Spatial Economic Analysis, Vol. 5, No. 1, March 2010

Routledge
Taylor & Francis Group

Applied Spatial Econometrics: Raising the Bar

J. PAUL ELHORST

(Received December 2009; accepted December 2009)

ABSTRACT *This paper places the key issues and implications of the new 'introductory' book on spatial econometrics by James LeSage & Kelley Pace (2009) in a broader perspective: the argument in favour of the spatial Durbin model, the use of indirect effects as a more valid basis for testing whether spatial spillovers are significant, the use of Bayesian posterior model probabilities to determine which spatial weights matrix best describes the data, and the book's contribution to the literature on spatio-temporal models. The main conclusion is that the state of the art of applied spatial econometrics has taken a step change with the publication of this book.*

Relever le niveau de l'économetrie spatial appliquée

RÉSUMÉ *La présente communication place les principales questions et implications du nouvel ouvrage d'introduction sur l'économétries spatiale de James LeSage & Kelley Pace (2009) dans un contexte plus général: l'argument favorisant le modèle spatial de Durbin, l'emploi d'effets indirects comme base plus valable pour évaluer l'aspect significatif des déversements spatiaux, l'emploi des probabilités d'un modèle baysien postérieur pour évaluer laquelle des matrices de poids spatiaux décrit le mieux les donnes, et la contribution de l'ouvrage la documentation sur les modèles spatio-temporels. La principale conclusion est qu'avec la publication de cet ouvrage, l'état de l'art de l'économétries spatiale applique a effectué un grand pas en avant.*

Alzar el nivel de la econometría espacial aplicada

RÉSUMÉ *Este trabajo plantea las cuestiones e implicaciones clave del nuevo libro introductorio sobre económetra espacial de James LeSage & Kelley Pace (2009) dentro de una perspectiva más amplia: el argumento a favor del modelo espacial Durbin, el uso de efectos indirectos como una base más válida para poner a prueba si los desbordamientos espaciales son significativos, el uso de probabilidades posteriores bayesianas para descubrir que matriz de pesos espaciales describe mejor los datos, y la contribución del libro a la bibliógrafa sobre modelos espaciotemporales. La principal conclusión es que la econometría espacial aplicada más avanzada ha experimentado un cambio radical con la publicación de este libro.*

Faculty of Economics and Business, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands. Email: j.p.elhorst@rug.nl. The author would like to thank Jan Jacobs and Jan Oosterhaven and the editors of this journal Bernard Fingleton and Harry Garretsen for valuable comments on a previous version of this paper.

ISSN 1742-1772 print; 1742-1780 online/10/010009-20 DOI: 10.1080/17421770903541772

应用空间计量经济学：门槛提高

摘要：本文通过一个更宽广的视角讨论James LeSage和Kelley Pace（2009）所著“入门”新书的关键问题和意义：支持空间德宾模型、将间接影响作为检测空间溢出是否重要的更有效的基础、用贝叶斯后验模型概率确定哪个空间加权矩阵最符合数据，以及该书对时空模型文献的贡献。主要结论是，随着这本书的出版，应用空间计量经济学的发展水平产生的突变。

KEYWORDS: *Spatial Durbin model; spatial spillovers; posterior model probabilities; spatio-temporal models*
JEL CLASSIFICATION: C21*;* R10

1. Introduction

The year 2007 marks a sea change in spatial econometricians' way of thinking. Prior to this they were interested mainly in models containing one type of spatial interaction effect: the spatial lag model and the spatial error model. The first model contains a spatially lagged dependent variable, while the second model incorporates a spatial autoregressive process in the error term. The seminal book by Anselin (1988) and the testing procedure for a spatial error or a spatial lag model based on robust Lagrange multiplier tests developed by Anselin *et al.* (1996) may be considered as the main pillars behind this way of thinking. After 2007 the interest in models containing more than one spatial interaction effect increased. In his keynote speech at the first World Conference of the Spatial Econometrics Association in 2007, Harry Kelejian advocated models that include both a spatially lagged dependent variable and a spatially autocorrelated error term (based on Kelejian & Prucha, 1998 and related work), while James LeSage, in his presidential address at the 54th North American Meeting of the Regional Science Association International in 2007, advocated models that include both a spatially lagged dependent variable and spatially lagged explanatory variables. In analogy to Durbin (1960) for the time series case, Anselin (1988) labelled the latter model the spatial Durbin model.

The argument in favour of the spatial Durbin model is now laid down in a new 'introductory' book on spatial econometrics by James LeSage & Kelley Pace (2009), and may be considered a landmark in raising the bar in the field of applied spatial econometrics. One strength of the spatial Durbin model is that it produces unbiased coefficient estimates also if the true data-generation process is a spatial lag or a spatial error model. Another strength is that it does not impose prior restrictions on the magnitude of potential spatial spillover effects. In contrast to other spatial regression specifications, these spillover effects can be global or local and be different for different explanatory variables. These and other important issues put forward in LeSage & Pace's book (hereinafter with page numbers shown in parentheses) will be summarized in this paper. This paper, however, is more than just a book review. In each of the following sections, I will first give my own view of the state of the art of one theme in applied spatial econometrics, and then I will discuss the contribution of LeSage & Pace's book.

2. A Taxonomy of Linear Spatial Dependence Models for Cross-section Data

To give a full explanation of the claim that the spatial Durbin model produces unbiased coefficient estimates, also if the true data-generation process is a spatial lag or a spatial error model, I first consider a taxonomy of linear spatial dependence models for cross-section data.

The standard approach in most empirical work is to start with a non-spatial linear regression model and then to test whether or not the model needs to be extended with spatial interaction effects. This approach is known as the specific-to-general approach. The non-spatial linear regression model takes the form

$$Y = \alpha \iota_N + X\beta + \varepsilon, \tag{1}$$

where Y denotes an $N \times 1$ vector consisting of one observation on the dependent variable for every unit in the sample ($i = 1, \ldots, N$), ι_N is an $N \times 1$ vector of ones associated with the constant term parameter α, X denotes an $N \times K$ matrix of exogenous explanatory variables, with the associated parameters β contained in a $K \times 1$ vector, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)^T$ is a vector of disturbance terms,[1] where ε_i are independently and identically distributed error terms for all i with zero mean and variance σ^2. Since the linear regression model is commonly estimated by ordinary least squares (OLS), it is often labelled the OLS model. Furthermore, even though the OLS model in most studies focusing on spatial interaction effects is rejected in favour of a more general model, its results often serve as a benchmark.

The opposite approach is to start with a more general model containing, nested within it as special cases, a series of simpler models that ideally should represent all the alternative economic hypotheses requiring consideration. Manski (1993) points out that three different types of interaction effects may explain why an observation associated with a specific location may be dependent on observations at other locations: (i) endogenous interaction effects, where the decision of a spatial unit (or its economic decision makers) to behave in some way depends on the decision taken by other spatial units; (ii) exogenous interaction effects, where the decision of a spatial unit to behave in some way depends on independent explanatory variables of the decision taken by other spatial units—if the number of independent explanatory variables in a linear regression model is K, then the number of exogenous interaction effects is also K, provided that the intercept is considered as a separate variable; and (iii) correlated effects, where similar unobserved environmental characteristics result in similar behaviour.

The Manski model takes the form

$$Y = \rho WY + \alpha \iota_N + X\beta + WX\theta + u, \tag{2a}$$

$$u = \lambda Wu + \varepsilon, \tag{2b}$$

where the variable WY denotes the endogenous interaction effects among the dependent variables, WX the exogenous interaction effects among the independent variables, and Wu the interaction effects among the disturbance terms of the different spatial units. ρ is called the spatial autoregressive coefficient, λ the spatial autocorrelation coefficient, while θ, just as for β, represents a $K \times 1$ vector of fixed but unknown parameters.

