specifically, in standard least-squares regression where the dependent variable vector contains *independent* observations,  $y = \sum_{r=1}^k x_r \beta_r + \varepsilon$ , and the partial derivatives of  $y_i$  with respect to  $x_{ir}$  have a simple form :  $\partial y_i / \partial x_{ir} = \beta_r$  for all  $i, r$  ; and  $\partial y_i / \partial x_{jr} = 0$ , for  $j \neq i$  and all variables  $r$ .

One way to think about this is that the information set in least-squares for an observation  $i$  consists only of exogenous or predetermined variables associated with observation  $i$ . Thus, a linear regression specifies :  $E(y_i) = \sum_{r=1}^k x_{ir} \beta_r$ , and takes a restricted view of the information set by virtue of the independence assumption.

#### 4.1. Spatial regression model partial derivatives

In our spatial dependence model, the least-squares interpretation is not valid. In essence, dependence expands the information set to include information from neighboring regions. To see the impact of this, consider the SDM model from (15) expressed as :

$$(I_n - \rho W)y = X\beta + WX\theta + \iota_n\alpha + \varepsilon$$

$$y = \sum_{r=1}^k S_r(W)x_r + V(W)\iota_n\alpha + V(W)\varepsilon \quad (25)$$

$$S_r(W) = V(W)(I_n\beta_r + W\theta_r)$$

$$V(W) = (I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 + \dots$$

To illustrate the role of  $S_r(W)$ , consider the expansion of the data generating process in (26) as shown in (27).

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \sum_{r=1}^k \begin{pmatrix} S_r(W)_{11} & S_r(W)_{12} & \dots & S_r(W)_{1n} \\ S_r(W)_{21} & S_r(W)_{22} & & \\ \vdots & \vdots & \ddots & \\ S_r(W)_{n1} & S_r(W)_{n2} & \dots & S_r(W)_{nn} \end{pmatrix} \begin{pmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{nr} \end{pmatrix} \quad (26)$$

$$+ V(W)\iota_n\alpha + V(W)\varepsilon$$

To make the role of  $S_r(W)$  clear, consider the determination of a single dependent variable observation  $y_i$  shown in (27).

$$y_i = \sum_{r=1}^k [S_r(W)_{i1}x_{1r} + S_r(W)_{i2}x_{2r} + \dots + S_r(W)_{in}x_{nr}]$$

$$+ V(W)\iota_n\alpha + V(W)\varepsilon \quad (27)$$

It follows from (28) that the derivative of  $y_i$  with respect to  $x_{jr}$  takes a much more complicated form :

$$\frac{\partial y_i}{\partial x_{jr}} = S_r(W)_{ij} \quad (28)$$

In contrast to the least-squares case, the derivative of  $y_i$  with respect to  $x_{ir}$  usually does not equal  $\beta_r$ , and the derivative of  $y_i$  with respect to  $x_{jr}$  for  $j \neq i$  usually does not equal 0. Therefore, any change to an explanatory variable in a single region (observation) can affect the dependent variable (commuting time) in all regions (observations). This is of course a logical consequence of our simultaneous spatial dependence model since it takes into account other regions' commuting times, and these are determined by the characteristics of those regions. Any change in the characteristics of neighboring regions that set



in motion changes in commuting times will impact commuting times of neighboring regions, and so on.

In the case of the own derivative for the  $i$ th region,

$$\frac{\partial y_i}{\partial x_{ir}} = S_r(W)_{ii} \quad (29)$$

$S_r(W)_{ii}$  expresses the impact on the dependent variable observation  $i$  from a change in  $x_{ir}$  as a combination of direct and indirect (neighborhood) influences. These spatial spillovers arise as a result of impacts passing through neighboring regions and back to the region itself. The magnitude of this type of feedback will depend upon : (1) the position of the region in space (or in general in the connectivity structure), (2) the degree of connectivity among regions governed by the weight matrix  $W$  used in the model, (3) the parameter  $\rho$  measuring the strength of spatial dependence, and (4) the magnitude of the coefficient estimates for  $\beta$  and  $\theta$ .

