REM WORKING PAPER SERIES

The Inversion of the Spatial Lag Operator in Binary Choice Models: Fast Computation and a Closed Formula Approximation

Luís Silveira Santos and Isabel Proença

REM Working Paper 011-2017

October 2017

REM – Research in Economics and Mathematics
Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

The inversion of the spatial lag operator in binary choice models: fast computation and a closed formula approximation[☆]

Luís Silveira Santos[a], Isabel Proença[a,*]

[a] *ISEG-School of Economics and Management, Universidade de Lisboa; REM-Research in Economics and Mathematics; CEMAPRE*

Abstract

This paper presents a new method to approximate the inverse of the spatial lag operator matrix, used in the estimation of a spatial lag model with a binary dependent variable. The method is based on an approximation of the high order terms of the inverse series expansion. The proposed method is also applied to approximate other complex matrix operations and closed formulas for the elements of the approximated matrices are deduced. The approximated matrices are used in the gradients of a variant of Klier and McMillen's full GMM estimator, allowing to reduce the overall computational complexity of the estimation procedure. Monte Carlo experiments show that the new estimator performs well in terms of bias and root mean square error and exhibits a minimum trade-off between time and unbiasedness within a class of spatial GMM estimators. The new estimator is also applied to the analysis of competitiveness in the Metropolitan Statistical Areas of the United States of America. A new definition of binary competitiveness is proposed. Estimation of the spatial dependence parameter and the environmental effects are addressed as central issues.

Keywords: Matrix approximation, matrix factorization, Spatial binary choice models, Spatial lag operator inverse, Spatial nonlinear models

1. Introduction

Spatial binary choice models deal with dichotomous dependent variables that reflect the introduction of spatial dependence in choice outcomes. In fact, since profit maximizing or utility maximizing agents are able to interact in space, the observed choice of a given agent can be determined by the observed choices of neighboring agents. Hence, due to the nature of this decision,

[☆]The authors gratefully acknowledge the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia) for the financial support through national funds: Luís Silveira Santos for the Ph.D. scholarship SFRH/BD/92277/2013 and Isabel Proença for the CEMAPRE-UID/MULTI/00491/2013 program. The authors are also grateful to the participants of the XI World Conference of the Spatial Econometrics Association (SEA) for very helpful comments. An earlier draft of this paper, with the title "The inversion of the spatial lag operator in nonlinear models: fast computation and a closed formula approximation", has received the award for the "Young Researcher Paper Competition", by the Spatial Econometrics Association, Singapore 2017. Any errors or omissions that remain are the sole responsibility of the authors.

[*]Corresponding author

Email addresses: lsantos@iseg.ulisboa.pt (Luís Silveira Santos), isabelp@iseg.ulisboa.pt (Isabel Proença)

the dependent variable can only take two possible values (say 0 and 1). Many applications can be found using such dependent variables, across distinct research areas. Examples are: the choice on participating in environmental policies (Murdoch et al., 2003, Beron et al., 2003), the adoption of new technologies in agriculture (Case, 1992, Holloway et al., 2002, Wollni and Andersson, 2014), the implementation of state income taxes (Beron and Vijverberg, 2004, Fiva and Rattsø, 2007), the location choice (Miyamoto et al., 2004, Klier and McMillen, 2008), the decision to (re)open a business (Holloway and Lapar, 2007, LeSage et al., 2011) or the existence of high crime rates (McMillen, 1992).

Besides the interest on the effects of the explanatory variables over the binary dependent variable, measuring the strength of the spatial dependence is crucial. However, the estimation of spatial binary choice models is known to be complex and computationally burdensome, specially for large data sets (Anselin, 2007). In order to deal with the estimation of spatial binary choice models, several methods have been proposed. These methods can be grouped into three main families of estimators: Maximum Likelihood (ML) based methods – the EM algorithm (McMillen, 1992), the RIS simulator (Beron and Vijverberg, 2004), partial ML estimation based on pairwise correlations (Bhat, 2011, Wang et al., 2013), the GHK simulator (Pace and LeSage, 2016) and the Mendel-Elston approximation (Martinetti and Geniaux, 2017) –, Bayesian based methods – the Gibbs sampler (LeSage, 2000) and Markov Chains Monte Carlo (Smith and LeSage, 2004) – and Generalized Method of Moments (GMM) estimators (Pinkse and Slade, 1998, Klier and McMillen, 2008). On one hand, Maximum Likelihood methods and Bayesian methods are, in general, computationally burdensome for moderate and large data sets, once they rely on assumptions about the CDF for the conditional distribution of the errors, resulting in the computation (or simulation) of a high dimensional integral. Even if the methods of Beron and Vijverberg (2004), Pace and LeSage (2016) or Martinetti and Geniaux (2017) are used to approximate the high dimensional integral, the estimation can still be time demanding and computationally infeasible, specially if the spatial units are influenced by many neighbors. On the other hand, under the GMM approach, the distributional assumptions can be relaxed, only requiring the correct specification of the moment conditions. The major drawback of the GMM approach is related to the N-dimensional matrix operations, that are also computationally infeasible for moderate and large samples. In practice, the literature is particularly scarce in terms of new methods that can accurately tackle estimation under a large sample framework.

