HOW TO GET RID OF W?

A LATENT VARIABLES APPROACH TO MODELING SPATIALLY LAGGED VARIABLES

Henk Folmer [1]

Johan Oud [2]

[1]Department of Spatial Sciences, University of Groningen, P.O. Box 800, NL-9700AV Groningen, The Netherlands
and
Department of Social Sciences, Wageningen University, P.O. Box 8130, NL-6700 EW Wageningen, The Netherlands, E-mail: Henk.Folmer@wur.nl

[2]Behavioural Science Institute, Radboud University Nijmegen, P.O. Box 9104, NL-6500 HE Nijmegen, The Netherlands, E-mail: j.oud@pwo.ru.nl

Abstract

In this paper we propose a Structural Equation Model (SEM) approach with latent variables to model spatial dependence. Rather than using the spatial weights matrix W, we propose to use latent variables to represent spatial dependence and spill-over effects, of which the observed spatially lagged variables are indicators. This approach allows to incorporate and test more information on spatial dependence and offers more flexibility than the representation in terms of Wy or Wx. Furthermore, we adapt the estimators included in the software packages Mx and LISREL 8 to estimate SEMs with spatial dependence. We present an illustration based on Anselin's (1988) Columbus, Ohio, crime data set.

Keywords: Structural equation model, Spatial weights matrix, Latent variable, Maximum likelihood estimator, LISREL 8, Mx, Columbus Ohio crime data set

1. Introduction

Spatial dependence and spill-over effects are conventionally represented by means of the spatial weights matrix W. This matrix records the contiguity relations for each region in its corresponding row. The selection of spatial weights is a crucial feature of spatial models because it imposes a priori a structure of spatial dependence on the model and affects estimates (Bhattacharjee and Jensen-Butler, 2006; Anselin, 2002 and Fingleton, 2003) and substantive interpretation of the research (Hemple, 1995). Therefore, it is not surprising that the last two decades have seen major theoretical and methodological developments in specifying the basic structure of the weights matrix. For instance, Bavaud (1998) discusses some theoretical issues related to spatial weights while Hemple (1995a,b) develops a Bayesian posterior probabilities approach for the comparison of spatial weights matrices for both the systematic and the disturbance components. Getis and Aldstadt (2003) focus on constructing the spatial weights matrix using a local statistic while Aldstadt and Getis (2006) present a multidirectional optimum eco-type-based algorithm for the construction of a spatial weights matrix. Bhattacharjee and Jensen-Butler (2006) develop a method for estimating spatial weights matrices that are consistent with an observed pattern of spatial dependence rather than assuming a priori the nature of spatial interactions. The interest in contiguity relations is not restricted to spatial sciences but extends to other social sciences, particularly those subfields that focus on network analysis (see, amongst others, Leenders (2002) and Gould (1991)).

In spite of the theoretical and methodological developments mentioned above, most applied work still proceeds via a priori specification of a spatial weights matrix W based on first order double rook contiguity. Other types of contiguity, higher order contiguity or spatial clustering based on theoretical considerations or that are consistent with an observed pattern of spatial dependence are still rather rare. Similarly for the inverse distance matrix. Moreover, little effort has been invested in constructing and testing alternative weights matrices.

The standard, W-based approach to modelling contiguity relations is rigorous and inflexible. For instance, if only one spatial weights matrix is used (which usually is the case in practice), it is not possible to simultaneously include contiguity to first order neighbours as well as spillover from e.g. core regions. Moreover, spatial dependence of a given kind is captured by one parameter only which measures the average influence of neighbouring observations on observations of some dependent variable.

This paper proposes an alternative representation of spatial dependence that allows for the inclusion of more detailed information on the impact of spatial dependence , for testing of the a priori imposed structure and for detailed substantive interpretation. Particularly, we introduce the class of Structural Equation Models (SEM) with latent variables to model spatial dependence. In order to further clarify the purpose of the paper, we briefly describe the notion of a latent variable and some characteristics of SEMs.

Latent variables or theoretical constructs refer to those phenomena that are supposed to exist but cannot be directly observed. A well-known example a of latent variable is regional welfare. Observable variables on the other hand possess direct empirical meanings derived from experience. Latent variables can only be observed and measured by means of observable variables. For instance, the latent variable regional welfare is measured by observed variables at the regional level such as per capita GDP, income distribution aspects, employment opportunities, features of the housing market, health indicators, indicators of environmental quality, etc. The simultaneous use of both latent and observable variables in one modelling framework has, amongst others, the advantages that latent variables are given empirical meanings by means of operational definitions; that a closer correspondence between theory and empirics is obtained; that measurement errors are accounted for, and that the impacts of multicollinearity can be mitigated. (See amongst others Blalock (1971) and Folmer (1986) and the references therein for further details.)

The class of SEMs proposed here makes it possible to simultaneously estimate theoretical statements (which contain latent variables only) and correspondence statements (which contain both latent and observable variables). Particularly, a SEM is made up of two related sub-models:

- A structural model which represents the relationships between the latent variables.

- A latent variables measurement model which represents the relationships between the latent variables and their observable indicators.