W is an $N \times N$ matrix describing the spatial arrangement of the spatial units in the sample. Lee (2004) shows that W should be a non-negative matrix of known constants. The diagonal elements are set to zero by assumption, since no spatial unit can be viewed as its own neighbour. The matrices $I - \rho W$ and $I - \lambda W$ should be non-singular, where I represents the identity matrix of order N. For a symmetric W, this condition is satisfied as long as ρ and λ are in the interior of $(1/\omega_{\min}, 1/\omega_{\max})$, where $\omega_{\min}$ denotes the smallest (i.e. most negative) and $\omega_{\max}$ the largest real characteristic root of W. If W is row normalized subsequently, the latter interval takes the form $(1/\omega_{\min}, 1)$, since the largest characteristic root of W equals unity in this situation. If W is an asymmetric matrix before it is row normalized, it may have complex characteristic roots. LeSage & Pace (pp. 88–89) demonstrate that in that case ρ and λ are restricted to the interval $(1/r_{\min}, 1)$, where $r_{\min}$ equals the most negative purely real characteristic root of W after this matrix is row normalized. Finally, one of the following two conditions should be satisfied: (a) the row and column sums of the matrices W, $(I - \rho W)^{-1}$ and $(I - \lambda W)^{-1}$ before W is row normalized should be uniformly bounded in absolute value as N goes to infinity, or (b) the row and column sums of W before W is row normalized should not diverge to infinity at a rate equal to or faster than the rate of the sample size N. Condition (a) originates from Kelejian & Prucha (1998, 1999), and condition (b) from Lee (2004). Both conditions limit the cross-sectional correlation to a manageable degree, i.e. the correlation between two spatial units should converge to zero as the distance separating them increases to infinity.

When the spatial weights matrix is a binary contiguity matrix, (a) is satisfied. Normally, no spatial unit is assumed to be a neighbour to more than a given number, say q, of other units. By contrast, when the spatial weights matrix is an inverse distance matrix, (a) may not be satisfied. Consider an infinite number of spatial units that are arranged linearly. The distance of each spatial unit to its first left- and right-hand neighbour is d; to its second left- and right-hand neighbour, the distance is $2d$; and so on. When W is an inverse distance matrix and the off-diagonal elements of W are of the form $1/d_{ij}$, where d_{ij} is the distance between two spatial units i and j, each row sum is $2 \times (1/d + 1/2d + 1/3d + \ldots)$, representing a series that is not finite. This is perhaps the reason why some empirical applications introduce a cut-off point d^* such that $w_{ij} = 0$ if $d_{ij} > d^*$. However, since the ratio $2 \times (1/d + 1/2d + 1/3d + \ldots)/N \to 0$ as N goes to infinity, condition (b) is satisfied, which implies that an inverse distance matrix without a cut-off point does not necessarily have to be excluded in an empirical study for reasons of consistency.

The opposite situation occurs when all cross-sectional units are assumed to be neighbours of each other and are given equal weights. In that case all off-diagonal elements of the spatial weights matrix are $w_{ij} = 1$. Since the row and column sums are $N - 1$, these sums diverge to infinity as N goes to infinity. In contrast to the previous case, however, $(N - 1)/N \to 1$ instead of 0 as N goes to infinity. This implies that a spatial weights matrix that has equal weights and that is row normalized subsequently, $w_{ij} = 1/(N - 1)$, must be excluded for reasons of consistency.

Figure 1 summarizes a family of eight linear spatial econometric models, among which are the non-spatial model in (1) on the right-hand side and the Manski model in (2) on the left-hand side. Each model to the right of the Manski model can be obtained from that model by imposing restrictions on one or more of its parameters. The kinds of restrictions are reported alongside the arrows in Figure 1.

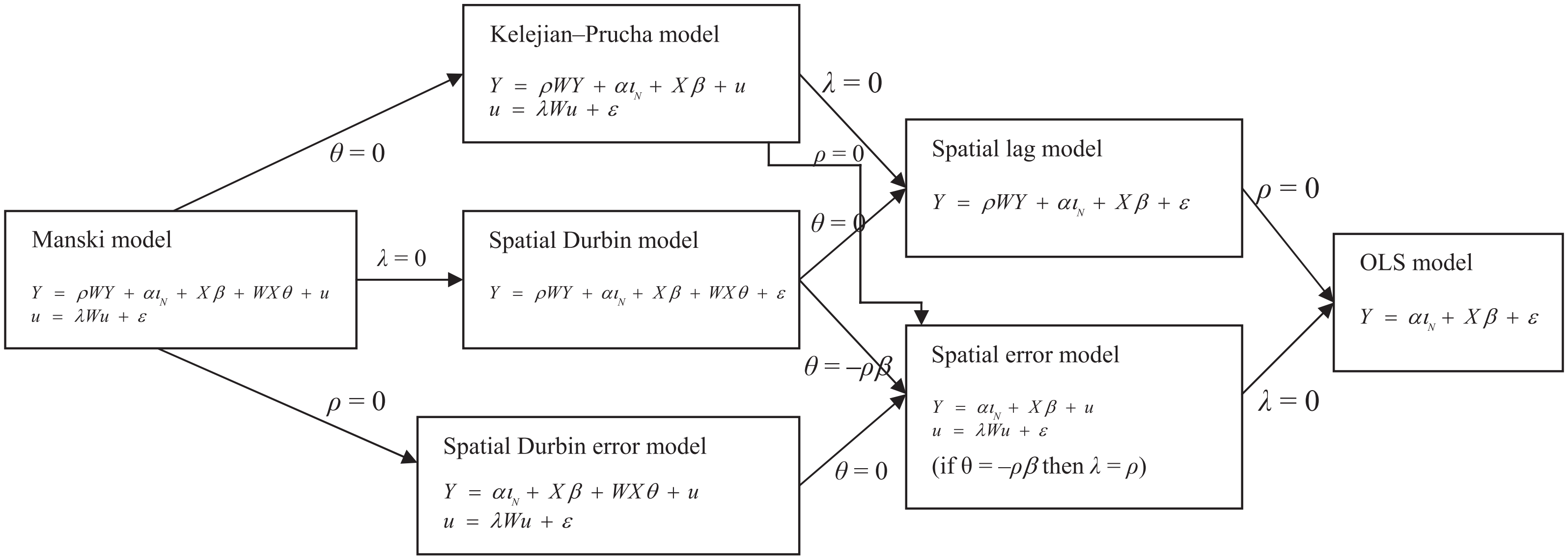

Figure 1. The relationships between different spatial dependence models for cross-section data.

Some of the models recorded in Figure 1 are well known and frequently used in applied research, while other models are not. LeSage & Pace (p. 32) denote the model with a spatially lagged dependent variable (WY) and a spatially autocorrelated error term ($W\varepsilon$) by the term SAC, though without pointing out what this acronym is standing for. Since Kelejian & Prucha (1998) have been the main advocates of this model, it is therefore renamed the Kelejian–Prucha model in this paper. The model with a spatially lagged dependent variable (WY) and spatially lagged independent variables (WX) has been introduced by Anselin (1988) and is labelled the spatial Durbin model. A model with spatially lagged independent variables (WX) and a spatially autocorrelated error term has hardly been used in the literature. LeSage & Pace (pp. 41–42) label it the spatial Durbin error model.