Since the impact of changes in an explanatory variable differs over all regions, it seems desirable to find a summary measure of these varying impacts. Pace and LeSage (2006) set forth the following scalar summary measures that can be used to average these impacts across all institutions.

The **Average Direct effect** – averaged over all  $n$  regions/observations providing a summary measure of the impact arising from changes in the  $i$ th observation of variable  $r$ . For example, if region  $i$  increases the number of commuters who use public transportation, what will be the average impact on the commuting times in region  $i$  ? This measure will take into account feedback effects that arise from the change in the  $i$ th region's public transportation usage on commuting times of neighboring regions in the system of spatially dependent regions.

The **Average Total effect = Average Direct effect + Average Indirect effect**. This scalar summary measure has two interpretations. Interpretation 1), if all regions raise public transportation usage, what will be the average total impact on commuting times of the typical region ? This total effect will include both the **average direct impact** plus the **average indirect impact**. Interpretation 2) measures the total cumulative impact arising from one region  $j$  raising its public transportation usage on commuting times of all other regions (on average).

(Pace and LeSage, 2006) show that the numerical magnitudes arising from calculation of the average total effect summary measure using interpretation 1) or 2) are equal. They argue this is a feature of spatial lag regression models that has been ignored by practitioners using these models.

Finally, the **Average Indirect effect = Average Total effect - Average Direct effect** by definition. As an example, this effect could be used to measure the impact of all other regions raising their public transportation usage on the commuting times of an individual region, again averaged over all regions.

We note that these summary measures of the impacts arising from changes in the explanatory variables of the model average over all institutions or observations in the sample, as is typical of regression model interpretations of the parameters  $\hat{\beta}_r$ . Of course, one could examine impacts for an individual region  $i$  arising from changes in explanatory variables of region  $i$ , or a neighboring region  $j$  without averaging. For example, Hondroyannis, Kelejian, Tavlas (2006) examine the impact of financial contagion arising from a single country on other countries in the model, reflecting a situation where interest is in the total impact that arises from a change in an explanatory variable  $r$  in a particular country  $i$ . Other examples include : Anselin and Le Gallo (2006) who examine diffusion of point source air pollution, and Le Gallo, Ertur and Baumont (2003), and Dall'erba and Le Gallo (2007) where the impact of changing explanatory variables (such as European Union structural funds) in strategic regions is considered on overall economic growth. In these cases the impacts take the form of an  $n$  by 1 vector, greatly complicating interpretation.