The approach we are proposing here is to replace the spatially lagged variables Wy (spatial lag model) or Wx (spatially lagged exogenous variables) in the structural model by latent variables and to model the relationship between a latent spatially lagged variable and its observed variables in the measurement model. Since one latent variable can be measured by several indicators, this approach allows for the straightforward inclusion of several kinds of spatial dependence in the model. For instance, for a given latent variable representing spatial dependence (denoted latent spatial dependence variable in the sequel) , say investment, investments of the nearest and next nearest neighbour and distance weighted investments in the core regions could be indicators. In addition to testing overall spatial dependence in the structural model via a test of the significance of the coefficient of the latent spatial dependence variable (similar to the conventional approach of testing the significance of the coefficient of Wy or Wx), the model set up would allow further testing, e.g. of distance decay via the measurement model (for example, whether or not the third nearest neighbour and beyond exert influence). We observe that this is in line with Getis and Aldstadt's (2004) suggestion that spatial structure should be considered in a two-part framework: those units that evoke a distance effect, and those that do not.

Our approach also allows for several different types of contiguity including non-spatial contiguity. For instance, it allows the inclusion of similarity relationships, e.g. the relationships between regions that are economically and demographically similar (see Case et al., 1993). This is because the framework allows measurement by several sets of indicators, e.g. contiguity and similarity. Alternatively, several latent spatial dependence variables can be included in the structural model with corresponding observed variables in the measurement model.

The following observations apply. First, conventional W-based spatial econometric modelling also allows inclusion of various types of spatial dependence in the model, e.g. first and second order contiguity. However, as observations on neighbours tend to be strongly correlated, this is likely to lead to multicollinearity. In our SEM approach this problem is mitigated because of the presence of the latent spatially lagged variable in the structural model rather than the observed variables that are responsible for the multicollinearity problem. Since the observed variables are dependent variables in the measurement model (see equations (2) and (3) below), multicollinearity is not a problem in this model either. Secondly, the approach proposed here is not only instrumental to obtaining consistent, unbiased and efficient estimators by controlling for spatial dependence but also provides substantive information. Particularly, it does not only allow estimating and testing of the overall impact of spatial dependence via the coefficient of the latent spatial dependence variable but also of e.g. distance decay via the coefficients in the measurement model.

This paper is organized as follows. In section 2 we briefly introduce the class of SEMs. In section 3, in a bid to show that standard spatial dependence models can be routinely estimated by SEM software (after correction of the likelihood), we specify the lag model as a SEM and estimate it for Anselin's (1988) Columbus, Ohio, crime data set applying the SEM software package Mx. In section 4 we present the latent variable approach to the spatially lagged dependent variable model such that in the structural model Wy is replaced by a latent variable while in the measurement model the latent spatially lagged dependent variable is related to its indicators, i.e. observables in neighbouring units of observation. The model is applied to the Columbus, Ohio, crime data set again and compared to the benchmark model estimated in section 3. Section 5 concludes the paper.

2. The Structural Equation Model (SEM)

A SEM, as introduced by notably Jöreskog (1977), reads:

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \qquad \text{with} \qquad cov(\boldsymbol{\xi}) = \boldsymbol{\Phi}, cov(\boldsymbol{\zeta}) = \boldsymbol{\Psi} \ , \tag{1}$$

$$\mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon} \qquad \text{with} \qquad cov(\boldsymbol{\varepsilon}) = \boldsymbol{\Theta}_\varepsilon \ , \tag{2}$$

$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta} \qquad \text{with} \qquad cov(\boldsymbol{\delta}) = \boldsymbol{\Theta}_\delta \ . \tag{3}$$

where (1) is the structural model and (2) and (3) are the measurement models of the endogenous and exogenous latent variables, respectively. In the structural model the vector $\boldsymbol{\eta}$ contains the endogenous latent variables and the vector $\boldsymbol{\xi}$ the exogenous latent variables, $\boldsymbol{B}$ specifies the structural relationships among the latent endogenous variables and $\boldsymbol{\Gamma}$ contains the effects of the latent exogenous on the latent endogenous variables. $\boldsymbol{\Phi}$ is the covariance matrix of $\boldsymbol{\xi}$ and $\boldsymbol{\Psi}$ of the errors in $\boldsymbol{\zeta}$. In the measurement equations (2) and (3) the $\boldsymbol{\Lambda}$-matrices contain the loadings or regression coefficients of the observed variables on the latent variables, and the $\boldsymbol{\Theta}$-matrices contain the measurement error covariance matrices. The measurement errors in $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are assumed to be uncorrelated with one another as well as with the structural errors in $\boldsymbol{\zeta}$ and the $\boldsymbol{\xi}$. Moreover, the structural errors are assumed uncorrelated with the exogenous latent variables in $\boldsymbol{\xi}$. Finally, all errors are assumed to have expected value equal to zero.

We observe that several or all of the latent exogenous or endogenous in the structural model may be observed variables. For those cases an identity relationship holds in the corresponding measurement equation.

Several estimators for SEMs have been developed including instrumental variables (IV), two-stage least squares (TSLS), unweighted least squares (ULS), generalized least squares (GLS), fully weighted (WLS) and diagonally weighted least squares (DWLS), and maximum likelihood (ML). These estimators are available in the software packages Mx (Neale et al., 2003) and LISREL 8 (Jöreskog and Sörbom, 1996). These packages also include procedures to check model identification, to evaluate the estimation results and to calculate indirect and total effects. Especially Mx, a program that can be downloaded for free from internet, is very flexible and offers an extensive matrix algebraic toolbox. It also allows to impose various linear and nonlinear constraints on the model parameters and to modify and extend the likelihood-function in a user-defined way. Finally, it makes it possible to account for missing values by an individual likelihood procedure (Neale, 2000).