Figure 1 seems to indicate that the best strategy to test for spatial interaction effects is to start with the most general model, i.e. the model that includes a spatially lagged dependent variable, spatially lagged independent variables, and a spatially autocorrelated error term simultaneously. However, as Manski (1993) notes, at least one of the $K+2$ interaction effects must be excluded, because otherwise the parameters are unidentified. To verify this, I carried out a simple Monte Carlo experiment generating Y by (2a) and (2b), one X variable drawn from a uniform distribution on the interval [–1,1], $\rho = \alpha = \beta = \theta = \lambda = 0.25$, $\sigma^2 = 0.01$, $N = 60$, and a spatial weights matrix W corresponding to the corners of the seams in a soccer ball (in Matlab known as the Bucky Ball). Based on 1,000 repetitions, I found biases in the parameter estimates of the endogenous and exogenous interaction effects (ρ and θ, respectively) that may be as great as 0.0423 and, related to that, standard deviations that may be as great as 0.2677. These results corroborate Manski's (1993, p. 534) finding that there are no technical obstacles to estimating a model with interaction effects among the dependent variable, the independent variables and the disturbance terms, but that the parameter estimates cannot be interpreted in a meaningful way since the endogenous and exogenous effects cannot be distinguished from each other.

Following LeSage & Pace (pp. 155–158), the best option in such circumstances is to exclude the spatially autocorrelated error term.[2] The cost of ignoring spatial dependence in the dependent variable and/or in the independent variables is relatively high since the econometrics literature has pointed out that if one or more relevant explanatory variable are omitted from a regression equation, the estimator of the coefficients for the remaining variables is biased and inconsistent (Greene, 2005, pp. 133–134). In contrast, ignoring spatial dependence in the disturbances, if present, will only cause a loss of efficiency. Furthermore, the spatial Durbin model produces unbiased coefficient estimates also if the true data-generation process is any of the other spatial regression specifications recorded in Figure 1, except for the Manski model. By contrast, if the Kelejian–Prucha model is taken as the point of departure, it will suffer from omitted variables bias if the true data-generation process is a spatial Durbin or a spatial Durbin error model. Similarly, if the spatial Durbin error model is taken as the point of departure, it will suffer from omitted variables bias if the true data-generation process is a spatial lag, Kelejian–Prucha or spatial Durbin model.

A concomitant advantage of the spatial Durbin model is that it produces correct standard errors or t-values of the coefficient estimates also if the true data-generating process is a spatial error model. This is because the spatial error model is a special case of the spatial Durbin model (see Figure 1), as a result of which error dependence is correctly taken into account in its variance–covariance matrix. Whether inference regarding dispersion of the explanatory variables is also correct if the true data-generating process is a Kelejian–Prucha or spatial Durbin error model is an issue that, according to LeSage & Pace, needs further exploration (p. 158). This is because neither of these models is a special case of the others (see Figure 1), indicating that the implicit specification of spatial error dependence in the spatial Durbin model is different from that in the other two models.

3. Methods of Estimation

Three methods have been developed in the literature to estimate models that include spatial interaction effects. One is based on maximum likelihood (ML), one on instrumental variables or generalized method of moments (IV/GMM), and one on the Bayesian Markov Chain Monte Carlo (MCMC) approach. One important limitation of LeSage & Pace's book is that it discusses the ML method (Ch. 3) and the Bayesian method (Ch. 5) extensively, but does not pay any attention to the IV/GMM method.

One advantage of IV/GMM estimators is that they do not rely on the assumption of normality of the disturbances. One disadvantage is the possibility of ending up with a coefficient estimate for ρ or for λ outside its parameter space. Whereas these coefficients are restricted to the interval $(1/r_{min}, 1)$ by the Jacobian term in the log-likelihood function of ML estimators or in the conditional distribution of the spatial parameter of Bayesian estimators, they are unrestricted using IV/GMM since these estimators ignore the Jacobian term.

One of the reasons for developing IV/GMM estimators was as a response to perceived computational difficulties (Kelejian & Prucha, 1998, 1999). Estimation of spatial econometric models involves the manipulation of $N \times N$ matrices, such as matrix multiplication, matrix inversion, the computation of characteristic roots and/or Cholesky decomposition. These manipulations may be computationally intensive and/or may require significant amounts of memory if N is large. Since IV/GMM estimators ignore the Jacobian term, many of these problems could be avoided. In Chapter 4 of their book, however, LeSage & Pace produce conclusive evidence that these computational difficulties have become a thing of the past, as is the case for ML and Bayesian estimators.

In spite of this, Fingleton & Le Gallo (2007, 2008) show that IV/GMM estimators are extremely useful in those cases where linear spatial dependence models contain one or more endogenous explanatory variables (other than the spatially lagged dependent variable) that need to be instrumented. In applied econometrics work, the presence of endogenous variables on the right-hand side is a common occurrence, as endogeneity may be the result of measurement errors on explanatory variables, of omitted variables correlated with included explanatory variables or of the existence of an unknown set of simultaneous structural equations. ML or Bayesian estimators of models with a spatial lag (i.e. the spatial lag model and the spatial Durbin model) and additional endogenous variables do not feature in the spatial econometrics literature and would be difficult, if not impossible, to derive. The same applies to models with a spatial error process (i.e. the spatial error model and the spatial Durbin error model). By contrast, models including a spatial lag and additional endogenous variables can be straightforwardly estimated by two-stage least squares (2SLS). To instrument the spatially lagged dependent variable, Kelejian *et al.* (2004) suggest $[X_t \;\; WX_t \; \dots \; W^g X_t]$, where g is a pre-selected constant.[3] Typically, one would take $g = 1$ or $g = 2$, dependent on the number of regressors and the type of model. For example, in the case of the spatial Durbin model g should be greater than one, since this model already contains the variables X and WX on the right-hand side. If one or more of the explanatory variables are endogenous, the set of instruments must be limited to $[X_t^{ex} \;\; WX_t^{ex} \;\; \ldots \;\; W^d X_t^{ex}]$, where 'ex' denotes the X variables that are exogenous. Another difference is that this set may also be used to instrument the additional endogenous explanatory variables. A similar type of extension applies to Kelejian & Prucha's (1999) GMM estimator for models including a spatial error process together with endogenous explanatory variables (Fingleton & Le Gallo, 2007). In addition, Fingleton & Le Gallo (2008) consider a mixed 2SLS/GMM estimator of the Kelejian–Prucha model extended to include endogenous explanatory variables.

4. Model Comparison

Many practitioners are in two minds about whether to apply the general-to-specific or the specific-to-general approach. This is understandable. Whereas LeSage & Pace argue that the spatial Durbin model is the best point of departure, Florax *et al.* (2003) have found that the expansion of a linear regression equation with spatially lagged variables, conditional on the results of misspecification tests, outperforms the general-to-specific approach for finding the true data-generation process.[4] I therefore propose the following test procedure to find out which model is the most likely candidate to explain the data.[5]

First estimate the OLS model and test whether the spatial lag model or the spatial error model is more appropriate to describe the data. For this purpose, one may use the classic LM-tests proposed by Anselin (1988), and the robust LM-tests proposed by Anselin *et al.* (1996).[6] Both the classic and the robust tests are based on the residuals of the OLS model and follow a chi-squared distribution with one degree of freedom.

If the OLS model is rejected in favour of the spatial lag, the spatial error model or in favour of both models, then the spatial Durbin model should be estimated. If these models are estimated by maximum likelihood, a likelihood ratio (LR) test can subsequently be used to test the hypotheses H_0: $\theta = 0$ and H_0: $\theta + \rho\beta = 0$. The first hypothesis examines whether the spatial Durbin can be simplified to the spatial lag model, and the second hypothesis whether it can be simplified to the spatial error model (see Figure 1). Both tests follow a chi-squared distribution with K degrees of freedom.