In this paper, the estimation problem of spatial binary choice models is addressed by an approximation method that allows to reduce the computational complexity of N-dimensional matrix operations, in the context of GMM. The new approximation method explores the eigenstructure characteristics of normalized spatial weights matrices and the limiting properties of their high order powers. Focusing on spatial lag models with a binary dependent variable, the most complex matrix operation, the spatial lag operator inverse, is approximated by a sum of known matrices

and by a simple matrix-vector product. The approximation of the spatial lag operator inverse is applied to the gradients of the iterative GMM procedure of Klier and McMillen (2008), allowing to approximate other matrix operations, as well. The overall computational time and computational complexity of the estimation is drastically reduced. Moreover, a closed formula for the elements of the approximated spatial lag operator inverse matrix and for the elements of the covariance matrix is available.

The remainder of this paper is organized as follows. In section 2, the specification and estimation of spatial lag models with a binary dependent variable is reviewed. The new method to approximate the spatial lag operator inverse is presented in section 3 and used in the gradients of the iterative GMM procedure of Klier and McMillen (2008) in section 4. In section 5, a set of Monte Carlo experiments are conducted to address the statistical properties and the computational ability of the new approximation-based GMM estimator, which is also compared to other two GMM estimators frequently used in the literature. In section 6, the adequacy of the previous GMM estimators to real data is assessed and compared using an empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas. Finally, section 7 concludes. The results of the Monte Carlo experiments are summarized in appendix A, while the estimation results of the empirical application are shown in appendix B.

2. Spatially lagged latent dependent variable model for binary outcomes

A spatial binary choice model can be derived based on the following spatially lagged latent variable specification:

$$Y_i^* = \alpha \sum_{i \neq j} w_{i,j} Y_j^* + \mathbf{X}_i \boldsymbol{\beta} + \xi_i, \qquad i = 1, 2, \ldots, N \tag{1}$$

where Y_i^* is a general dependent variable (possibly not observable) for the unit i and N denotes the total number of spatial units. The coefficients $w_{i,j}$ are known non-negative scalars that refer to the spatial weights of unit j on unit i, with $j \neq i$ and $j = 1, 2, \ldots, N$. By convention, $w_{i,i} = 0$, for all i. The scalar parameter α is the spatial lag parameter. The $1 \times K$ vector $\mathbf{X}_i$ includes the observations for a set of K exogenous explanatory variables and a constant, for the unit i. The $K \times 1$ vector $\boldsymbol{\beta}$ is the corresponding vector of regression parameters. The disturbance term, ξ_i, is an i.i.d. random error for the unit i.

Stacking over the cross-sectional units, the spatial lag model can be written in matrix form and rearranged as a reduced form for the dependent variable:

$$\mathbf{Y}^* = \alpha \mathbf{W} \mathbf{Y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi} = (\mathbf{I} - \alpha \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

where $\mathbf{Y}^* = [Y_1^*, Y_2^*, \ldots, Y_N^*]^\intercal$ and $\mathbf{X} = [\mathbf{X}_1^\intercal, \mathbf{X}_2^\intercal, \ldots, \mathbf{X}_N^\intercal]^\intercal$. The error is now $\boldsymbol{\varepsilon} = (\mathbf{I} - \alpha \mathbf{W})^{-1} \boldsymbol{\xi}$, where $(\mathbf{I} - \alpha \mathbf{W})^{-1}$ is the spatial lag operator inverse and $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_N]^\intercal$. The matrix $\mathbf{I}$ is

the $N \times N$ identity matrix and the matrix $\mathbf{W}$ is the $N \times N$ spatial weights matrix, with generic element $w_{i,j}$. It is assumed that the row and column sums of $\mathbf{W}$ are uniformly bounded in absolute value and that $|\alpha|$ is also bounded. For a non-normalized $\mathbf{W}$, the parameter space of α is set to $-1/|\lambda|_{\max} < \alpha < 1/|\lambda|_{\max}$, where $|\lambda|_{\max}$ is the largest absolute eigenvalue of $\mathbf{W}$. For the case where $\mathbf{W}$ is normalized in such a way that the rows, columns or both rows and columns sum up to one, Kelejian and Robinson (1995) suggest to set the parameter space of α to $-1 < \alpha < 1$, similar to time-series frameworks. In addition, the matrix $(\mathbf{I} - \alpha\mathbf{W})$ is non-singular for all α in the parameter space.