Below we restrict ourselves to the ML estimator, which maximizes the loglikelihood function of the free elements in the 8 parameter matrices in

(1)-(3) for given data $\mathbf{Y}$:

$$\ell(\boldsymbol{\theta}|\mathbf{Y}) = -\frac{N}{2}\ln|\boldsymbol{\Sigma}| - \frac{N}{2}\mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \frac{pN}{2}\ln 2\pi \ . \tag{4}$$

$\boldsymbol{\theta}$ in (4) contains the parameters to be estimated in the 8 matrices , $\mathbf{Y}_{(N\times p)}$ is the data matrix (N rows of independent replications of the p-variate vector $\mathbf{y}$, typically originating from a sample of randomly drawn subjects) and $\boldsymbol{\Sigma}_{(p\times p)}$ is the model-implied covariance or moment matrix (which is a function $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$):

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Lambda}_y\mathbf{H}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})\mathbf{H}'\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\varepsilon & \boldsymbol{\Lambda}_y\mathbf{H}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' \\ \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Gamma}'\mathbf{H}'\boldsymbol{\Lambda}_y' & \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta \end{bmatrix} \text{with} \ \ \mathbf{H} = \mathbf{I} - \boldsymbol{B}, \tag{5}$$

Finally, $\mathbf{S}_{(p\times p)} = \frac{1}{N}\mathbf{Y}'\mathbf{Y}$ is the sample covariance or moment matrix.

The ML-estimator $\hat{\boldsymbol{\theta}} = \text{argmax}\ \ell(\boldsymbol{\theta}|\mathbf{Y})$ chooses that value of $\boldsymbol{\theta}$ which maximizes $\ell(\boldsymbol{\theta}|\mathbf{Y})$. If the observed variables follow a multivariate standard normal distribution, maximization of $\ell(\boldsymbol{\theta}|\mathbf{Y})$ gives genuine maximum likelihood estimates. In the case of deviation from normality the standard errors produced by LISREL 8 and most other SEM programs should be interpreted with caution. The same applies to various statistics for model fit judgement, especially χ^2.

Instead of maximizing the loglikelihood function in (4) standard software for SEM estimation usually minimizes the fit function

$$F_{ML} = \ln|\boldsymbol{\Sigma}| + \mathrm{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \ln|\mathbf{S}| - p \tag{6}$$

or $\chi^2 = (N-1)F_{ML}$ with the same result. Because the data-based matrix $\mathbf{S}$ is a constant, (4) and (6) relate linearly.

As an introduction to the next section, we consider the vector of exogenous variables $\boldsymbol{\xi}$ in (1) as fixed and observed. In that case the loglikelihood function (4) reduces to (Oud, 2004):

$$\ell(\boldsymbol{\theta}|\mathbf{Y_0}) = -\frac{N}{2}\ln \mid \boldsymbol{\Sigma}_0 \mid -\frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_{0i}-\boldsymbol{\mu}_{0i})'\boldsymbol{\Sigma}_0^{-1}(\mathbf{y}_{0i}-\boldsymbol{\mu}_{0i}) - \frac{p_0 N}{2}\ln 2\pi \ , \qquad (7)$$

where the subscript 0 of $\mathbf{Y}_0$, $\mathbf{y}_0$, and $\boldsymbol{\mu}_0$ indicates that the exogenous variables are fixed and observed, p_0 is the number of observed variables in $\mathbf{Y}_0$ and $\mathbf{y}_0$, and

$$\boldsymbol{\mu}_0 \ = \ E(\mathbf{y}_0) = \boldsymbol{\Lambda}_y(\mathbf{I}-\boldsymbol{B})^{-1}\boldsymbol{\Gamma}\mathbf{x} \ , \qquad (8)$$

$$\boldsymbol{\Sigma}_0 \ = \ E[(\mathbf{y}_0-\boldsymbol{\mu}_0)(\mathbf{y}_0-\boldsymbol{\mu}_0)'] = \boldsymbol{\Lambda}_y\boldsymbol{\Psi}\boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\varepsilon \ . \qquad (9)$$

3. SEM representation of the observed spatial lag model

Before going into detail we observe that below it is important to distinguish between the traditional definition of a SEM in terms of variables (as in (1)-(3)) and a SEM defined in units of observation. If a vector refers to units of observation, it will denoted by a tilde ˜. Otherwise it refers to a vector of variables. Similarly for matrices.

We consider the standard (one equation) spatial lag model in units of observation:

$$\tilde{\boldsymbol{y}} = \rho \mathbf{W} \tilde{\boldsymbol{y}} + \tilde{\mathbf{X}} \boldsymbol{\gamma} + \tilde{\boldsymbol{\epsilon}} \ , \tag{10}$$

where

$\tilde{\boldsymbol{y}}$ is the $(N \times 1)$ vector with observations of the dependent variable y;

$\mathbf{W}$ is the $(N \times N)$ contiguity matrix;

$\tilde{\mathbf{X}}$ is the $(N \times q)$ matrix of observations of the q explanatory variables;

$\tilde{\boldsymbol{\epsilon}}$ is the $(N \times 1)$ vector of stochastic disturbances;

ρ is the spatial dependence parameter measuring the average influence of contiguous observations on y;

$\boldsymbol{\gamma}$ is the $q \times 1$ vector of regression coefficients of the q explanatory variables.

In SEM notation (10) reads

$$y = \rho\, y_W + \boldsymbol{\gamma}' \mathbf{x} + \epsilon \ , \tag{11}$$

where y_W is the spatially lagged dependent variable and $\mathbf{x}$ the vector of exogenous explanatory variables.