If both hypotheses H_0: $\theta = 0$ and H_0: $\theta + \rho\beta = 0$ are rejected, then the spatial Durbin best describes the data. Conversely, if the first hypothesis cannot be rejected, then the spatial lag model best describes the data, provided that the (robust) LM tests also pointed to the spatial lag model. Similarly, if the second hypothesis cannot be rejected, then the spatial error model best describes the data, provided that the (robust) LM tests also pointed to the spatial error model. If one of these conditions is not satisfied, i.e. if the (robust) LM tests point to a model other than the LR tests, then the spatial Durbin model should be adopted. This is because this model generalizes both the spatial lag and the spatial error model (see Figure 1).

If the OLS model is estimated and not rejected in favour of both the spatial lag and the spatial error model,[7] the OLS model should be re-estimated including spatially lagged independent variables (WX) or a particular selection of these K variables to be able to test the hypothesis H_0: $\theta = 0$. If this hypothesis cannot also be rejected, then it may be concluded that the OLS model best describes the data. In that case there is no empirical evidence in favour of any type of spatial interaction effect. By contrast, when this hypothesis must be rejected, the spatial Durbin model should be estimated to be able to test the additional hypothesis H_0: $\rho = 0$. If the latter hypothesis is also rejected, then the spatial Durbin model best describes the data. If it is not, it may be concluded that a model with spatially lagged independent variables only suffices.

5. Selection of the Spatial Weights Matrix

One major weakness of spatial econometric models is that the spatial weights matrix W cannot be estimated but needs to be specified in advance and that economic theory underlying spatial econometric applications often has little to say about the specification of W (Leenders, 2002). For this reason, it has become common practice to investigate whether the results are robust to the specification of W. A recent influential paper by Ertur & Koch (2007) is illustrative of this approach. By including spatial dependence structures in theoretical economic relationships explaining economic growth per capita, they end up with an empirical model that takes the form of a spatial Durbin model. However, since the theoretical model offers no guidance as to how to specify the spatial weights matrix, they consider two alternatives: one matrix whose non-diagonal elements are measured by $1/d^2$, where d reflects the distance between two units, and another matrix whose elements are measured by e^{-2d}. Since many other specifications have been considered in the spatial econometrics literature, among which are a binary contiguity matrix, an inverse distance matrix with or without a cut-off point, a q nearest-neighbour matrix (where q is a positive integer, such as 5 or 10), a question that recurs is how the selection procedure of the spatial weights matrix might be improved.

A recent Monte-Carlo study by Stakhovych & Bijmolt (2009) demonstrates that a weights matrix selection procedure that is based on 'goodness-of-fit' criteria increases the probability of finding the true specification. The most widely used criterion is the log-likelihood function value. If a spatial interaction model is estimated based on S different spatial weights matrices and the log-likelihood function value of every model is estimated, one may select the spatial weights matrix exhibiting the highest log-likelihood function value. Harris & Kravtsova (2009) criticize this approach, because it would only find a local maximum among the competing models and not necessarily a correctly specified W (unless it is unknowingly included in the set of competing models considered). However, the Monte Carlo results found by Stakhovych & Bijmolt (2009) partly refute this critique. Although there is a serious probability of selecting the wrong spatial weights matrix if spatial dependence is weak, the consequences of this poor choice are limited because the coefficient estimates are quite close to the true ones. Conversely, although the wrong choice of a spatial weights matrix can distort the coefficient estimates severely, the probability that this really happens is small if spatial dependence is strong.

One of the merits of LeSage & Pace's book is that they offer another criterion to select models, namely the Bayesian posterior model probability. Whereas tests for significant differences between log-likelihood function values, such as the LR-test, can formally not be used if models are non-nested (i.e. based on different spatial weights matrices), Bayesian posterior model probabilities do not require nested models to carry out these comparisons (p. 162). The basic idea is to set prior probabilities equal to $1/S$, making each model equally likely a priori, to estimate each model by Bayesian methods, and then to compute posterior probabilities based on the data and the estimation results of this set of S models. Although the mathematics of this approach might deter potential users (Chs 5 and 6), my experience with this approach is positive. First, posterior model probabilities may differ widely even if the estimation results appear to be quite robust to different specifications of the spatial weights matrix. In a study I did on cross-country differences in governance (Seldadyo *et al.*, 2010), the posterior model probability of a 10 nearest-neighbour matrix appeared to be more than six times as large as that of an inverse distance matrix, more than three times as large as that of a five nearest-neighbour matrix, and more than twice as large as an inverse distance matrix with a cut-off point. Second, since Matlab routines applying Bayesian methods to the spatial lag, spatial error and spatial Durbin models are made downloadable for free on LeSage's website (www.spatial-econometrics.com), these kinds of comparisons can be carried out relatively easily. Furthermore, since LeSage also provides Matlab routines applying the Bayesian method to the spatial lag, spatial error and spatial Durbin models of limited dependent variables, similar types of selection procedures as discussed in this and the previous sections can be used for empirical problems requiring a probit or tobit approach (Ch. 10).

Another approach to capture spatial interaction effects is to extend the linear regression model by a finite number of unobserved factors that affect all units (Pesaran, 2006). Factor models are potentially powerful in that they do not require strong and unverifiable assumptions on the nature of spatial dependence. On the other hand, common factors only model interaction effects among the error terms and, often, these factors are difficult to interpret. Nevertheless, a careful elaboration of the relative merits of the multifactor approach, for example, in combination with the spatial Durbin model is an interesting topic for further research.

6. Direct, Indirect and Spatial Spillover Effects

Many empirical studies use point estimates of one or more spatial regression model specifications to test the hypothesis as to whether or not spatial spillovers exist. One of the key contributions of LeSage & Pace's book (p. 74) is the observation that this may lead to erroneous conclusions, and that a partial derivative interpretation of the impact from changes to the variables of different model specifications represents a more valid basis for testing this hypothesis. To illustrate this, they give an example of a spatially lagged independent variable WX whose coefficient is negative and insignificant (table 3.3), while its spatial spillover effect is positive and significant (table 3.4).

If the spatial Durbin model is taken as the point of departure and rewritten as

$$Y = (I - \rho W)^{-1}\alpha\iota_N + (I - \rho W)^{-1}(X\beta + WX\theta) + (I - \rho W)^{-1}\varepsilon, \qquad (3)$$

the matrix of partial derivatives of Y with respect to the kth explanatory variable of X in unit 1 up to unit N (say x_{ik} for $i = 1, \ldots, N$, respectively) is relatively easy to obtain:

$$\left[\frac{\partial Y}{\partial x_{1k}} \quad \cdot \quad \frac{\partial Y}{\partial x_{Nk}}\right] = \begin{bmatrix} \frac{\partial y_1}{\partial x_{1k}} & \cdot & \frac{\partial y_1}{\partial x_{Nk}} \\ \cdot & \cdot & \cdot \\ \frac{\partial y_N}{\partial x_{1k}} & \cdot & \frac{\partial y_N}{\partial x_{Nk}} \end{bmatrix}$$

$$= (I - \rho W)^{-1} \begin{bmatrix} \beta_k & w_{12}\theta_k & \cdot & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdot & w_{2N}\theta_k \\ \cdot & \cdot & \cdot & \cdot \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdot & \beta_k \end{bmatrix}, \tag{4}$$

where w_{ij} is the (i, j)th element of W. However, it is the unfamiliarity with or the complexity of this expression that troubles many practitioners. To provide a better understanding of the properties of the partial derivatives, I will give the simplest example possible without loss of generality, and use this example to explain LeSage & Pace's claim step by step.