If Y_i^* is observable, the conditional expectation is given by $\mathrm{E}(Y_i^* \,|\, \mathbf{X}, \mathbf{W}) = \mathbf{X}_i^{\#}\boldsymbol{\beta}$, where $\mathbf{X}_i^{\#}$ is the ith row of the matrix product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{X}$, and equation (2) defines a linear spatial lag model. Here, however, Y_i^* is not observable. The observed dependent variable is Y_i, a function of particular characteristics of Y_i^*, such as $Y_i = 1$ if $Y_i^* \geq 0$ and $Y_i = 0$ if $Y_i^* < 0$ and Y_i is a binary dependent variable. The conditional expectation of a spatial lag model when Y_i^* is not observable and Y_i is a binary dependent variable is given by:

$$\begin{aligned} &\mathrm{E}(Y_i \,|\, \mathbf{X}, \mathbf{W}) = \mathrm{P}(Y_i = 1 \,|\, \mathbf{X}, \mathbf{W}) = \mathrm{P}(Y_i^* > 0 \,|\, \mathbf{X}, \mathbf{W}) \\ &= \mathrm{P}\left(\mathbf{X}_i^{\#}\boldsymbol{\beta} + \varepsilon_i > 0 \,\middle|\, \mathbf{X}, \mathbf{W}\right) = \mathrm{P}\left(\varepsilon_i > -\mathbf{X}_i^{\#}\boldsymbol{\beta} \,\middle|\, \mathbf{X}, \mathbf{W}\right) \\ &= 1 - \mathrm{P}\left(\varepsilon_i \leq -\mathbf{X}_i^{\#}\boldsymbol{\beta} \,\middle|\, \mathbf{X}, \mathbf{W}\right) = G\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right), \qquad i = 1, 2, \ldots, N \end{aligned} \tag{3}$$

where $G(\eta)$ is a function that takes on values in the interval $0 < G(\eta) < 1$, for all $\eta \in \mathbb{R}$, and it is twice continuously differentiable, for all $\eta \in \mathbb{R}$ as well. Usually $G(\eta)$ is called the link function and η is called the index. It is further assumed that $G(\eta)$ is known[1] and given by the cumulative distribution function (CDF) of ξ_i conditional on $(\mathbf{X}, \mathbf{W})$. The parameter σ_i is the square root of the conditional variance of ε_i, for each i, obtained from the diagonal elements of the conditional covariance matrix of $\boldsymbol{\varepsilon}$:

$$\mathrm{Var}(\boldsymbol{\varepsilon} \,|\, \mathbf{X}, \mathbf{W}) = \left[(\mathbf{I} - \alpha\mathbf{W})^{\intercal}(\mathbf{I} - \alpha\mathbf{W})\right]^{-1} \mathrm{Var}(\boldsymbol{\xi} \,|\, \mathbf{X}, \mathbf{W}) = \boldsymbol{\Sigma} \tag{4}$$

Under this framework, the conditional variance of $\boldsymbol{\xi}$ is fixed, to ensure identification. Further, $\sigma_i^2 > 0$, for all i, and the row and column sums of $(\mathbf{I} - \alpha\mathbf{W})^{-1}$ are uniformly bounded in absolute value. This guarantees that σ_i^2 is also finite.

In most of the applications using binary response models, the conditional distribution of ξ_i is assumed to be the standard Normal distribution, $G(\eta) = \Phi(\eta)$, or the standard Logistic distribution, $G(\eta) = \Lambda(\eta)$. This implies that $\mathrm{Var}(\xi_i \,|\, \mathbf{X}, \mathbf{W}) = 1$, and $\mathrm{Var}(\xi_i \,|\, \mathbf{X}, \mathbf{W}) = \pi^2/3$, for all i, respectively. In both cases, the probability distribution functions (PDFs) of the link functions are

[1] Generally the link function, $G(\eta)$, is unknown and can be estimated by nonparametric and semiparametric methods, see Härdle et al. (2004) and Horowitz (2009) for details.

symmetric about zero and $1 - G(-\eta) = G(\eta)$, but this is not generally the case for other possible links[2]. Under these assumptions, the spatial lag model with a binary dependent variable follows as:

$$Y_i = G\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right) + u_i, \qquad i = 1, 2, \ldots, N \tag{5}$$

where u_i differs from ε_i because $u_i = Y_i - \mathrm{E}(Y_i \,|\, \mathbf{X}, \mathbf{W})$ and $\varepsilon_i = Y_i^* - \mathrm{E}(Y_i^* \,|\, \mathbf{X}, \mathbf{W})$. Hence, u_i is a discrete random variable assuming only two values, $1 - G(\cdot)$ and $-G(\cdot)$.

With regard to estimation, is it addressed by a GMM approach based on the works of Pinkse and Slade (1998) and Klier and McMillen (2008). The first author uses GMM to estimate a Probit with spatially lagged errors, assuming that the off-diagonal information of the conditional covariance matrix of $\boldsymbol{\varepsilon}$ is neglegible. The second author uses GMM to estimate a spatial lag Logit, considering the full dependence structure of the data, and show how this procedure can be applied to the estimation of models with spatially lagged errors as well. The moment condition considered by the former authors follows as:

$$\mathrm{E}(\mathbf{Z}^{\intercal}\mathbf{u}_*) = \mathbf{0} \tag{6}$$

where $\mathbf{Z}$ is the $N \times (K + p)$ matrix of instruments, with p the number of additional instruments other than the K explanatory variables. Following Kelejian and Prucha (1998), $\mathbf{Z} = [\mathbf{X}\ \mathbf{WX}]$. The $N \times 1$ vector $\mathbf{u}_*$ are the "generalized residuals" vector (Gourieroux et al., 1987):

