



Flexible dependence modeling using convex combinations of different types of connectivity structures

Nicolas Debarsy^{a,*}, James LeSage^b

^a CNRS, LEM - UMR 9221, Université de Lille, Cité Scientifique, Bat. SH2, F-59655, Villeneuve-d'Ascq, France

^b Texas State University-San Marcos, Department of Finance & Economics, 601 University Drive, San Marcos, TX 78666, USA

ARTICLE INFO

JEL:

C11
C21
C51
M12
L84

Keywords:

Spatial econometrics
Connectivity matrix
Salary benchmarking models
Markov Chain Monte Carlo estimation
Bayesian model probabilities
Convex combination

ABSTRACT

There is a great deal of literature regarding use of non-geographically based connectivity matrices or combinations of geographic and non-geographic structures in spatial econometrics models. We explore alternative approaches for constructing convex combinations of different types of dependence between observations. Pace and LeSage (2002) as well as Hazir et al. (2016) use convex combinations of different connectivity matrices to form a single weight matrix that can be used in conventional spatial regression estimation and inference. An example for the case of two weight matrices, W_1, W_2 reflecting different types of dependence between a cross-section of regions, firms, individuals etc., located in space would be: $W_c = \gamma_1 W_1 + (1 - \gamma_1) W_2, 0 \leq \gamma_1 \leq 1$. The matrix W_c reflects a convex combination of the two weight matrices, with the scalar parameter γ_1 indicating the relative importance assigned to each type of dependence. We explore issues that arise in producing estimates and inferences from these more general cross-sectional regression relationships in a Bayesian framework. We propose two procedures to estimate such models and assess their finite sample properties through Monte Carlo experiments. We illustrate our methodology in an application to CEO salaries for a sample of nursing homes located in Texas. Two types of weights are considered, one reflecting spatial proximity of nursing homes and the other peer group proximity, which arise from the salary benchmarking literature.

1. Introduction

Spatial regression models typically rely on spatial proximity to specify weight matrices, for example first-order neighbors (those with borders in common) or some number (say m) of nearest neighboring regions, or points (e.g., firms, consumers, houses) located in space. This approach has two advantages: 1) geographical location of observations is objective and easy to determine, and 2) weight matrices based on geographical space can be viewed as fixed over time and in most cases exogenous.¹ There has been a great deal of criticism of weight matrices based solely on spatial location of observations, (e.g., Corrado and Fingleton, 2012). This criticism in part derives from application of spatial regression models to broader contexts involving interregional flows of goods (e.g., trade), population (e.g., migration), knowledge (e.g., patent citations); student peer groups, social networks, etc., where geographical location of observations does not seem intuitively or theoretically appealing.

Further, early concerns of Fingleton (2003) regarding the theoretical and empirical basis for assumptions about the spatial reach of externalities, and methods for explicitly modeling knowledge spillovers between interacting firms or modeling knowledge flows due to job switching in labor market areas, etc. still remain largely unexplored. There are a limited number of studies where weight matrices reflecting connectivity of observations have been motivated by underlying theoretical considerations. For example Behrens et al. (2012) derive a quantity-based structural gravity equation system where both trade flows and error terms are cross-sectionally correlated based on population shares of regions in the sample, and Koch and LeSage (2015) show that the multilateral resistance concept from trade theory (Anderson and van Winkoop, 2003, 2004) can be viewed as a simultaneous autoregressive dependence structure involving gross domestic product shares of the sample regions as well as other types of generalized distance factors.

One response to dissatisfaction regarding use of connectivity structures based solely on spatial location has been the introduction of

* Corresponding author.

E-mail addresses: nicolas.debarsy@cnrs.fr (N. Debarsy), jlesage@spatial-econometrics.com (J. LeSage).

¹ There are cases where location of firms or consumers in space could result from an endogenous sorting process or some other underlying economic mechanism.

simultaneous autoregressive models that rely on more than one weight matrix (see Lacombe, 2004; Badinger and Egger, 2011; Lee and Liu, 2010; Elhorst et al., 2012; Liu et al., 2014). In these models, different weight matrices are introduced in an effort to capture different types of cross-sectional dependence within the same spatial autoregressive specification. Specifically, multiple spatial lags of the dependent variable vector y are introduced as right-hand side variables in an effort to extend conventional spatial autoregressive models to include what have been labeled ‘higher-order’ terms, as shown in (1), where L of such terms are introduced.

$$y = \left(\sum_{\ell=1}^L \rho_{\ell} W_{\ell} \right) y + X\beta + \varepsilon \quad (1)$$

In (1), y is an $n \times 1$ vector of dependent variable outcomes, X is an exogenous $n \times k$ explanatory variables matrix, with β the associated $k \times 1$ parameters, ρ_{ℓ} , $\ell = 1, \dots, L$ are scalar dependence parameters measuring the strength of each type of dependence modeled by the $n \times n$ connectivity matrices W_{ℓ} , $\ell = 1, \dots, L$. The $n \times 1$ disturbance vector is assumed to have independent error terms with zero mean and constant scalar variance (σ^2) across all n observations.

One example of a higher-order specification is Lacombe (2004), who explored a county-level cross-sectional spatial relationship involving program participation of state residents, where a spatial matrix W_1 is used to identify nearby counties located within the same state and a second spatial weight matrix W_2 captures the influence of nearby counties located in neighboring states. Liu et al. (2014) in a model of social interaction that explores peer effects rely on one weight matrix to capture local-average (social norm) influences of peers and a second weight matrix for influences reflected by local-aggregate peer effects (social multiplier). However, LeSage and Pace (2011) point to a number of estimation and interpretive issues that arise for models of the type in (1), and Elhorst et al. (2012) point out complications that arise regarding the parameter space for the dependence parameters ρ_{ℓ} , $\ell = 1, \dots, L$.