6.1. Properties of Partial Derivatives

Suppose we have three spatial units that are arranged linearly: unit 1 is a neighbour of unit 2, unit 2 is a neighbour of both units 1 and 3, and unit 3 is a neighbour of unit 2.[8] Then the row-normalized spatial weights matrix W and the spatial multiplier matrix $(I - \rho W)^{-1}$ are[9]

$$W = \begin{bmatrix} 0 & 1 & 0 \\ w_{21} & 0 & w_{23} \\ 0 & 1 & 0 \end{bmatrix} \text{ and } (I - \rho W)^{-1}$$

$$= \frac{1}{1-\rho^2} \begin{bmatrix} 1 - w_{23}\rho^2 & \rho & \rho^2 w_{23} \\ \rho w_{21} & 1 & \rho w_{23} \\ \rho^2 w_{21} & \rho & 1 - w_{21}\rho^2 \end{bmatrix}, \tag{5}$$

where $w_{12} = w_{32} = 1$ since units 1 and 3 have only one neighbour, and $w_{21} + w_{23} = 1$ (w_{21} and w_{23} might be different). Substituting the second matrix of (5) into (4), the impact of a change in successively units 1, 2 and 3 of variable x_k on units 1, 2 and 3 of the dependent variable turns out to be

$$\left[\frac{\partial Y}{\partial x_{1k}} \quad \frac{\partial Y}{\partial x_{2k}} \quad \frac{\partial Y}{\partial x_{3k}}\right] =$$

$$\frac{1}{1-\rho^2} \begin{bmatrix} (1 - w_{23}\rho^2)\beta_k + (w_{21}\rho)\theta_k & \rho\beta_k + \theta_k & (w_{23}\rho^2)\beta_k + (\rho w_{23})\theta_k \\ (w_{21}\rho)\beta_k + w_{21}\theta_k & \beta_k + \rho\theta_k & (w_{23}\rho)\beta_k + w_{23}\theta_k \\ (w_{21}\rho^2)\beta_k + (w_{21}\rho)\theta_k & \rho\beta_k + \theta_k & (1 - w_{21}\rho^2)\beta_k + (w_{23}\rho)\theta_k \end{bmatrix}. \tag{6}$$

The results obtained in (5) and (6) illustrate that the partial derivatives of Y with respect to the kth explanatory variable in the spatial Durbin model in (4) have three important properties. First, if a particular explanatory variable in a particular unit changes, not only will the dependent variable in that unit itself change but also the dependent variables in other units. The first is called a *direct effect* and the second an *indirect effect*. Note that every diagonal element of the matrix of partial derivatives represents a direct effect, and that every non-diagonal element represents an indirect effect. Consequently, indirect effects do not occur if both $\rho = 0$ and $\theta_k = 0$, since all non-diagonal elements will then be zero [see (4) and (6)].

Second, direct and indirect effects are different for different units in the sample. Direct effects are different because the diagonal elements of the matrix $(I - \rho W)^{-1}$ are different for different units, provided that $\rho \neq 0$ [see the diagonal elements of (5) and (6)]. Indirect effects are different because both the non-diagonal elements of the matrix $(I - \rho W)^{-1}$ and of the matrix W are different for different units, provided that $\rho \neq 0$ and/or $\theta_k \neq 0$ [see the non-diagonal elements of (5) and (6)].

Third, indirect effects that occur if $\theta_k \neq 0$ are known as *local effects*, as opposed to indirect effects that occur if $\rho \neq 0$ and that are known as *global effects*. Local effects got their name because they arise only from a unit's neighbourhood set; if the element w_{ij} of the spatial weights matrix is non-zero (zero), then the effect of x_{jk} on y_i is also non-zero (zero). Global effects got their name because they also arise from units that do not belong to a unit's neighbourhood set. This follows from the fact that the matrix $(I - \rho W)^{-1}$, in contrast to W, does not contain zero elements (provided that $\rho \neq 0$) [see W and $(I - \rho W)^{-1}$ in (5)].

6.2. Implications for Reporting Direct and Indirect Effects

Since both the direct and indirect effects are different for different units in the sample, the presentation of these effects is a problem. If we have N spatial units and K explanatory variables, we obtain K different $N \times N$ matrices of direct and indirect effects. Even for small values of N and K, it may already be rather difficult to report these results compactly. To improve the surveyability of the estimation results of spatial regression model specifications, LeSage & Pace therefore propose to report one direct effect measured by the average of the diagonal elements of the matrix on the right-hand side of (4), and one indirect effect measured by the average of either the row sums or the column sums of the non-diagonal elements of that matrix. The average row effect quantifies the impact on a particular element of the dependent variable as a result of a unit change in all elements of an exogenous variable, while the average column effect quantifies the impact of changing a particular element of an exogenous variable on the dependent variable of all other units. However, since the numerical magnitudes of these two calculations of the indirect effect are the same, it does not matter which one is used (pp. 33–42).

Table 1 reports the direct and indirect effects of the models that have been recorded in Figure 1 for the example with $N = 3$ and the spatial weights matrix in (5). In the case of the spatial Durbin model [using (6)], we obtain a direct effect of

$$\frac{(3 - \rho^2)}{3(1 - \rho^2)}\beta_k + \frac{2\rho}{3(1 - \rho^2)}\theta_k$$

and an indirect effect of

$$\frac{3\rho + \rho^2}{3(1 - \rho^2)}\beta_k + \frac{3 + \rho}{3(1 - \rho^2)}\theta_k.$$

Unfortunately, since every empirical application will have its own unique number of observations (N) and spatial weights matrix (W), these formulae cannot be

Table 1. Direct and indirect effects of different model specifications

Type of model	Direct effect	Indirect effect
Spatial Durbin model/ Manski model	$\frac{(3-\rho^2)}{3(1-\rho^2)}\beta_k + \frac{2\rho}{3(1-\rho^2)}\theta_k$	$\frac{3\rho+\rho^2}{3(1-\rho^2)}\beta_k + \frac{3+\rho}{3(1-\rho^2)}\theta_k$
Spatial lag model/ Kelejian–Prucha model	$\frac{(3-\rho^2)}{3(1-\rho^2)}\beta_k$	$\frac{3\rho+\rho^2}{3(1-\rho^2)}\beta_k$
Spatial Durbin error model	β_k	θ_k
OLS model/ Spatial error model	β_k	0

Note: $N = 3$, W as in (4).

generalized. Nevertheless, the results in Table 1 do illustrate that the direct and indirect effects of different model specifications have the general properties as described below.

6.3. Properties of the Proposed Direct and Indirect Effects

If the OLS model is adopted, the direct effect of an explanatory variable is equal to the coefficient estimate of that variable (β_k), while its indirect effect is zero by construction. If the OLS model is augmented with a spatially autocorrelated error term, the direct and the indirect effects remain the same. This is because the spatial autoregressive model for the disturbances does not come into play when considering the partial derivative of the dependent variable with respect to changes in the explanatory variables [see (3) and (4)]. These properties also hold for the extension of the spatial lag model and of the spatial Durbin model with spatial autocorrelation, i.e. the Kelejian–Prucha and the Manski models, respectively.

If the spatial Durbin error model is adopted, the direct effect of an explanatory variable is equal to the coefficient estimate of that variable (β_k), while its indirect effect is equal to the coefficient estimate of its spatial lagged value (θ_k). This means that the interpretation of the coefficients of the OLS model, the spatial error model and the spatial Durbin error model is straightforward and that a comparison of these coefficients is valid.

The interpretation of the coefficients gets complicated when moving to one of the other models. If the direct effect and the indirect effect of the spatial lag model (as well as the Kelejian–Prucha model) are compared with those of the OLS model, two notable differences can be observed. First, whereas the direct effect of the *k*th explanatory variable in the OLS model is β_k, the direct effect in the spatial lag model is β_k pre-multiplied with a number that will be greater than or equal to unity. Table 1 shows that this number in the example amounts to $(3-\rho^2)/3(1-\rho^2)$. Second, whereas the indirect effect in the OLS model is zero by construction, it is non-zero in the spatial lag model.