$$u_{*,i} = \frac{\left[Y_i - G\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)\right] g\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)}{G\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)\left[1 - G\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)\right]}, \qquad i = 1, 2, \ldots, N \tag{7}$$

where the function $g(\cdot)$ is the first derivative of $G(\cdot)$ w.r.t. the index. The GMM estimates of the parameter vector, $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \alpha)^{\intercal}$, are obtained by minimizing the objective function:

$$\mathcal{Q}(\boldsymbol{\beta}, \alpha) = \mathbf{u}_*^{\intercal}\mathbf{Z}\boldsymbol{\Xi}\mathbf{Z}^{\intercal}\mathbf{u}_* \tag{8}$$

where $\boldsymbol{\Xi}$ is a $(K + p) \times (K + p)$ symmetric positive definite matrix. Following Klier and McMillen (2008), $\boldsymbol{\Xi}$ is set to $(\mathbf{Z}^{\intercal}\mathbf{Z})^{-1}$. This way, the GMM estimator reduces to nonlinear two stages least squares (N2SLS). However, as the objective function in (8) does not have a closed formula, an iterative procedure is required to obtain the parameter estimates. The following steps are considered:

1. Assume initial values for the parameter vector $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \alpha)^{\intercal}$, $\boldsymbol{\Theta}^{(0)}$, and compute the gradients evaluated at the initial values $\boldsymbol{\Gamma}_i^{(0)} = (\partial u_{*,i}/\partial\boldsymbol{\Theta})|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(0)}}$, $i = 1, 2, \ldots, N$.

[2] See, for example, the complementary log-log and the Weibull links.

2. Regress $\boldsymbol{\Gamma}^{(0)}$ on $\mathbf{Z}$, in a similar fashion to (linear) 2SLS. Obtain $\hat{\boldsymbol{\Gamma}}^{(0)}$.

3. Construct new estimates as $\boldsymbol{\Theta}^{(1)} = \boldsymbol{\Theta}^{(0)} + \left[\left(\hat{\boldsymbol{\Gamma}}^{(0)}\right)^{\intercal}\left(\hat{\boldsymbol{\Gamma}}^{(0)}\right)\right]^{-1}\left(\hat{\boldsymbol{\Gamma}}^{(0)}\right)^{\intercal}\mathbf{u}_{*}^{(0)}$, where $\mathbf{u}_{*}^{(0)}$ are the generalized residuals evaluated at the estimates of step 0.

4. Repeat steps 1. to 3., using the estimates from the last iteration, until the algorithm converges.

The estimated asymptotic variance of the (iterative) GMM estimator is:

$$\widehat{Avar}\left(\hat{\boldsymbol{\Theta}}\right) = \left(\hat{\boldsymbol{\Gamma}}^{\intercal}\hat{\boldsymbol{\Gamma}}\right)^{-1}\left[\sum_{i=1}^{N}\hat{u}_{i}^{2}\hat{\boldsymbol{\Gamma}}_{i}^{\intercal}\hat{\boldsymbol{\Gamma}}_{i}\right]\left(\hat{\boldsymbol{\Gamma}}^{\intercal}\hat{\boldsymbol{\Gamma}}\right)^{-1} \tag{9}$$

The individual gradients for each parameter are:

$$\left(\boldsymbol{\Gamma}_{\boldsymbol{\beta}}\right)_{i} = \frac{\partial u_{*,i}}{\partial \boldsymbol{\beta}^{\intercal}} = -u_{*,i}\left(\frac{g'\left(\frac{\mathbf{X}_{i}^{\#}\boldsymbol{\beta}}{\sigma_{i}}\right)}{g\left(\frac{\mathbf{X}_{i}^{\#}\boldsymbol{\beta}}{\sigma_{i}}\right)} - u_{*,i}\right)\frac{\mathbf{X}_{i}^{\#}}{\sigma_{i}}, \qquad i = 1, 2, \ldots, N \tag{10}$$

and

$$\left(\boldsymbol{\Gamma}_{\alpha}\right)_{i} = \frac{\partial u_{*,i}}{\partial \boldsymbol{\beta}^{\intercal}} = -u_{*,i}\left(\frac{g'\left(\frac{\mathbf{X}_{i}^{\#}\boldsymbol{\beta}}{\sigma_{i}}\right)}{g\left(\frac{\mathbf{X}_{i}^{\#}\boldsymbol{\beta}}{\sigma_{i}}\right)} - u_{*,i}\right)\left[\frac{1}{\sigma_{i}}\left(\mathbf{H}_{i}\boldsymbol{\beta} - \frac{\mathbf{X}_{i}^{\#}\boldsymbol{\beta}}{2\sigma_{i}^{2}}\Upsilon_{ii}\right)\right], \qquad i = 1, 2, \ldots, N \tag{11}$$