To develop a consistent and unbiased estimator of model (11) we need to take into account that the spatial lag variable y_W is a transformation $\mathbf{W}\tilde{\boldsymbol{y}}$ of the dependent variable and therefore cannot be assumed to be uncorrelated with the error term. As a first step we write model (11) as

$$(\mathbf{A}\tilde{\boldsymbol{y}} - \tilde{\mathbf{X}}\boldsymbol{\gamma}) = \tilde{\boldsymbol{\epsilon}} \qquad \text{where} \qquad \mathbf{A} = \mathbf{I} - \rho\mathbf{W} \tag{12}$$

Transformation of the vector of error terms to the vector of dependent variables leads to the addition of the Jacobian term $\ln|\mathbf{A}|$ to the loglikelihood function (8). For the one-equation case this leads to

$$\ell(\boldsymbol{\theta}|\tilde{\boldsymbol{y}}) \;=\; \ln \mid \mathbf{A} \mid - \frac{N}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \mu_i)^2 - \frac{N}{2}\ln 2\pi \quad , \tag{13}$$

with $\mu_i = \rho\, y_{Wi} + \boldsymbol{\gamma}'\mathbf{x}_i$ and the subscript i denoting individual observation i. (13) shows that the component $\ln|\mathbf{A}|$ is just added to the standard univariate loglikelihood. The SEM program Mx allows this component to be added to the standard fit function straightforwardly. Observe that as Mx minimizes $\chi^2 = (N-1)F_{ML} = -2(\frac{N-1}{N})[\ell(\boldsymbol{\theta}|\mathbf{Y}) + \text{ constant}]$, the correction to be applied to obtain the maximum likelihood solution by means of the Mx program is

$$-2(\tfrac{N-1}{N})\ln|\mathbf{A}| \ . \tag{14}$$

We apply the SEM model to the Anselin (1988) crime data set which relates crime to income and housing value for 49 contiguous neighbourhoods in Columbus, Ohio. The data matrix $\mathbf{Y}$ and contiguity matrix $\mathbf{W}$ have been obtained from Anselin (1988) and from website http://www.spatial-econometrics.com. The spatial SEM analyses are performed by the ML options of the SEM programs Mx and LISREL 8. We observe that the scope of this data set is limited. However, it is well-known, frequently used for illustrative purposes (e.g. Hepple, 1995) and limited in number of variables. These features facilitate highlighting the basic objectives of the illustration, and comparison of the estimates obtained by Mx to the outcome obtained by conventional software packages.

The (49×5) data matrix $[\tilde{\boldsymbol{y}}\ \mathbf{W}\tilde{\boldsymbol{y}}\ \tilde{\mathbf{X}}]$ consists of the five columns y (crime), y_W (spatially lagged crime), x_1 (income), x_2 (housing value), and 1 (unit variable). Because of the presence of the unit variable, the sample moment matrix $\mathbf{S}$ has the variable means in the last row and last column. In a one-equation model with observables only, $\boldsymbol{\Lambda}_y = \boldsymbol{\Lambda}_x = \mathbf{I}$, $\boldsymbol{\Theta}_\varepsilon = \boldsymbol{\Theta}_\delta = \mathbf{0}$, $\boldsymbol{B} = \mathbf{0}$, such that the model implied moment matrix (5) reduces to

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Gamma\Phi\Gamma}' + \boldsymbol{\Psi} & \boldsymbol{\Gamma\Phi} \\ \boldsymbol{\Phi\Gamma}' & \boldsymbol{\Phi} \end{bmatrix} . \tag{15}$$

The matrices $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Psi}$ in (15) read

$$
\begin{array}{ll}
 & \begin{array}{cccc} y_W & x_1 & x_2 & 1 \end{array} \\
\mathbf{\Gamma} = & \left[\begin{array}{cccc} \rho & \gamma_1 & \gamma_2 & \gamma_0 \end{array} \right] ,
\end{array}
$$

$$
\begin{array}{ll}
 & \begin{array}{cccc} y_W & x_1 & x_2 & 1 \end{array} \\
\mathbf{\Phi} = & \left[\begin{array}{cccc} E(y_W^2) & & & \\ E(y_W x_1) & E({x_1}^2) & & \\ E(y_W x_2) & E(x_2 x_1) & E({x_2}^2) & \\ \mu_{y_W} & \mu_{x_1} & \mu_{x_1} & 1 \end{array} \right] ,
\end{array}
$$

$$
\mathbf{\Psi} = \sigma^2 .
$$

Observe that in a model with observables only, the estimated moment matrix of the explanatory variables is equal to the corresponding sample moment matrix.

The total number of parameters to be estimated (including the 1 in $\mathbf{\Phi}$) is 15, while also the number of nonidentical elements in the 5×5 sample moment matrix $\mathbf{S}$ is 15. This would make the model just identified. However, the presence of $\ln |\mathbf{A}|$ in the loglikelihood puts a restriction on the model such that the degrees of freedom $(df) = 1$.

In Table 1 we present the estimates for model (10), specified as a SEM and estimated by Mx. We also present the results obtained by Anselin (1988) and Anselin and Bera (1999). In addition to the parameter estimates the Mx program also computes likelihood based confidence intervals for the parameters (Neale and Miller, 1997). However, we restrict ourselves to point estimates and ignore the confidence intervals. (They can be obtained from the authors upon request.) From Table 1 we conclude that the differences between the values obtained from the SEM-Mx procedure and Anselin's procedure are very small and within rounding errors. This means that the correction of the SEM loglikelihood function for spatial dependence and subsequent estimation by SEM software like Mx produces virtually the same estimates as standard spatial econometrics procedures and software.