#### 4.2. Statistical significance of the impacts

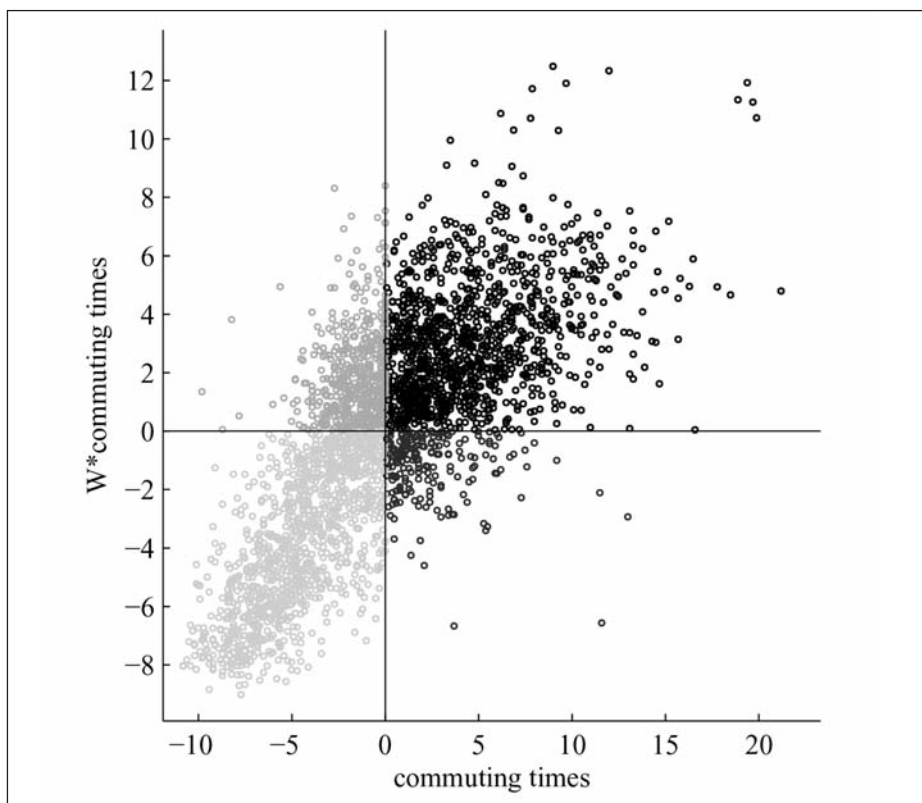
For inference regarding the significance of these impacts, we need the distribution in addition to the point estimates discussed above. There are two ways to proceed here, one involves simulating impacts based on the model estimates and expression (29). The other approach is to rely on Bayesian Markov Chain Monte Carlo estimation that provides a large number of draws for the model parameters,  $\rho$ ,  $\beta$ ,  $\theta$ ,  $\alpha$  and  $\sigma^2$ . As shown by Gelfand and Smith (1990), using these draws of the parameters to evaluate (29) will produce estimates of dispersion based on simple variance calculations applied to the results.

### V. — AN APPLIED ILLUSTRATION

To illustrate the ideas discussed we present maximum likelihood estimates from the SDM spatial regression model using logged commuting to work times (in minutes) for 3,110 US counties taken from the year 2000 census as the dependent variable. A simple way to examine the extent of spatial dependence in commuting times is a Moran scatter plot, shown in Figure 1 (see next page), with an accompanying map shown in Figure 2 (see next page).

The Moran scatter plot shows the relation between the commuting times dependent variable vector  $y$  (in deviation from means form) and the average values of neighboring observations in the spatial lag vector  $Wy$ , where we relied on the ten nearest neighboring counties to form the matrix  $W$ . The moti-

FIGURE 1 : Moran scatter plot of commuting times



variation for using ten nearest neighbors will become clear when we compare models based on different spatial weight matrices. By virtue of the transformation to deviation from means, we have four Cartesian quadrants in the scatter plot centered on zero values for the horizontal and vertical axes. These four quadrants reflect :

Quadrant I (upper right) counties that have commuting times above the mean, where the average of neighboring counties commuting times are also above the mean,

Quadrant II (lower right) counties that are below the mean commuting time, but the average of neighboring county commuting times is above the mean,

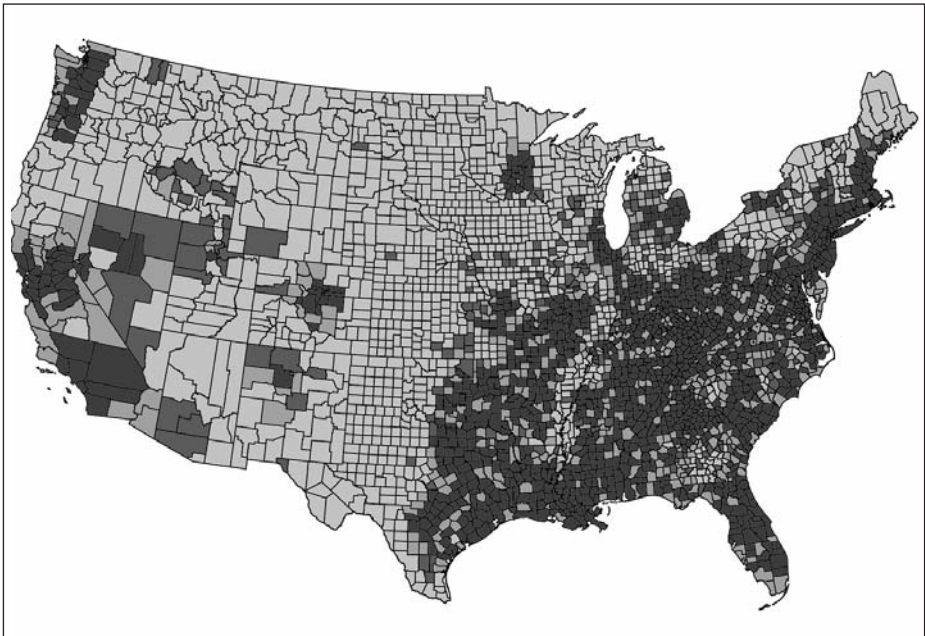
Quadrant III (lower left) counties that are below the mean commuting time, where the average of neighbors is also below the mean,