where $g'(\cdot)$ is the first derivative of the function $g(\cdot)$ w.r.t. the index, $\mathbf{H}_{i}$ is the ith row of the matrix product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{W}\mathbf{X}^{\#}$ and Υ_{ii} is the ith element of the diagonal of the matrix:

$$\begin{aligned}
\boldsymbol{\Upsilon} &= (\mathbf{I} - \alpha\mathbf{W})^{-1}\left\{\mathbf{W}(\mathbf{I} - \alpha\mathbf{W})^{-1} + \left[\mathbf{W}(\mathbf{I} - \alpha\mathbf{W})^{-1}\right]^{\intercal}\right\}\left[(\mathbf{I} - \alpha\mathbf{W})^{-1}\right]^{\intercal} \\
&= (\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I} - \alpha\mathbf{W})^{-1}\left[(\mathbf{I} - \alpha\mathbf{W})^{-1}\right]^{\intercal} + \\
&\quad + \left\{(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I} - \alpha\mathbf{W})^{-1}\left[(\mathbf{I} - \alpha\mathbf{W})^{-1}\right]^{\intercal}\right\}^{\intercal} \\
&= \left[\mathrm{Var}(\boldsymbol{\xi} \mid \mathbf{X}, \mathbf{W})\right]^{-2} \times \left[(\mathbf{F}\boldsymbol{\Sigma}) + (\mathbf{F}\boldsymbol{\Sigma})^{\intercal}\right]
\end{aligned} \tag{12}$$

where the matrix $\mathbf{F}$ is equal to $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{W}$ and the matrix $\boldsymbol{\Sigma}$ is the conditional variance of $\boldsymbol{\varepsilon}$.

The closed forms for the gradients help to accelerate the optimization process. However, there is a computational problem, because $(\mathbf{I} - \alpha\mathbf{W})^{-1}$ has to be computed on each iteration. To solve this issue, Klier and McMillen (2008) also suggests a linearized variant of the full GMM estimator, which consists in a linearization of the model around $\alpha = 0$. This choice is obvious, once no matrices need to be inverted and none of the gradients are equal to zero. In addition, $\sigma_{i} = \sqrt{\mathrm{Var}(\xi_{i} \mid \mathbf{X}, \mathbf{W})}$, for all i, and $\mathbf{X}_{i}^{\#} = \mathbf{X}_{i}$. Under this approach, the estimation procedure is rather simple. The initial estimates of the regression parameters, $\boldsymbol{\beta}^{(0)}$, can be obtained by standard Probit or standard Logit, because the gradients no longer depend on α. Next, regress $\mathbf{u}_{*}^{(0)} + (\boldsymbol{\Gamma}_{\boldsymbol{\beta}})^{\intercal}\boldsymbol{\beta}^{(0)}$ on $\boldsymbol{\Gamma}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Gamma}_{\alpha}$ and the

corresponding coefficients are the estimated values of $\boldsymbol{\beta}$ and α. The estimated asymptotic variance of the Linearized GMM estimator is:

$$\widehat{Avar}\left(\hat{\boldsymbol{\Theta}}\right) = \left(\sum_{i=1}^{N} \hat{u}_i^2\right)\left(\hat{\boldsymbol{\Gamma}}^{\intercal}\hat{\boldsymbol{\Gamma}}\right)^{-1} \tag{13}$$

As the Linearized GMM only requires a single iteration to obtain estimates for the parameter vector, it outstands all estimation methods in terms of computational time. However, this method only provides good estimates for the spatial lag parameter if the true α is less than 0.5. Then, a middle-ground approach that allows the approximation of the computationally demanding matrix operations by simpler operations, may yield better results for all admissible values of α. In particular, the approximation of the most complex matrix operation, the spatial lag operator inverse, $(\mathbf{I} - \alpha\mathbf{W})^{-1}$, may help to reduce the overall computational burden of the iterative GMM procedure, while not penalizing consistency. The details of the new approximation method are presented in the sections that follow.

3. Approximation of the spatial lag operator inverse

With regard to the approximation of the spatial lag operator inverse, it is usual to consider the series expansion of the inverse:

$$(\mathbf{I} - \alpha\mathbf{W})^{-1} = \mathbf{I} + \alpha\mathbf{W} + \alpha^2\mathbf{W}^2 + \alpha^3\mathbf{W}^3 + \ldots = \sum_{h=0}^{\infty} \alpha^h \mathbf{W}^h \tag{14}$$

which converges absolutely for all α in the parameter space. Several authors (LeSage and Pace, 2009, Arbia, 2014, Elhorst, 2014, to name a few) suggest to approximate $(\mathbf{I} - \alpha\mathbf{W})^{-1}$ by considering a finite lower-order series. However, this approximation implies that there are as many matrix operations as the number of powers of $\mathbf{W}$ considered. Also, the approximate functional form of the elements of $(\mathbf{I} - \alpha\mathbf{W})^{-1}$ is complicated. Alternatively, if the powers $h \geq 2$ of $\mathbf{W}$ are approximated by a particular matrix, the computational complexity of this inverse can be drastically reduced and the approximate functional form can be simplified as well.