	SEM-Mx	Anselin
ρ	0.431	0.431
γ_1	-1.031	-1.032
γ_2	-0.266	-0.266
γ_0	45.057	45.079
σ^2	95.504	95.495
$\ell(\boldsymbol{\theta}\|\mathbf{Y})$	-165.413	-165.408
$-2(\frac{N-1}{N}) \ln \|\mathbf{A}\|$	2.287	
corrected χ^2	3.036	
df	1	

Table 1: ML estimates of the observed spatial lag model for the Columbus, Ohio, crime data set by SEM-Mx and Anselin (1988).

4. Spatial dependence by means of a latent variable

The model in this section will be referred to as "latent dependence model". It replaces the spatially lagged variable y_W in the standard spatial lag model by a latent variable η:

$$y = \rho\,\eta + \boldsymbol{\gamma}'\mathbf{x} + \zeta \;, \tag{16}$$

and is completed by a measurement equation:

$$\mathbf{y} = \boldsymbol{\Lambda}\eta + \boldsymbol{\epsilon} \tag{17}$$

with

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{bmatrix}, \quad \boldsymbol{\Theta} = \begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & 0 & 0 \\ 0 & \sigma^2_{\varepsilon_2} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma^2_{\varepsilon_m} \end{bmatrix} . \tag{18}$$

(Observe that without an appropriate constraint, the loadings λ_i will not be identified. For that reason, one often chooses $\lambda_1 = 1$.)

We assume that the spatially lagged observed variables are chosen on the basis of theoretical or ad hoc considerations. Moreover, as pointed out in the introduction more than one spatial feature can be taken into account. For instance, both the relationships to neighboring regions as relationships to the core regions could be included.

We first turn to the measurement model. This model is constructed by means of selection functions or selection matrices $\mathbf{S}_i$ which select relevant observations from the vector of observations as follows:

$$\begin{aligned}
\tilde{\boldsymbol{y}}_1 &= \mathbf{S}_1\tilde{\boldsymbol{y}} \\
\tilde{\boldsymbol{y}}_2 &= \mathbf{S}_2\tilde{\boldsymbol{y}} \\
&\vdots \\
\tilde{\boldsymbol{y}}_m &= \mathbf{S}_m\tilde{\boldsymbol{y}}.
\end{aligned} \tag{19}$$

That is, $\mathbf{S}_1$ selects the values for the first indicator $\tilde{\boldsymbol{y}}_1$, $\mathbf{S}_2$ for the second indicator $\tilde{\boldsymbol{y}}_2$, etc. For example, $\mathbf{S}_1$ may be defined as the selector of the observations on, say crime, in the the nearest contiguous neighbours, $\mathbf{S}_2$ of the observations on crime in the next nearest contiguous neighbours, $\mathbf{S}_3$ as the spillover of crime from the core, measured, for example, as crime in the core times aneighbours distance from the core, etc.

For the measurement model we thus obtain:

$$\begin{aligned}
\tilde{\boldsymbol{y}}_1 &= \mathbf{S}_1\tilde{\boldsymbol{y}} = \lambda_1\tilde{\boldsymbol{\eta}} + \tilde{\boldsymbol{\epsilon}}_1 \\
\tilde{\boldsymbol{y}}_2 &= \mathbf{S}_2\tilde{\boldsymbol{y}} = \lambda_2\tilde{\boldsymbol{\eta}} + \tilde{\boldsymbol{\epsilon}}_2 \\
&\vdots \\
\tilde{\boldsymbol{y}}_m &= \mathbf{S}_m\tilde{\boldsymbol{y}} = \lambda_m\tilde{\boldsymbol{\eta}} + \tilde{\boldsymbol{\epsilon}}_m.
\end{aligned} \tag{20}$$

From the above one observes that spatial dependence is captured by two kinds of parameters, ρ and the λ_is, whereas in the standard lag model only the "average" effect ρy_W shows up. This means that a much richer representation and testing of the spatial structure can be obtained than by way of standard spatial econometric approaches. For instance, it allows for determining those units that evoke a distance effect and those that do not, as suggested by Getis and Aldstadt (2004) by testing the significance of the λ_i coefficients.

As in the preceding section, the standard SEM loglikelihood function needs correction so as to account for the presence of the spatially lagged dependent variable among the explanatory variables. To develop the appropriate loglikelihood function, we first use (20) to express $\tilde{\boldsymbol{\eta}}$ in terms of $\tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{\epsilon}}_i$s. Since there are m measurement equations in (20) we get

$$\begin{aligned} \tilde{\boldsymbol{\eta}} = (\frac{1}{m\lambda_1}\mathbf{S}_1 + \frac{1}{m\lambda_2}\mathbf{S}_2 \; + \; \cdots + \frac{1}{m\lambda_m}\mathbf{S}_m)\tilde{\boldsymbol{y}} \\ - \frac{1}{m\lambda_1}\tilde{\boldsymbol{\epsilon}}_1 - \frac{1}{m\lambda_2}\tilde{\boldsymbol{\epsilon}}_2 - \cdots - \frac{1}{m\lambda_m}\tilde{\boldsymbol{\epsilon}}_m \; . \end{aligned} \tag{21}$$

Next we write (16) in observation unit form to obtain

$$\tilde{\boldsymbol{y}} = \rho\,\tilde{\boldsymbol{\eta}} + \tilde{\mathbf{X}}\boldsymbol{\gamma} + \tilde{\boldsymbol{\zeta}} \; . \tag{22}$$