Quadrant IV (upper left) counties that have commuting times above the mean, and the average of neighboring county commuting times is below the mean.

From the scatter plot, we see a positive association between points  $y$  associated with the horizontal axis and points  $Wy$  from the vertical axis, suggesting positive spatial dependence in county-level commuting times. In fact, the magnitude of the slope from a line fitted through the points in the Moran scatter plot would equal Moran's  $I$ -statistic often used to formally test for spatial dependence. (The alternative hypothesis in this test is that the slope equals zero, indicating no spatial dependence). Another way to consider the strength of positive association is to note that there are very few Quadrant II and Quadrant IV in the scatter plot. Quadrant II represent observations from the vector  $y$  below the mean and those from  $Wy$  above the mean. The converse is true of the Quadrant IV, that is observations from  $y$  are above the mean and those from  $Wy$  are below the mean. A large number of points in quadrants II and IV with few points in quadrants I and III would suggest negative spatial dependence.

Points in the scatter plot can be placed on a map using the same coding scheme, as in Figure 2. Dark counties are those with higher than average commuting times where the average of neighboring county commuting times is also above the mean. The map shows a clustering of counties in the midwest that have lower than average commuting times and are surrounded by neighboring counties that also have low commuting times. The east and west coasts as well as the southeastern counties have higher than average commuting times, and

*FIGURE 2 : Moran plot map of county-level commuting times*



are surrounded by counties that also have higher than average commuting times.

As explanatory variables in our SDM regression model we use population density, a constant term, and in- and out-migration of population to the county over the 1995-2000 period. Since this is a spatial Durbin model, the explanatory variables also include the average of these variables from neighboring counties, which we label as  $W \cdot$  Population density,  $W \cdot$  in-migration,  $W \cdot$  out-migration.

Bayesian model comparison methods were used to compare models based on spatial weight matrices ranging between 1 and 15 nearest neighbors, with results for 4 to 14 neighbors shown in Table 1.

TABLE 1 : Models based on 6 to 14 nearest neighbors

# nearest neighbors	Model Probabilities	log-marginal likelihood	difference in log-marginals
6	0.0000	1200.9201	36.8444
7	0.0000	1214.5424	23.2220
8	0.0000	1227.1382	10.6262
9	0.1142	1235.9867	1.7778
10	0.6864	1237.7645	0.0000
11	0.0890	1235.7055	2.0590
12	0.1063	1235.8700	1.8945
13	0.0041	1232.5908	5.1737
14	0.0000	1227.4054	10.3591

From the table we see that a weight matrix based on ten neighbors produces the highest posterior model probability. Differences between the log-marginal likelihood function values are *log Bayes factors* for a comparison of two models. Specifically, the log-marginal for model 1 minus the log-marginal for model 2 provides evidence in favor of model 1 versus model 2. According to a scale developed by Jeffreys (1961), differences in the ranges of (0,1.15), provide « very slight » evidence in favor of model 1 versus 2, differences between (1.15,3.45), provide « slight evidence », whereas the range (3.45,4.60), represents « strong evidence » and the range (4.60,  $\infty$ ) indicates « decisive evidence » in favor of model 1 versus 2. The last column of the table shows the log Bayes factors based on a comparison of the best model based on ten nearest neighbors and other models. These provide « slight evidence » in favor of the model with ten neighbors relative to models based on 9, 11 and 12 neighbors since the log Bayes factors are in the range (1.15,3.45) for these, and «strong evidence» relative to all others, since these are in the range (3.45,4.60).