Let $\mathbf{W}$ be an $N \times N$ non-negative spatial weights matrix with both row and column sums uniformly bounded in absolute value. Assuming that $\mathbf{W}$ is row normalized and diagonalizable, the eigendecomposition is available and the series expansion of the spatial lag operator inverse can be written as:

$$\begin{aligned}
(\mathbf{I} - \alpha\mathbf{W})^{-1} &= \mathbf{I} + \alpha\mathbf{W} + \alpha^2\mathbf{W}^2 + \alpha^3\mathbf{W}^3 + \ldots \\
&= \mathbf{I} + \alpha\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} + \alpha^2\left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\right)^2 + \alpha^3\left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\right)^3 + \ldots \\
&= \mathbf{I} + \alpha\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} + \alpha^2\mathbf{V}\boldsymbol{\Lambda}^2\mathbf{V}^{-1} + \alpha^3\mathbf{V}\boldsymbol{\Lambda}^3\mathbf{V}^{-1} + \ldots \\
&= \mathbf{V}\left(\mathbf{I} - \alpha\boldsymbol{\Lambda}\right)^{-1}\mathbf{V}^{-1}
\end{aligned} \tag{15}$$

where $\mathbf{\Lambda}$ is an $N \times N$ diagonal matrix whose diagonal elements are the corresponding eigenvalues of $\mathbf{W}$. The matrix $\mathbf{V}$ is an $N \times N$ matrix whose ith column corresponds to the eigenvector associated with the ith eigenvalue of $\mathbf{W}$. By definition, the largest absolute eigenvalue of $\mathbf{W}$, $|\lambda|_{\max}$, is equal to one. Then, as $h \to \infty$:

$$\mathbf{\Lambda}^{\infty} = \lim_{h \to \infty} \mathbf{\Lambda}^{h} = \lim_{h \to \infty} \operatorname{diag}\left(1^{h}, \lambda_{2}^{h}, \lambda_{3}^{h} \ldots, \lambda_{N}^{h}\right) = \operatorname{diag}\left(1, 0, 0, \ldots, 0\right) \tag{16}$$

once the eigenvalues $\lambda_2, \lambda_3 \ldots, \lambda_N$ are, in absolute value, less than one. In addition, $|\lambda|_{\max}$ is assumed to have algebraic multiplicity equal to one[3].

Replacing the powers $h \geq 2$ of $\mathbf{\Lambda}$ by $\mathbf{\Lambda}^{\infty}$ yields:

$$\begin{aligned}
(\mathbf{I} - \alpha\mathbf{W})^{-1} &\approx \mathbf{V}\left(\mathbf{I} + \alpha\mathbf{\Lambda} + \frac{\alpha^2}{1-\alpha}\mathbf{\Lambda}^{\infty}\right)\mathbf{V}^{-1} \\
&= \mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty} \\
&= \mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\operatorname{col}\left(\mathbf{V}\right)_1 \operatorname{row}\left(\mathbf{V}^{-1}\right)_1
\end{aligned} \tag{17}$$

where $\mathbf{W}^{\infty}$ is the "long run" matrix, $\operatorname{col}\left(\mathbf{V}\right)_1$ is the first column of $\mathbf{V}$ and $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$ is the first row of $\mathbf{V}^{-1}$. By definition, the largest eigenvector of $\mathbf{W}$ associated with its largest eigenvalue is equal to the $N \times 1$ vector of ones, $\boldsymbol{\iota}$, so $\operatorname{col}\left(\mathbf{V}\right)_1 = \boldsymbol{\iota}$. But, to obtain $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$, an additional problem is posed, because the entire linear system has to be solved. Nevertheless, $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$ can be identified without additional computational burden if $\mathbf{W}$ is a function of a matrix with orthogonal eigenvectors.

Let $\mathbf{W}_0$ be an initial $N \times N$ spatial weights matrix with non-zero row sums (every spatial unit has neighbors) and let $\mathbf{D}_R$ be an $N \times N$ diagonal matrix whose diagonal elements are the row sums of $\mathbf{W}_0$. Since the row sums of $\mathbf{W}_0$ are different from zero, $\mathbf{D}_R$ is invertible. Let $\mathbf{W}$ be such that $\mathbf{W} = \mathbf{D}_R^{-1}\mathbf{W}_0$, then $\mathbf{W}$ is row normalized. Now, consider the transformation $\mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2}$, hence:

$$\mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2} = \mathbf{D}_R^{1/2}\mathbf{D}_R^{-1}\mathbf{W}_0\mathbf{D}_R^{-1/2} = \mathbf{D}_R^{-1/2}\mathbf{W}_0\mathbf{D}_R^{-1/2} = \mathbf{W}_{sim} \tag{18}$$

is similar to $\mathbf{W}$. This implies that all eigenvalues of $\mathbf{W}$ and $\mathbf{W}_{sim}$ are equal and their eigenvectors are related. For a symmetric $\mathbf{W}_{sim}$, both $\mathbf{W}_{sim}$ and $\mathbf{W}$ can be decomposed into orthogonal eigenvectors. For the case where $\mathbf{W}_{sim}$ is not symmetric, a "symmetrization" procedure is required, such that the decomposition into orthogonal eigenvectors is available. Because $\mathbf{W}$ is always non-symmetric and the similarity transformation has no impact on symmetry, it is the $\mathbf{W}_0$ that determines whether $\mathbf{W}_{sim}$ is or it is not symmetric. The outcomes of these two scenarios are discussed in detail in the following subsections.