The property that the number with which β_k is pre-multiplied when calculating the direct effect is greater than or equal to unity is due to the spatial multiplier matrix $(I-\rho W)^{-1}$ that can be decomposed as follows (pp. 40–41):

$$(I-\rho W)^{-1} = \left(\sum_{q=0}^{\infty} \rho^q W^q\right) = I + \rho W + \rho^2 W^2 + \ \ldots \tag{7}$$

Since the non-diagonal elements of the first matrix term on the right-hand side (the identity matrix I) are zero, this term represents a direct effect of a change in X only. Conversely, since the diagonal elements of the second matrix term on the right-hand side (ρW) were assumed to be zero (see Section 2), this term represents an indirect effect of a change in X only. Furthermore, since W is taken to the power 1 here, this indirect effect is limited to first-order neighbours only, i.e. the units that belong to the neighbourhood set of every spatial unit. All other terms on the right-hand side represent second- and higher-order direct and indirect effects. Higher-order direct effects arise as a result of feedback effects, i.e. impacts passing through neighbouring units and back to the unit itself (e.g. $1 \rightarrow 2 \rightarrow 1$ and $1 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1$). It is these feedback effects that are responsible for the fact that the overall direct effect eventually increases.[10]

One important limitation of the spatial lag model is that the ratio between the indirect and the direct effect of a particular explanatory variable is independent of β_k. Table 1 illustrates that β_k in the numerator and β_k in the denominator of this ratio cancel each other out:

$$\frac{(3-\rho^2)}{3(1-\rho^2)}\beta_k \bigg/ \frac{3\rho+\rho^2}{3(1-\rho^2)}\beta_k = \frac{3-\rho^2}{3\rho+\rho^2}.$$

This property implies that the ratio between the indirect and direct effects in the spatial lag model is the same for every explanatory variable, and that its magnitude depends on the spatial autoregressive parameter ρ and the specification of the spatial weights matrix W only. In many empirical applications, this is not very likely.

Finally, if the spatial Durbin model is adopted, both the direct effect and the indirect effect of a particular explanatory variable will also depend on the coefficient estimate θ_k of the spatially lagged value of that variable (see Table 1). The result is that no prior restrictions are imposed on the magnitude of both the direct and indirect effects and thus that the ratio between the indirect and the direct effect may be different for different explanatory variables. As a result of this property, the spatial Durbin is a more attractive point of departure in an empirical study than other spatial regression specifications.

6.4. Testing for Spatial Spillovers

The estimated indirect effects of the independent explanatory variables should eventually be used to test the hypothesis as to whether or not spatial spillovers exist, rather than the coefficient estimate of the spatially lagged dependent variable and/or the coefficient estimates of the spatially lagged independent variables. However, one difficulty is that it cannot be seen from the coefficient estimates and the corresponding standard errors or t-values (derived from the variance–covariance matrix) whether the indirect effects in the spatial Durbin model are significant (note: the same applies to the spatial lag model). This is because the indirect effects are composed of different coefficient estimates according to complex mathematical formulae and the dispersion of these indirect effects depends on the dispersion of all coefficient estimates involved (see Table 1). For example, if the coefficients ρ, β_k and θ_k in the spatial Durbin model happen to be significant, this does not automatically mean that the indirect effect of the kth explanatory variable is also significant. Conversely, if one or two of these coefficients are insignificant, the indirect effect may still be significant.

One possible way to calculate the dispersion of the direct and indirect effects is to apply formulae for the sum, the difference, the product and the quotient of random variables (see, among others, Mood *et al.*, 1974, pp. 178–181). However, owing to the complexity of the matrix of partial derivatives [see (6)] and because every empirical application will have its own unique number of observations (N) and spatial weights matrix (W), it is almost impossible to derive one general approach that can be applied under all circumstances. In order to draw inferences regarding the statistical significance of the direct and indirect effects, LeSage & Pace (p. 39) therefore suggest simulating the distribution of the direct and indirect effects using the variance–covariance matrix implied by the maximum likelihood estimates.

Elhorst & Fréret (2009) show that the variance–covariance matrix of the parameter estimates of the spatial Durbin model, a matrix that is not reported in LeSage & Pace's book, takes the form:

$$\mathrm{Var}(\hat{\rho}, \hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\sigma}^2) = \begin{bmatrix} \mathrm{trace}(\tilde{W}\tilde{W} + \tilde{W}^T\tilde{W}) + \frac{1}{\hat{\sigma}^2}\hat{\gamma}\tilde{X}^T\tilde{W}^T\tilde{W}\tilde{X}\hat{\gamma} & . & . \\ \tilde{X}^T\tilde{W}\tilde{X}\hat{\gamma} & \frac{1}{\hat{\sigma}^2}\tilde{X}^T\tilde{X} & . \\ \frac{1}{\hat{\sigma}^2}\mathrm{trace}(\tilde{W}) & 0 & \frac{N}{2\hat{\sigma}^4} \end{bmatrix}^{-1}, \tag{8}$$

where $\tilde{W} = W(I - \hat{\rho}W)^{-1}$, $\tilde{X} = [\iota_N \quad X \quad WX]$ and $\hat{\gamma} = [\hat{\alpha} \quad \hat{\beta}^T \quad \hat{\theta}^T]^T$ to simplify notation. Since this matrix is symmetric the upper diagonal elements are not shown.

Using the Matlab routine 'sar' posted on LeSage's website (www.spatial-econometrics.com), one particular parameter combination drawn from this variance–covariance matrix (indexed by d) can be obtained by:

$$[\rho_d \quad \alpha_d \quad \beta_d^T \quad \theta_d^T \quad \sigma_d^2]^T = P^T\vartheta + [\hat{\rho} \quad \hat{\alpha} \quad \hat{\beta}^T \quad \hat{\theta}^T \quad \hat{\sigma}^2]^T, \tag{9}$$

where P denotes the upper-triangular Cholesky decomposition of $\mathrm{Var}(\hat{\rho}, \hat{\alpha}, \hat{\beta}, \hat{\theta}, \hat{\sigma}^2)$ and ϑ is a vector of length $3+2K$ (the number of parameters that have been estimated) containing random values drawn from a normal distribution with mean zero and standard deviation one. If D parameter combinations are drawn like this and the (in)direct effect of a particular explanatory variable is determined for every parameter combination, the overall (in)direct effect can be approximated by computing the mean value over these D draws and its significance level (t-value) by dividing this mean by the corresponding standard deviation. If μ_{kd} denotes the indirect effect of the kth explanatory variable of draw d,[11] the overall indirect effect over all draws and the corresponding t-value will be:

$$\bar{\mu}_k \text{ (ind. eff. } k\text{th var.)} = \frac{1}{D}\sum_{d=1}^{D}\mu_{kd}, \tag{10a}$$

$$t\text{-value (of ind. eff. } k\text{th var.)} = \bar{\mu}_k \bigg/ \left[\frac{1}{D-1}\sum_{d=1}^{D}(\mu_{kd} - \bar{\mu}_k)^2\right]. \qquad (10b)$$

Given the t-value of this indirect effect, one can finally test whether the kth variable causes spatial spillover effects.