Table 2 presents least-squares and SDM model estimates based on ten nearest neighbors. Since the estimate for the parameter  $\rho$  is significantly different from zero, least-squares estimates are biased and inconsistent. From the table, it seems clear that the bias in least-squares is upward. Typically, non-spatial models tend to attribute variation in the dependent variable to the explanatory variables leading to larger (in absolute value terms) estimates, whereas the SDM model assigns this variation to the spatial lag of the dependent variable, producing smaller estimates.

As noted in section 41, the SDM model estimates cannot be interpreted as partial derivatives in the typical regression model fashion. To assess the signs and magnitudes of impacts arising from changes in the three explanatory variables, we turn to the summary measures of direct, indirect and total impacts presented in Table 3 (see next page).

From the table we see that the direct impact of increasing population density is not significant, suggesting this will not have an impact on commuting times. However, the indirect effect of increasing population density in neighboring regions is positive and significant. This suggests that increased population density in neighboring counties has a positive impact on commuting times, which seems intuitively plausible. The total effect from population density is positive and comprised mostly of the indirect impact.

In-migration also exerts a positive direct and indirect impact on commuting times, suggesting that we would see increased commuting times in counties experiencing positive in-migration. The indirect impacts from in-migration in

**TABLE 2 :**  
*Least-squares versus Spatial estimates for the commuting time model*

	SDM model		Least-squares	
	coefficient	t-statistic	coefficient	t-statistic
Intercept	0.9990	10.89	3.912	63.90
Population Density	-0.0005	-0.09	0.1080	24.11
In-migration	0.1246	11.87	0.2334	19.31
Out-migration	-0.1649	-15.15	-0.2959	-24.20
W · Population Density	0.0337	4.16	na	na
W · In-migration	-0.0096	-0.50	na	na
W · Out-migration	0.0572	2.92	na	na
$\rho$	0.6837	36.27	na	na
$\sigma^2$	0.0230		0.0431	
R-squared	0.4903		0.3530	
na = not applicable				

TABLE 3 : Effects of Changes in the Regressors on Commuting Times

	Mean	t-statistic	t-probability
Direct effects			
Population density	0.0031	0.4923	0.6225
In-migration	0.1331	12.6698	0.0000
Out-migration	-0.1711	-15.7163	0.0000
Indirect effects			
Population density	0.1021	6.0220	0.0000
In-migration	0.2319	4.1527	0.0000
Out-migration	-0.1708	-2.9921	0.0028
Total effects			
In-migration	0.1052	6.5284	0.0000
Out-migration	0.3650	6.3123	0.0000
Total effects	-0.3420	-5.7814	0.0000

nearby counties is nearly twice the magnitude of the direct impact, suggesting a large spillover impact from in-migration. The total impact is positive, with about two-third of this comprised of the spillover effects from in-migration in neighboring counties.

Finally, the direct and indirect impacts of out-migration are negative, as we would expect. These two impacts are equal in magnitude, leading to a total impact that consists of equal parts direct and indirect impact. The net direct impact of migration (in-migration minus out-migration) is negative, since the negative direct impact from out-migration exceeds that of the positive impact from in-migration. In contrast, the net indirect impact of migration is positive, since the positive indirect impact from in-migration is greater than the negative indirect impact from out-migration. The net total impact of migration is small but positive ( $0.3650 - 0.3420 = 0.023$ ).

## VI. — CONCLUSIONS

Spatial autoregressive processes represent a parsimonious way to model structured dependence between observations that often arise in economic research. One type of structured dependence is spatial dependence, but other examples include situations where actions of one economic agent depend on that of others. For example, Blankmeyer *et al.* (2007) model dependence between CEO compensation using these models in a structured dependence set-

ting. Links to work in the area of social networking were also made in our presentation.