[3] For the case where the algebraic multiplicity of $|\lambda|_{\max}$ is greater than one and $\mathbf{W}$ is not block diagonal, $\mathbf{\Lambda}^{\infty}$ can still be given by (16), but at a cost of computational accuracy.

3.1. Case 1: symmetric $\mathbf{W}_0$

For a symmetric $\mathbf{W}_0$, the matrix $\mathbf{W}_{sim}$ is also symmetric and its eigendecomposition is given by $\mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{\intercal}$, where $\mathbf{V}_{sim}$ is the eigenvectors matrix of $\mathbf{W}_{sim}$. Rewrite $\mathbf{W}_{sim}$ as a function of $\mathbf{W}$ and eigendecompose both matrices. This follows as:

$$\mathbf{W}_{sim} = \mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2} \Leftrightarrow \mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{\intercal} = \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \tag{19}$$

Because $\mathbf{W}_{sim}$ is similar to $\mathbf{W}$, the matrix $\mathbf{\Lambda}$ is equal in both sides of the equation. Then, the eigenvectors of both matrices are related as $\mathbf{V}_{sim} = \mathbf{D}_R^{1/2}\mathbf{V}$ and $\mathbf{V}_{sim}^{\intercal} = \mathbf{V}^{-1}\mathbf{D}_R^{-1/2}$. But, $\mathbf{V}_{sim}^{\intercal} = \left(\mathbf{D}_R^{1/2}\mathbf{V}\right)^{\intercal} = \mathbf{V}^{\intercal}\mathbf{D}_R^{1/2}$. Replacing $\mathbf{V}_{sim}$ in equation (19) by the previous equalities, yields:

$$\begin{aligned}\mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{\intercal} = \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} &\Leftrightarrow \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\intercal}\mathbf{D}_R^{1/2} = \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \\ &\Leftrightarrow \mathbf{D}_R^{-1/2}\left(\mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\intercal}\mathbf{D}_R^{1/2}\right)\mathbf{D}_R^{1/2} = \mathbf{W}\end{aligned} \tag{20}$$

and in the "long run":

$$\begin{aligned}\mathbf{W}^{\infty} &= \frac{1}{\|\mathbf{D}_R^{1/2}\mathrm{col}\left(\mathbf{V}\right)_1\|_2\|\mathrm{col}\left(\mathbf{V}\right)_1^{\intercal}\mathbf{D}_R^{1/2}\|_2} \times \mathbf{D}_R^{-1/2}\left(\mathbf{D}_R^{1/2}\mathrm{col}\left(\mathbf{V}\right)_1\mathrm{col}\left(\mathbf{V}\right)_1^{\intercal}\mathbf{D}_R^{1/2}\right)\mathbf{D}_R^{1/2} \\ &= \frac{1}{\left[\sqrt{\left(d_{R,1}^{1/2}\right)^2 + \left(d_{R,2}^{1/2}\right)^2 + \ldots + \left(d_{R,N}^{1/2}\right)^2}\right]^2} \times \boldsymbol{\iota}\boldsymbol{\iota}^{\intercal}\mathbf{D}_R = \left[\sum_{i=1}^{N} d_{R,i}\right]^{-1}\mathbf{J}\mathbf{D}_R\end{aligned} \tag{21}$$

where $d_{R,i}$ is the sum of the ith row of $\mathbf{W}_0$ and $\mathbf{J}$ is the $N \times N$ matrix of ones. The "long run" matrix, $\mathbf{W}^{\infty}$, is rescaled by the sum of all rows of $\mathbf{W}_0$ because $\mathbf{W}$ has an implicit decomposition into orthogonal eigenvectors. There are two major advantages related to the expression of $\mathbf{W}^{\infty}$. Firstly, the matrix product $\mathbf{J}\mathbf{D}_R$ can be simplified to a, less demanding, matrix-vector product, because $\mathbf{D}_R$ is a diagonal matrix. Secondly, each element of the matrix $\mathbf{W}^{\infty}$ have an exact closed formula given by:

$$w_{i,j}^{\infty} = \left[\sum_{i=1}^{N} d_{R,i}\right]^{-1} \times d_{R,j} \tag{22}$$

This exact closed formula implies that the rows of $\mathbf{W}^{\infty}$ are all equal. In addition, the ith element of each row vector of $\mathbf{W}^{\infty}$ is given by the sum of the ith row of $\mathbf{W}_0$, that is $\mathrm{row}\left(\mathbf{W}^{\infty}\right)_i = (d_{R,1}, d_{R,2}, \ldots, d_{R,N})$, for all i.