Another thread in the literature is to simply replace the spatial weight matrix with more appropriate types of connectivity structures, for example weight matrices based on friendship ties. Patacchini and Zenou (2012) analyze the role played by teenagers conformity to their peers’ behavior in producing juvenile crime outcomes in a social networking application. In the area of international finance, connectivity matrices may reflect real transmission channels for risk across countries, which might arise from trade or financial ties between countries. Alternatively, information transmission channels for risk might be reflected by financial market information that captures perceptions of market participants regarding own- and other-country risks (see Debarsy et al., 2017, and references therein). This suggests weight matrices based on trade, financial flows, or socio-economic similarities between countries.

A related literature is on methods for assessing different weight structures for their consistency with a specific economic relationship and set of sample data. Since models based on alternative weight matrices are likely to be non-nested, one approach in this literature uses the non-nested J test developed by Davidson and MacKinnon (1981) extended to spatial regression models by Anselin (1988). The power of alternative predictions for a host of spatial regression model specifications are explored in Kelejian (2008) and Kelejian and Piras (2011). Burridge and Fingleton (2010) and Burridge (2012) propose respectively to rely on bootstrap procedures for inference on the J test and to use maximum likelihood estimation rather than instrumental variables in the first step of the J test computation, within the Kelejian (2008) framework. Liu et al. (2014) propose an extension of the Kelejian (2008) J test to differentiate between the local-aggregate and the local-average endogenous peer effects in an econometric network model with network fixed effects. Debarsy and Ertur (2016) build on the J tests of Kelejian (2008) and Kelejian and Piras (2011) to allow for heteroskedasticity in a spatial autoregressive specification and further propose a procedure based on Hagemann (2012) to circumvent the deci-

sion problem inherent to non-nested models tests (the decision problem arises when non-nested tests do not lead to a clear choice between competing models). Alternatively, Jin and Lee (2013) consider a spatial model extension of the Cox test (Cox, 1961, 1962) for the case of non-nested models. In the context of determining the most relevant geographically based spatial weight matrix, Géniaux and Martinetti (2017) suggest to use different distance kernels with a single parameter h (representing the bandwidth or the number of neighbors, depending on the kernel). Identification of the matrix W is then based on a moment estimator that tries to minimize the residual sum of squares of the model estimation with respect to $W(h)$.

A Bayesian alternative to non-nested model tests is proposed by LeSage and Pace (2009) in order to select the most appropriate spatial weight matrix. In contrast to the J tests that rely on specific model estimates and their associated predictions, the Bayesian approach to model comparison integrates over all model parameters to calculate the log-marginal likelihood and associated model probabilities. This approach makes inference regarding the best weight matrix unconditional on any particular set of estimates. Note that parameter estimates based on models that rely on the wrong weight matrix will be biased, making it desirable to draw model comparison conclusions that are unconditional on the parameter estimates. LeSage (2014, 2015), extends this approach to simultaneously calculate log-marginal likelihoods and associated model probabilities for both cross-sectional and panel data model specifications and weight matrices. Again, inferences drawn based on posterior model probabilities are unconditional on parameter estimates from the host of alternative models considered.

Finally, Harris et al. (2011) provide a wide ranging discussion of techniques aimed at searching over alternative weight matrices for the best fit, approaches to estimating the weight matrix using non-parametric methods, correlation and iterative approaches, along with an illustration focused on establishment level R&D in the UK. While noting approaches based on hybrid combinations, their focus is on finding a single most appropriate weight matrix.

Our contribution to the spatial econometric literature regarding alternative weight matrices is to pursue an approach considered by Pace and LeSage (2002) as well as in Hazir et al. (2016), that relies on convex combinations of different connectivity matrices to form a single weight matrix. An advantage of this approach is that the resulting weight matrix can be used in conventional spatial regression methods to produce estimates and inference. This approach also avoids several issues raised in LeSage and Pace (2011) regarding estimation and interpretation of higher-order models that include spatial lags involving multiple different W matrices.

This convex combination approach proposes using $W_c = \sum_{\ell=1}^L \gamma_{\ell} W_{\ell}$, with $0 \leq \gamma_{\ell} \leq 1$, $\ell = 1, \dots, L$ and $\sum_{\ell=1}^L \gamma_{\ell} = 1$, in a standard spatial econometrics specification. The matrix W_c reflects cross-section dependence specified using a convex combination of L different types of connectivity between observations.² The scalar parameters γ_{ℓ} indicate the relative importance assigned to each type of dependence in the cross-sectional dependence scheme. When each W_{ℓ} , $\ell = 1, \dots, L$, is row-normalized, then W_c obeys the conventional row-normalization, which allows use of conventional spatial regression model specifications and estimation methods.³

In Section 2, we explore two alternative estimation strategies for determining estimates of γ_{ℓ} , $\ell = 1, \dots, L$ in this convex combination approach, one that calculates Bayesian posterior model probabilities

² In this paper, we do not address the potential endogeneity issue that may arise when weight matrices are not based on geographic proximity. The matrices entering the convex combination are thus assumed to be exogenous. For standard SAR cross-section models, Qu and Lee (2015) develop an estimator robust to endogeneity of the connectivity structure. However, in our context, we leave this question for further research.

³ Alternative types of normalization for connectivity matrices W_{ℓ} are possible (see Kelejian and Prucha, 2010). However, with the exception of special cases, normalization of each connectivity matrix W_{ℓ} by one of the convex norms proposed in Kelejian and Prucha (2010) does not result in a normalized convex combination matrix W_c .

Download English Version:

<https://daneshyari.com/en/article/7383658>

Download Persian Version:

<https://daneshyari.com/article/7383658>

[Daneshyari.com](https://daneshyari.com)