For the case of spatial dependence considered here, we show how basic regression models can be augmented with spatial autoregressive processes to produce models that incorporate simultaneous feedback between regions located in space. Least-square regressions that ignore this feedback result in biased and inconsistent estimates.

Interpretation of estimates and inferences regarding the relationships modeled require a steady-state view, where changes in the explanatory variables lead to a series of simultaneous feedbacks that produce a new steady-state equilibrium. Because we are working with cross-sectional sample data, these model adjustments appear as if they are simultaneous, but we demonstrated that an implicit time dimension exists in these models.

The availability of public domain software to implement estimation and inference for the models described here should make these methods widely accessible (Anselin, 2006 ; Bivand, 2002 ; LeSage, 1999 ; Pace, 2003).

A number of problems are likely to arise when attempting to extend the spatial regression methods described here to the case of non-spatial structured dependence that would be more appropriate for sample data involving individual firms. These issues outlined below require further research.

One controversial issue would be criterion used to quantify the dependence structure between firms. Related to this, the observations identified as « neighbors or peer firms » and the associated dependence structure must be exogenous with respect to the explanatory variables in the model. This means that any variables used to produce the « generalized distance » matrix cannot be included as explanatory variables in the regression equation itself. For example, inclusion of a size variable used to define neighboring firms and the associated dependence structure as an explanatory variable in the regression would lead to interpretative complications when considering the partial derivative impact of changes in this variable. Any change in the size variable would have two impacts on the model, one associated with the change in this explanatory variable, and another arising from the implied change in the dependence structure. These issues do not arise in spatial regression modeling where Euclidean distance between observations/regions is the accepted criterion for defining neighbors, and it is relatively easy to exclude this variable from the explanatory variable set. Further work needs to be done in this area.

Another potential issue for structured versus spatial dependence is that the bounds on the scalar dependence parameter  $\rho$  are determined by the minimum and maximum eigenvalues of the matrix  $W$  in the model. For spatial models, a symmetric matrix  $W$  (or a *similar matrix* having the same eigenvalues) is typically employed. In the case of structured dependence between firms, asymmetric relations are likely to be more realistic, and an asymmetric weight matrix



can have complex eigenvalues. Blankmeyer *et al.* (2007) suggest use of an overly restrictive bound  $-1 < \rho < 1$  as a sufficient condition to ensure a positive-definite variance covariance matrix in these models.

Working with samples of individual firms also leads to issues regarding sample selection bias and would tend to increase heterogeneity and perhaps outliers. Note that we have assumed homoscedastic disturbance variances in our spatial regression model. For a Bayesian generalization that accommodates non-constant variance and downweights outliers see LeSage (1997). The heterogeneity may also lead to variation in the parameters of the relationship being explored, an issue explored in a spatial setting by Ertur, Le Gallo and LeSage (2007).

For the case of analyzing binary decisions made by firms in a structured dependence setting, each firm's decision may depend on those of « neighboring » firms. Since this violates the usual independence assumption, use of conventional logit and probit models is inappropriate. Spatial regression models have been extended for these situations by LeSage (2000), Smith and LeSage (2004) and for the case of multinomial probit see Autant-Bernard, LeSage and Parent (2007).

Summarizing, there is a great deal of potential for use of spatial regression models in modeling interdependence between cross-sectional and panel data samples of firms. However, extending the spatial methods requires caution to avoid potential pitfalls that have not been explored in the spatial econometrics literature. Nonetheless, the expanding literature on spatial econometrics should provide a good starting point for those interested in use of these methods in non-spatial settings.