7. Spatio-temporal Models

Figure 1 uncovered the relationships between eight different spatial dependence models for cross-section data. A similar type of scheme for space-time data has been reported in Elhorst (2001), with as the most general model a first-order autoregressive distributed lag model in both space and time:

$$Y_t = \tau Y_{t-1} + \rho WY_t + \eta WY_{t-1} + X_t\beta + WX_t\theta + X_{t-1}\varphi + WX_{t-1}\phi + \varepsilon_t. \quad (11)$$

In this specification Y_t denotes an $N \times 1$ vector consisting of one observation on the dependent variable for every unit in the sample ($i = 1, \ldots, N$), and X_t an $N \times K$ matrix of exogenous explanatory variables, both measured at different points in time. The model in (11) subsumes several simpler econometric models, among which is the spatial Durbin model ($\tau = \eta = \varphi = \phi = 0$). However, just as with the Manski model discussed in Section 2, it suffers from an identification problem (see Anselin *et al.*, 2008). This raises the question as to which spatio-temporal models are feasible. Up to now, at least two interesting candidates have been proposed in the literature, including one by LeSage & Pace (Ch. 7). The names of these models are taken from Anselin *et al.* (2008).

The first candidate is the time-space dynamic model

$$Y_t = \tau Y_{t-1} + \rho WY_t + \eta WY_{t-1} + X_t\beta + \varepsilon_t, \quad (12)$$

whose properties have been analysed by Yu *et al.* (2008).[12] By continuous substitution of $Y_{t\text{-}1}$ up to $Y_{t-(T-1)}$ into this equation and rearranging terms, (12) can be rewritten as:

$$Y_t = (I - \rho W)^{-T}(\tau I + \eta W)^T Y_{t-T} + \sum_{p=1}^{T}(I - \rho W)^{-p}(\tau I + \eta W)^{p-1} \times (X_{t-(p-1)}\beta + \varepsilon_{t-(p-1)}). \quad (13)$$

Anselin *et al.* (2008) criticize this model because it might still create identification problems. The difficulty is that two global spatial multiplier matrices are at work at the same time, $(I - \rho W)^{-p}$ and $(\tau I + \eta W)^{p-1}$, rather than one process that produces global spatial spillover effects and another one that produces local spatial spillover effects. Anselin *et al.* (2008) therefore suggest setting $\eta = 0$ (the model so obtained is labelled the time-space simultaneous model). In his keynote speech at the third World Conference of the Spatial Econometrics Association in 2009, James LeSage suggested $\eta = -\tau\rho$. This restriction has the effect that all higher-order terms of the two spatial multiplier matrices cancel each other out, except for the factor τ, as a result of which (13) simplifies to:

$$Y_t = \tau^T Y_{t-T} + \sum_{p=1}^{T} \tau^{p-1} (I - \rho W)^{-1} (X_{t-(p-1)}\beta + \varepsilon_{t-(p-1)}). \quad (14)$$

Using (7), (14) can be rewritten as:

$$Y_t = \tau^T Y_{t-T} + \sum_{p=1}^{T} \left\{ \tau^{p-1} \left[\sum_{q=0}^{\infty} \rho^q W^q (X_{t-(p-1)}\beta + \varepsilon_{t-(p-1)}) \right] \right\}. \quad (15)$$

This equation shows that the impact of a change in one of the explanatory variables gradually diminishes over both space and time, provided that $|\tau| < 1 - \rho$ if $\rho \geq 0$ and $|\tau| < 1 - \rho r_{min}$ if $\rho < 0$ (see Elhorst, 2008). Furthermore, both effects can be separated from each other mathematically. The impact of a change in one of the explanatory variables over space falls by the factor ρW for every higher-order neighbour, and over time by the factor τ for every next time period. It is these kinds of schemes that might be used to uncover short-term and long-term direct effects, as well as short-term and long-term indirect effects. Furthermore, these kinds of schemes will further contribute to a better understanding (e.g. Groote *et al.*, 2009) and thus to the popularity of spatio-temporal models.

Since (15) does not contain local spillover effects, the original model in (12) may be extended to include spatially lagged independent variables $WX_t\theta$, provided that this model is estimated under the condition $\eta = -\tau\rho$ to avoid identification problems. In doing so, this spatio-temporal model can be used, just as with the spatial Durbin model for cross-sectional data, to estimate both global and local spatial spillover effects without imposing prior restrictions on the magnitude of these effects.

The second candidate is the time-space recursive model

$$Y_t = \tau Y_{t-1} + \eta W Y_{t-1} + X_t\beta + WX_t\theta + \varepsilon_t, \quad (16)$$

introduced by LeSage & Pace (pp. 190–191). By continuous substitution and rearranging terms, this model can be rewritten as:

$$Y_t = (\tau I + \eta W)^T Y_{t-T} + \sum_{p=0}^{T-1} (\tau I + \eta W)^p (X_{t-p}\beta + WX_{t-p}\theta + \varepsilon_{t-p}). \quad (17)$$

This equation shows that Y_t depends not only on present but also on past values of X_t and WX_t. This is exactly the reason why the extension of (16) with X_{t-1} and WX_{t-1} would create identification problems. If the stationarity conditions are satisfied,[13] the impact of the matrix $(\tau I + \eta W)^p$ diminishes over both space and time. To determine the extent of this, the matrix of partial derivatives with respect to the kth explanatory for every spatial unit may again be contemplated. One difference between this and the earlier spatio-temporal model is that the decomposition of the matrix $(\tau I + \eta W)^p$ is more complex. If τ is non-zero, we have:

$$\begin{aligned}(\tau I + \eta W)^p &= \tau^p (I + \tfrac{\eta}{\tau} W)^p \\ &= \tau^p [I + p\tfrac{\eta}{\tau} W + \tfrac{p(p-1)}{2!} (\tfrac{\eta}{\tau} W)^2 + \tfrac{p(p-1)(p-2)}{3!} (\tfrac{\eta}{\tau} W)^3 + \ldots]. \quad (18)\end{aligned}$$

This equation shows that the impact of a change in one of the explanatory variables over time falls by the factor τ for every next time period, and that the impact over space falls by the factor

$$\left(\frac{p+1}{O}-1\right)\frac{\eta}{\tau}W$$

for every higher-order neighbour, where O denotes the order number of that neighbour (e.g. when moving from first- to second-order neighbours, $O = 2$). Therefore, this spatio-temporal model can, just as the spatial Durbin for cross-sectional data and the earlier spatio-temporal model, be used to estimate both global and local spatial spillover effects without imposing prior restrictions on the magnitude of these effects.

8. Conclusions

The overall conclusion of this paper is that the state of the art of applied spatial econometrics has taken a step change with the publication of LeSage & Pace's book. One can no longer restrict oneself to the spatial lag and/or the spatial error model, or to simply interpreting their point estimates and testing whether these point estimates are robust to different specifications of the spatial weights matrix. A state-of-the-art application of spatial econometrics should also consider the spatial Durbin model, and interpret its direct and indirect effects, unless statistical tests show that simpler models suffice. There are two major reasons why the spatial Durbin model cannot be ignored. First, it is the only means of producing unbiased coefficient estimates, even if the true data-generation process is a spatial lag, spatial error, Kelejian–Prucha or spatial Durbin error model. Second, it produces both global and local spillover effects and, related to that, it does not impose prior restrictions on the magnitude of these effects.

The conclusion from the short overview of spatio-temporal models is that models with similar types of properties are also available for space-time data. An application and a comparison of these spatio-temporal models are the topic of further research.

One important limitation of LeSage & Pace's book is that it does not pay attention to instrumental variables or generalized method of moments (IV/GMM) techniques. Whereas IV/GMM estimators can easily handle linear spatial dependence models containing one or more endogenous explanatory variables (other than the spatially lagged dependent variable), single equation maximum likelihood and Bayesian estimators cannot.