Substituting the right-hand side of (21) for $\tilde{\boldsymbol{\eta}}$ in (22) we obtain

$$\begin{aligned} (\mathbf{I} - \frac{\rho}{m\lambda_1}\mathbf{S}_1 \; - \; \frac{\rho}{m\lambda_2}\mathbf{S}_2 - \cdots - \frac{\rho}{m\lambda_m}\mathbf{S}_m)\tilde{\boldsymbol{y}} \\ = \tilde{\mathbf{X}}\boldsymbol{\gamma} + \tilde{\boldsymbol{\zeta}} - \frac{\rho}{m\lambda_1}\tilde{\boldsymbol{\epsilon}}_1 - \frac{\rho}{m\lambda_2}\tilde{\boldsymbol{\epsilon}}_2 - \cdots - \frac{\rho}{m\lambda_m}\tilde{\boldsymbol{\epsilon}}_m \end{aligned} \tag{23}$$

For

$$\mathbf{A} = \mathbf{I} - \frac{\rho}{m\lambda_1}\mathbf{S}_1 \; - \; \frac{\rho}{m\lambda_2}\mathbf{S}_2 - \cdots - \frac{\rho}{m\lambda_m}\mathbf{S}_m \tag{24}$$

(23) can be written

$$\mathbf{A}\tilde{\boldsymbol{y}} - \tilde{\mathbf{X}}\boldsymbol{\gamma} = \tilde{\boldsymbol{\zeta}} - \frac{\rho}{m\lambda_1}\tilde{\boldsymbol{\epsilon}}_1 - \frac{\rho}{m\lambda_2}\tilde{\boldsymbol{\epsilon}}_2 - \cdots - \frac{\rho}{m\lambda_m}\tilde{\boldsymbol{\epsilon}}_m \quad . \tag{25}$$

Next we standard-normalize the vector of disturbances at the right-hand side of (25) by pre-multiplication with $\mathbf{\Omega}^{-\frac{1}{2}}$:

$$\mathbf{\Omega}^{-\frac{1}{2}}(\mathbf{A}\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\gamma}) \; = \; \tilde{\boldsymbol{\nu}} \tag{26}$$

where $\mathbf{\Omega}$ is the covariance matrix of the vector of disturbances

$$\mathbf{\Omega} = \mathbf{I}(\sigma^2_\zeta \; + \; \tfrac{\rho^2}{m^2\lambda_1^2}\sigma^2_{\epsilon_1} + \tfrac{\rho^2}{m^2\lambda_2^2}\sigma^2_{\epsilon_2} + \cdots + \tfrac{\rho^2}{m^2\lambda_m^2}\sigma^2_{\epsilon_m}) \tag{27}$$

and $\tilde{\boldsymbol{\nu}}$ the standard normal vector. From (26) it follows that we need to add ln $\mid \mathbf{A} \mid$ with $\mathbf{A}$ defined in (24) to the loglikelihood function.

The error variance in (27) is a linear combination of the variance of the disturbance term in the structural model σ^2_ζ and of the measurement error variances, transformed by the squares of ρ, the loadings λ_i and the number of indicators chosen, m.

Before turning to the example, we address the feasible parameter region. As shown by, amongst others, Anselin (1988, pp.78-79) the feasible range of ρ in the standard, W-based, model is determined by the constraint that the determinant of $\mathbf{A} = \mathbf{I} - \rho\mathbf{W}$ is positive, which implies that the feasible range of ρ is determined by the eigenvalues of $\mathbf{W}$. For instance, for row-standardized $\mathbf{W}$ the range for ρ is from $+1$ to $-(1/\omega_{max})$ where ω_{max} is the maximum eigenvalue of $\mathbf{W}$ (Anselin, 1988, pp. 78-79). In terms of estimation this implies that the non-linear search across ρ-values should be within the feasible parameter range. From (24) it follows that in the latent variable approach the determinant of $\mathbf{A}$ is dependent on ρ, m, as well as various λs. Moreover, instead of one $\mathbf{W}$ there are several selection matrices $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_m$. Constrained estimation such that the parameters are within the feasible parameter range can be achieved by imposing the constraint $\mid \mathbf{A} \mid > 0$. In addition, further constraints derived from theory can be imposed, e.g. distance decay such that $\lambda_1 > \lambda_2 > \ldots \lambda_m$. The Mx program allows imposing these constraints straightforwardly.

Rather than imposing the constraints ex ante, it is preferable to start with unconstrained estimation and to test whether or not the parameter estimates are within the feasible parameter region, since this allows detecting misspecification. As shown by, amongst others, Leamer (1978) misspecification may show up in implausible or infeasible estimates. Constrained estimation may invalidate this vehicle of diagnostic checking. Only if there is no evidence of misspecification and parameter estimates are outside the feasible parameter region, constrained estimation is an option.