## REFERENCES

- ANSELIN L. (1988), « Spatial Econometrics: Methods and Models », Dordrecht: Kluwer Academic Publishers.
- ANSELIN L. (2006), « GeoDa(tm) 0.9 User's Guide », at [geoda.uiuc.edu](http://geoda.uiuc.edu).
- ANSELIN L. and J. LE GALLO (2006). « Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects », *Spatial Economic Analysis*, 1:1, 31-52.
- AUTANT-BERNARD C., J.-P. LESAGE and O. PARENT (2007), « Firm innovation strategies: a spatial multinomial probit approach », forthcoming in *Annals of Economics and Statistics*.
- BALLESTER C., CALVÓ-ARMENGOL A. and Y. ZENOU (2006), « Who's who in networks. Wanted: The key player », *Econometrica* 74, 1403-1417.
- BARRY R. and R.-K. PACE (1999), « A Monte Carlo Estimator of the Log Determinant of Large Sparse Matrices », *Linear Algebra and its Applications*, 289, 41-54.
- BIVAND R. (2002), « Spatial econometrics functions in R: Classes and methods », *Journal of Geographical Systems*, 4:4, 405-421.
- BLACK W.-R. (1992), « Network autocorrelation in transport network and flow systems », *Geographical Analysis*, 24:3, 207-222.
- BLANKMEYER E., J.-P. LESAGE, J. STUTZMAN, K.-J. KNOX and R.-K. PACE (2007), « Statistical Modeling of Structured Peer Group Dependence Arising from Salary Benchmarking Practices », Available at SSRN: [ssrn.com/abstract=1020458](http://ssrn.com/abstract=1020458).
- BONACICH P.-B. (1987), « Power and centrality: a family of measures », *American Journal of Sociology*, 92, 1170-1182.
- DALL'ERBA S. and J. LE GALLO (2008), « Regional Convergence and the Impact of European Structural Funds over 1989-1999: A Spatial Econometric Analysis », *Papers in Regional Science* 87:2, 219-244.
- DEBREU G. and I.-N. HERSTEIN (1953), « Nonnegative Square Matrices », *Econometrica* 21, 597-607.
- ELHORST J.-P. (2003), « Specification and estimation of spatial panel data models », *International Regional Science Review* 26, 244-268.
- ERTUR C. and W. KOCH (2007), « Convergence, human capital and international spillovers », *Journal of Applied Econometrics*, 22:6 1033-1062.
- ERTUR C., J. LE GALLO and J.-P. LESAGE (2007), « Local versus Global Convergence in Europe: A Bayesian Spatial Econometric Approach », *Review of Regional Studies* 37:1, 82-108.
- FLORAX R.-J.-M. and H. FOLMER (1992), « Specification and estimation of spatial linear regression models: Monte Carlo evaluation of pre-test estimators », *Regional Science and Urban Economics*, 22, 405-432.
- FLORAX R.-J.-M., H. FOLMER and S.-J. REY (2003), « Specification searches in spatial econometrics: the relevance of Hendry's methodology », *Regional Science and Urban Economics*, 33, 557-579.
- GELFAND A.-E. and A.-F.-M. SMITH (1990), « Sampling-Based Approaches to Calculating Marginal Densities », *Journal of the American Statistical Association*, 85, 398-409.
- GRIFFITH D. (2003), « Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization », Berlin : Springer-Verlag.
- KELEJIAN H.-H., G.-S. TAVLAS and G. HONDRONYIANNIS (2006), « A Spatial Modeling Approach to Contagion Among Emerging Economies », *Open Economies Review*, 17:4/5, 423-442.
- KATZ L. (1953), « A new status index derived from sociometric analysis », *Psychometrika* 18, 39-43.
- KAZUHIKO K., W. POLASEK and H. WAGO (2007), « Model Choice for Panel Spatial Models: Crime Modeling in Japan », Advances in Data Analysis, Proceedings of the 30<sup>th</sup> Annual Conference of the Gesellschaft für Klassifikation e.V., Freie Universität Berlin, March 8-10, 2006, Reinhold Decker and Hans -J. Lenz (eds.), Springer Berlin Heidelberg, 237-244.