Notes

1. The superscript T denotes the transpose of a vector or matrix.
2. Lee (2007) has found that a Manski type of model is not beyond the bounds of possibility, provided that one is willing to accept that the interaction effects among the error terms have a different spatial weights matrix than the interaction effects among the dependent variable and among the independent variables. Lee considers G groups, each consisting of N_g cross-sectional units, and assumes that the elements of the spatial weights matrix measuring the endogenous and exogenous interaction effects are $w_{ij} = 1/(N_g - 1)$ if units i and j belong to the same group (except if $i = j$), and zero otherwise. To account for correlated errors among the members of each group, Lee considers group fixed effects. Mathematically, these group fixed effects can be represented by a spatial weights matrix whose elements are all equal to $w_{ij} = 1/N_g$ if units i and j belong to the same group, including the diagonal elements (i.e. if $i = j$), and zero otherwise. Starting with these spatial weights matrices, Lee proves that the parameters are identified either if both N and N_g tend to infinity, with at least two units in each group, or if the number of units in each group does not tend to infinity faster than or equal to the number of groups.

3. Lee (2003) introduces the optimal instrument 2SLS estimator, but Kelejian *et al.* (2004) show that the 2SLS estimator based on this set of instruments has quite similar small-sample properties.
4. One objection to this study is that the null rejection frequencies have not been standardized (Hendry, 2006). Another objection is that the model that has been used in Florax *et al.* (2003) as a point of departure did not include spatially lagged independent variables.
5. Based partly on Elhorst & Fréret (2009) and Seldadyo *et al.* (2010).
6. The latter tests are called robust because the existence of one type of spatial dependence does not bias the test for the other type of spatial dependence.
7. If the (robust) LM tests reject these extensions of the OLS model, it is still useful to estimate the spatial lag and the spatial error model. In case the spatial autoregressive coefficient ρ and/or the spatial autocorrelation coefficient λ turns out to be significant, we may again conclude that the OLS model must be rejected in favour of the spatial lag, the spatial error model or in favour of both models, and continue estimating the spatial Durbin model.
8. In order to obtain an asymmetric spatial weights matrix W, unit 1 is not assumed to be a neighbour of unit 3. If these units were also assumed to be neighbours, we would have a symmetric matrix which is not general enough.
9. I abstract here from the problem that a cross-sectional model with an intercept, WY, X and WX variables cannot be estimated on the basis of $N=3$ observations.
10. This also holds if the spatial autoregressive parameter is negative. The first term that produces feedback effects is $\rho^2 W^2$. This term will always be positive. The second term is $\rho^3 W^3$. Since ρ is restricted to the interval $(1/r_{min}, 1)$ and the non-negative elements of W after row normalization are smaller than or equal to 1, the diagonal elements of $\rho^3 W^3$ are smaller in absolute value than those of $\rho^2 W^2$. Since the series $\rho^2 W^2 + \rho^3 W^3 + \rho^4 W^4 + \ldots$ alternates in sign if ρ is negative, the sum of the diagonal elements of the matrix represented by this series will always be positive.
11. For example, the indirect effect of the kth variable in the spatial Durbin model of draw d that is used for illustration purposes in Table 1 would be equal to

$$\mu_{kd} = \frac{3\rho_d + \rho_d^2}{3(1-\rho_d^2)}(\beta_k)_d + \frac{3+\rho_d}{3(1-\rho_d^2)}(\theta_k)_d.$$

12. They also include spatial fixed effects.
13. $|\tau| < 1 - \eta$ if $\eta \geq 0$ and $|\tau| < 1 - \eta r_{min}$ if $\eta < 0$.

References

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

Anselin, L., Bera, A. K., Florax, R. & Yoon, M. J. (1996) Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics*, 26, 77–104.

Anselin, L., Le Gallo, J. & Jayet, H. (2008) Spatial panel econometrics, in: L. Matyas & P. Sevestre (eds) *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, 3rd edn, pp. 627–662, Dordrecht, Kluwer.

Durbin, J. (1960) Estimation of parameters in time-series regression models, *Journal of the Royal Statistical Society B*, 22, 139–153.

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

Elhorst, J. P. (2008) Serial and spatial autocorrelation, *Economics Letters*, 100, 422–424.

Elhorst, J. P. & Fréret, S. (2009) Evidence of political yardstick competition in France using a two-regime spatial Durbin model with fixed effects, *Journal of Regional Science*, 49, 931–951.

Ertur, C. & Koch, W. (2007) Growth, technological interdependence and spatial externalities: theory and evidence, *Journal of Applied Econometrics*, 22, 1033–1062.

Fingleton, B. & Le Gallo, J. (2007) Finite sample properties of estimators of spatial models with autoregressive, or moving average disturbances and system feedback, *Annales d'économie et de statistique*, 87/88, 39–62.

Fingleton, B. & Le Gallo, J. (2008) Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: finite sample properties, *Papers in Regional Science*, 87, 319–339.

Florax, R. J. G. M., Folmer, H. & Rey, S. J. (2003) Specification searches in spatial econometrics: the relevance of Hendry's methodology, *Regional Science and Urban Economics*, 33, 557–579.

Greene, W. H. (2005) *Econometric Analysis*, 6th edn, Upper Saddle River, NJ, Pearson Prentice Hall.

Groote, P., Elhorst, J. P. & Tassenaar, V. (2009) Standard of living effects due to infrastructure improvements in the 19th century, *Social Science Computer Review*, 27, 380–389.

Harris, R. & Kravtsova, V. (2009) In search of 'W', Paper presented to the third World Conference of the Spatial Econometric Association, Barcelona, 8–10 July.

Hendry, D. F. (2006) A comment on 'Specification searches in spatial econometrics: the relevance of Hendry's methodology', *Regional Science and Urban Economics*, 36, 309–312.

Kelejian, H. H. & Prucha, I. R. (1998) A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics*, 17, 99–121.

Kelejian, H. H. & Prucha, I. R. (1999) A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review*, 40, 509–533.

Kelejian, H. H., Prucha, I. R. & Yuzefovich, Y. (2004) Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results, in: J. P. LeSage & K. Pace (eds) *Spatial and Spatiotemporal Econometrics*, pp. 163–198, Amsterdam, Elsevier.

Lee, L.-F. (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, *Econometric Reviews*, 22, 307–335.

Lee, L.-F. (2004) Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models, *Econometrica*, 72, 1899–1925.

Lee, L.-F. (2007) Identification and estimation of econometric models with group interactions, contextual factors and fixed effects, *Journal of Econometrics*, 140, 333–374.

Leenders, R. T. A. J. (2002) Modeling social influence through network autocorrelation: constructing the weight matrix, *Social Networks*, 24, 21–47.

LeSage, J. P. & Pace, R. K. (2009) *Introduction to Spatial Econometrics*, Boca Raton, Taylor & Francis.

Manski, C. F. (1993) Identification of endogenous social effects: the reflection problem, *Review of Economic Studies*, 60, 531–542.

Mood, A. M., Graybill, F. & Boes, D. C. (1974) *Introduction to the Theory of Statistics*, 3rd edition, Tokyo, McGraw-Hill.

Pesaran, M. H. (2006) Estimation and inference in large heterogeneous panels with multifactor error structure, *Econometrica*, 74, 967–1012.

Seldadyo, H., Elhorst, J. P. & de Haan, J. (2010) Geography and governance: does space matter?, *Papers in Regional Science* (forthcoming).

Stakhovych, S. & Bijmolt, T. H. A. (2009) Specification of spatial models: a simulation study on weights matrices, *Papers in Regional Science*, 88, 389–408.

Yu, J., de Jong, R. & Lee, L. (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both *n* and *T* are large, *Journal of Econometrics*, 146, 118–134.