Below we illustrate the proposed approach using the Columbus crime data set again. We shall show that the latent dependence model allows a much richer representation and testing of the spatial structure. Moreover, we shall discuss how the estimation results compare to those obtained by standard procedures. For that purpose we present two models: one based on three first-order contiguity neighbours ordered by distance (Model3) and one on six (Model6). The first selection in Model3 is possible for all neighbourhoods except seven which only have two contiguous neighbours. For these cases we select the nearest non-contiguous neighbourhood. In Model6, we also select the fourth, fifth and sixth neighbour, when present. Again, if not present, we supplement the set of neighbours by the nearest non-contiguous neighbourhoods. So, the selection matrices (19) are based on contiguity and distance such that for each region the three (six) nearest contiguous or non-contiguous neighbourhoods are selected. Hence, $\mathbf{S}_1$ as the selector of the nearest contiguous neighbour, $\mathbf{S}_2$ of the next nearest contiguous neighbour, $\mathbf{S}_3$ of the third nearest contiguous or non-contiguous neighbour, etc. The variables thus formed will be called Crime-neighbour1, Crime-neighbour2, etc.

Whereas the observed spatial lag model in the previous section only includes the parameter matrices $\mathbf{\Gamma}$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$, the latent dependence model additionally contains the measurement model matrices $\mathbf{\Lambda}$ and $\mathbf{\Theta}$. The parameter matrices $\mathbf{\Gamma}$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$ are the same as in the observed spatial lag model (see below equation (15)). However, y_W is replaced by η. (Observe that as is customary in factor analysis, the mean of the latent η is fixed at zero.)

We now turn to the number of degrees of freedom (*df*). The matrices $\mathbf{\Gamma}$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$ contain 14 parameters that need to be estimated: ρ, γ_1, γ_2, and γ_0 in $\mathbf{\Gamma}$, 6 moments and 3 means in $\mathbf{\Phi}$, and σ_ζ^2 in $\mathbf{\Psi}$. For Model3 there are an additional 8 unknown parameters in the measurement model matrices (2 loadings λ_i, 3 measurement error variances $\sigma_{\varepsilon_i}^2$, and 3 means). Model3 contains 7 observed variables which gives 28 non-identical elements in the observed moment matrix. Since the number of degrees of freedom equals the difference between the number of non-identical elements in the observed moment or covariance matrix and the number of unknown parameters that need to be estimated we obtain: $df = 28 - 14 - 8 = 6$. Moreover, the Jacobian term results in an extra restriction on the likelihood function which gives an additional (1) *df*. Hence, the total *df* is 7. In a similar vein, for Model6 we obtain $df = 25$ (10 observed variables with 55 nonidentical elements in the observed moment matrix and 17 additional measurement parameters).

Before turning to the estimation results we present a path diagram of

Model3 in Figure 1. The most important estimates produced by the SEM-Mx program are presented in Table 2 (standard errors can be obtained from the authors upon request). We first turn to the overall fit. Because its χ^2 of 8.940 is slightly higher than its *df*, the fit of the Model3 is quite good. For Model6 the χ^2 of 23.482 is even lower than the *df* which means that its fit is very good.

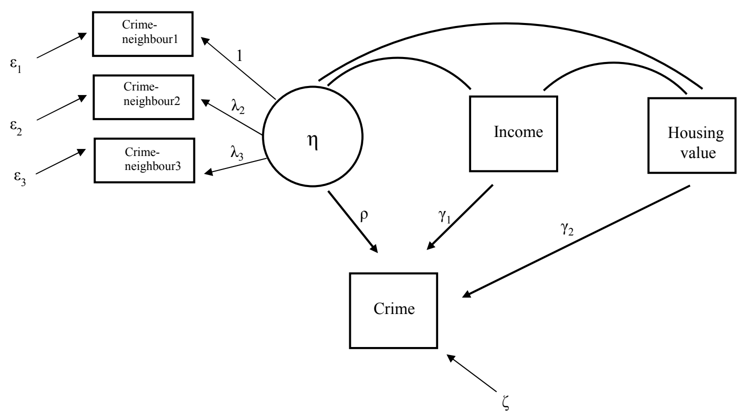

Figure 1: Path diagram of Model3: observed variables are indicated by squares, latent variables by circles, effects by arrows and covariances by curved lines.

Regarding the estimated parameters we first of all observe that λ_1 is fixed at 1 in both models so as to fix the measurement scale of the latent dependence variable. Next we observe that all parameter estimates are significant. We first turn to the measurement models. On the basis of distance decay, we expect the loadings of successive neighbours to go down. This does not happen uniformly due to the non-standardized character of the loadings, i.e. the measurement error variance is not

	Model3	Model6
ρ	0.521	0.491
γ_1	-0.864	-0.885
γ_2	-0.249	-0.259
γ_0	57.140	57.809
σ^2_ζ	79.559	83.432
σ^2_η	176.783	167.918
λ_1	1	1
λ_2	1.040	1.029
λ_3	0.959	1.013
λ_4	-	0.725
λ_5	-	0.756
λ_6	-	0.428
$\sigma^2_{\varepsilon_1}$	83.428 ($R^2 = 0.672$)	88.806 ($R^2 = 0.651$)
$\sigma^2_{\varepsilon_2}$	101.621 ($R^2 = 0.645$)	111.642 ($R^2 = 0.610$)
$\sigma^2_{\varepsilon_3}$	163.187 ($R^2 = 0.486$)	149.697 ($R^2 = 0.529$)
$\sigma^2_{\varepsilon_4}$	-	217.475 ($R^2 = 0.270$)
$\sigma^2_{\varepsilon_5}$	-	201.422 ($R^2 = 0.308$)
$\sigma^2_{\varepsilon_6}$	-	186.756 ($R^2 = 0.104$)
$-2(\frac{N-1}{N}) \ln \lvert \mathbf{A} \rvert$	3.768	3.616
corrected χ^2	8.940	23.482
df	7	25

Table 2: ML estimates of latent dependence models Model3 and Model6 for the Columbus, Ohio, crime data set by SEM-Mx.

taken into account. However, for the reliabilities, R^2=(squared loading)/(total variance)=1−(measurement error variance)/(total variance), we see a clearer downward trend (except for lambda 4 and lambda 5).