- KELEJIAN H. and I.-R. PRUCHA (1998), « A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances », *Journal of Real Estate and Finance Economics*, 17:1, 99-121.
- KELEJIAN H. and I.-R. PRUCHA (1999), « A generalized moments estimator for the autoregressive parameter in a spatial model », *International Economic Review* 40, 509-33.
- KOOP G. (2003), « Bayesian Econometrics », John Wiley & Sons, Ltd. (West Sussex : England).
- LACOMBE D. (2004), « Does Econometric Methodology Matter ? An Analysis of Public Policy Using Spatial Econometric Techniques », *Geographical Analysis*, 36, 87-89.
- LE GALLO J., ERTUR C., BAUMONT C. (2003), « A spatial econometric analysis of convergence across European regions, 1980-1995 », in Fingleton B. (ed.), « European Regional Growth », Springer-Verlag, Berlin.
- LESAGE J.-P. (1997), « Bayesian Estimation of Spatial Autoregressive Models », *International Regional Science Review*, 20, 113-129.
- LESAGE J.-P. (1999), « Spatial Econometrics using MATLAB » a manual for the spatial econometrics toolbox functions available at [spatial-econometrics.com](http://spatial-econometrics.com).
- LESAGE J.-P. (2000), « Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models », *Geographical Analysis*, 32:1, 19-35.
- LESAGE J.-P. (2005), « Spatial Econometrics », in « The Encyclopedia of Social Measurement », volume 3, edited by Kimberly Kempf-Leonard. Amsterdam, Netherlands: Elsevier: 613-619.
- LESAGE J.-P. M.-M. FISCHER and T. SCHERNGELL (2007), « Knowledge Spillovers across Europe, Evidence from a Poisson Spatial Interaction Model with Spatial Effects », *Papers in Regional Science*, 86:3, 393-421.
- LESAGE J.-P. and R.-K. PACE (2004a), « Introduction », « Advances in Econometrics : Volume 18: Spatial and Spatiotemporal Econometrics », (Oxford : Elsevier Ltd), J.-P. LeSage and R.-K. Pace (eds.), 1-32.
- LESAGE J.-P. and R.-K. PACE (2004b), « Using Matrix Exponentials to Estimate Spatial Probit/Tobit Models », in « Recent Advances in Spatial Econometrics », J. Mur, H. Zoller, and A. Getis (eds.) Palgrave Publishers, 105-131.
- LESAGE J.-P. and R.-K. PACE (2007), « A Matrix Exponential Spatial Specification », *Journal of Econometrics*, 140:1, 190-214.
- LESAGE J.-P. and O. PARENT (2007), « Bayesian Model Averaging for Spatial Econometric Models », *Geographical Analysis*, 39:3, 241-267.
- MARSH T.-L. and R.-C. MITTELHAMMER (2004), « Generalized Maximum Entropy Estimation Of A First Order Spatial Autoregressive Model », « Advances in Econometrics : Volume 18: Spatial and Spatiotemporal Econometrics », (Oxford: Elsevier Ltd), J.-P. LeSage and R.-K. Pace (eds.), 203-238.
- ORD J.-K. (1975), « Estimation Methods for Models of Spatial Interaction », *Journal of the American Statistical Association*, 70, 120-126.
- PACE R.-K. (2003), « Matlab Spatial Statistics Toolbox 2.0 » available at [spatial-statistics.com](http://spatial-statistics.com).
- PACE R.-K. and R. BARRY (1997), « Quick computation of spatial autoregressive estimators », *Geographical Analysis*, 29, 232-246.
- PACE R.-K. and J.-P. LESAGE (2006), « Interpreting Spatial Econometric Models », paper presented at the Regional Science Association International North American meetings, Toronto, CA.
- SONIS M., HEWINGS G.-J.-D. and OKUYAMA Y. (2001), « Feedback loops analysis of Japanese interregional trade, 1980-85-90 », *Journal of Economic Geography*, 1, 341-362.
- SMITH T.-E. and J.-P. LESAGE (2004), « A Bayesian Probit Model with Spatial Dependencies », in « Advances in Econometrics: Volume 18: Spatial and Spatiotemporal Econometrics », (Oxford: Elsevier Ltd), J.-P. LeSage and R.-K. Pace (eds.), 127-160.