The R^2 of Crime-neighbour3 in Model3 is 0.486 which indicates that more distant neighbours could probably considerably contribute to the explanation of crime in a neighbourhood. This is confirmed in Model6 where

Crime-neighbour4 and Crime-neighbour5 have R^2s of 0.207 and 0.309, respectively. However, the R^2 of Crime-neighbour6 of 0.104 indicates that neighbours beyond neighbour 6 do not exert any further influence on crime in a given neighbourhood.

For the structural models we find $\rho = 0.521$ and $\rho = 0.491$ for Model3 and Model6, respectively, which are somewhat larger than for Anselin's observed lag model (see Table 1). The coefficients γ_1 and γ_2 for income and housing are close to the ones obtained by Anselin, though in both of our models the effect of income on crime (γ_1) is somewhat smaller and that of housing value (γ_2) somewhat larger than in Anselins model. Finally, the explanation error variance (σ^2_ζ) is in both models smaller than in Anselin's model. Simulation is needed to shed further light on the differences between both approaches.

6. Conclusion

In this paper we present a structural equation (SEM) approach to spatial dependence models. As a first step, we adapt the standard SEM likelihood function such that the conventional lag model can be estimated in a straightforward fashion by standard SEM software packages. Application to the Columbus, Ohio, crime data set (Anselin, 1998) shows that these packages produce virtually the same estimates as obtained by the standard software for spatial dependence models. Special attention is paid to the SEM software package Mx that allows handling of nonlinearities in a straightforward fashion.

Next, we introduce the latent spatial dependence approach as an alternative to the conventional W-based approach to represent spatial dependence and spillover effects. Typical for this approach is that Wy is replaced by a latent dependence variable in the structural model of which observed variables of spatial dependence are indicators. The latent variables approach is both more flexible and more informative than the conventional approach based upon an a priori given spatial weight matrix W. Particularly, it allows handling several different types of spatial dependence in one model framework and formal testing, e.g. of distance decay so as to identify the spatial units that evoke a distance effect and those that do not. The model is applied to the Columbus, Ohio, crime data set again. Although several parameter estimates obtained are in line with the results obtained by Anselin (1998), further research, particularly Monte Carlo simulation, is needed.

The SEM approach to spatial dependence models has several potential advantages which are worthwhile further exploring. Particularly, it allows straightforward application to systems of equations. Moreover, it allows the introduction of several dynamic SEM features into spatial modelling.

Literature

Aldstadt, J. and Getis, A. (2006) Using AMOEBA to create a spatial weights matrix and identify spatial clusters. *Geographical Analysis, 38,* 327-343.

Aldstadt, J. and Getis, A. (2004). Constructing the spatial weights matrix using a local statistic. *Geographical Analysis, 36,* 2, 90-104

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

Anselin, L., & Hudak, A.K. (1992). Spatial econometrics in practice: A review of software options. *Regional Science and Urban Economics*, 22, 509-536.

Bavaud, F. (1998). Models for spatial weights: A systematic look. *Geographical Analysis, 30,* 153-171.

Bhattacharjee, A. & Jensen-Butler, C. (2006) Estimation of the spatial weights matrix in the spatial error model, with an application to diffusion in housing demand. http://www.econ.cam.ac.uk/panel2006/papers/Bhattacharjeepaper23.pdf

Blalock, H.M. (Ed.) (1971). *Causal models in the social sciences.* London: MacMillan.

Case, A.C., Rosen, S. & Hines, Jr., J.R. (1993). Budget spillovers and fiscal interdependence. Evidence from the states. *Journal of Public*

Economics, 52, 285-307

Folmer, H., (1986). *Regional economic policy.* Dordrecht: Kluwer Academic Publishers.

Gould, R. V. (1991). Multiple networks and mobilization in the Paris Commune, 1871. *American Sociological Review, 56,* 716-729.

Hepple, L. W. (1995). Bayesian techniques in spatial and network econometrics: 1. Model comparison and posterior odds. *Environment and Planning A, 27,* 447-469.

Hepple, L.W. (1995) Bayesian techniques in spatial and network econometrics:2. Computational methods and algorithms. Environment and planning A, 27, 615-644.

Jöreskog, K.G. (1977). Structural equation models in the social sciences: Specification, estimation and testing. In P.R. Krishnaiah (Ed.), *Applications of statistics* (pp. 265-288). Amsterdam: North-Holland.

Jöreskog, K.G., & Sörbom, D. (1996). *LISREL 8: User's reference guide.* Chicago: Scientific Software International. Leamer, E.E. (1978)Specification Searches. New York, Wiley

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

Neale, M.C. (2000). Individual fit, heterogeneity, and missing data in multigroup structural equation modeling. In T.D. Little, K.U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 219-240). Mahwah NJ: Lawrence Erlbaum.

Neale, M.C., Boker, S.M., Xie, G., & Maes, H.H. (2003). *Mx: Statistical Modeling* (6th ed.). Richmond, VA: Department of Psychiatry.

Oud, J. H. L. (2004). SEM state space modeling of panel data in discrete and continuous time and its relationship to traditional state space modeling. In K. van Montfort, J. Oud & A. Satorra (Eds.), *Recent developments on structural equations models: Theory and applications* (pp. 13-40). Dordrecht: Kluwer Academic Publishers.