Using (21) in the expression for the approximation of $\left(\mathbf{I} - \alpha\mathbf{W}\right)^{-1}$ yields:

$$\left(\mathbf{I} - \alpha\mathbf{W}\right)^{-1} \approx \mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\left[\sum_{i=1}^{N} d_{R,i}\right]^{-1}\mathbf{J}\mathbf{D}_R \tag{23}$$

which still converges absolutely for all α in the parameter space, because the expression for $\mathbf{W}^{\infty}$ is exact. In addition, an approximate closed formula is also available for each element of $(\mathbf{I}-\alpha\mathbf{W})^{-1}$:

$$\left[(\mathbf{I}-\alpha\mathbf{W})^{-1}\right]_{i,j} \approx \mathbb{1}_{i=j} + \alpha \times w_{i,j} + \frac{\alpha^2}{1-\alpha} \times \left[\sum_{i=1}^{N} d_{R,i}\right]^{-1} \times d_{R,j} \tag{24}$$

where $\mathbb{1}_{i=j}$ is the indicator function that is equal to one if $i=j$ and equal to zero if $i \neq j$, for all $i,j = 1,2,\ldots,N$. The quality of this approximation is improved as fast as the powers of the eigenvalues $\lambda_2, \lambda_3, \ldots, \lambda_N$ converge to zero. In fact, this is related to the approximation method proposed by Griffith (2000), for linear models.

The validity of these results also apply to a column normalized $\mathbf{W}$ and to a doubly stochastic $\mathbf{W}$. For the first case, the results hold for $\mathbf{W}^{\intercal}$, because it is row stochastic. For the second case, once $\mathbf{W}_0$ is symmetric, the doubly stochastic $\mathbf{W}$ is also symmetric, hence it can be straightforwardly decomposed into orthogonal eigenvectors and $\mathbf{W}^{\infty} = (1/n)\,\mathbf{J}$.

3.2. Case 2: non-symmetric $\mathbf{W}_0$

For the case where $\mathbf{W}_0$ is non-symmetric, the previous result for $\mathbf{W}^{\infty}$ is not valid. To see this, rewrite $\mathbf{W}_{sim}$ as a function of $\mathbf{W}$ and eigendecompose both matrices:

$$\mathbf{W}_{sim} = \mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2} \Leftrightarrow \mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{-1} = \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \tag{25}$$

where $\mathbf{V}_{sim} = \mathbf{D}_R^{1/2}\mathbf{V}$ and $\mathbf{V}_{sim}^{-1} = \mathbf{V}^{-1}\mathbf{D}_R^{-1/2}$. Since the eigenvectors of $\mathbf{W}_{sim}$ are no longer orthogonal, $\mathbf{V}_{sim}^{-1} \neq \mathbf{V}_{sim}^{\intercal}$. Therefore, to approximate $(\mathbf{I}-\alpha\mathbf{W})^{-1}$ without additional computational burden, it is crucial to obtain an expression for $\mathbf{V}_{sim}^{-1}$ based on a symmetric matrix.

Let $\mathbf{W}_0^*$ be the "symmetrized" variant of $\mathbf{W}_0$, such that all non-zero elements of $\mathbf{W}_0$ are equal above and below the main diagonal, for all $i,j = 1,2,\ldots,N$ and $i \neq j$. In other words, $\mathbf{W}_0^*$ is constructed such that if unit j is a neighbor of unit i, then unit i is also a neighbor of unit j with equal weight. This follows as:

$$\mathbf{W}_0^* = \mathbf{W}_0 - \frac{1}{2}\left\{\mathbf{W}_0 - \mathbf{W}_0^{\intercal} - \left[(\mathbf{W}_0 - \mathbf{W}_0^{\intercal})^{\circ 2}\right]^{\circ \frac{1}{2}}\right\} = \mathbf{W}_0 + \mathbf{A} \tag{26}$$

where $\mathbf{A}$ is the $N \times N$ "symmetrization" matrix. The operators "$\circ 2$" and "$\circ \frac{1}{2}$" are element-wise operations that correspond to the Hadamard square and to the Hadamard square root, respectively. As in the previous case, a row normalized matrix and a symmetric similar matrix can be defined, based on $\mathbf{W}_0^*$. The row normalized matrix is equal to $\mathbf{W}^* = \mathbf{D}_{R^*}^{-1}\mathbf{W}_0^*$, where $\mathbf{D}_{R^*}$ is a $N \times N$ invertible diagonal matrix whose diagonal elements are the row sums of $\mathbf{W}_0^*$. The symmetric similar matrix is given by $\mathbf{W}_{sim}^* = \mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0^*\mathbf{D}_{R^*}^{-1/2}$.

Assume that $\mathbf{A}$ is close to the null matrix, $\mathbf{0}$. Then it is straightforward that $\mathbf{W}_0^* \approx \mathbf{W}_0$. In other words, if the degree of non-symmetry is small, then $\mathbf{W}_0$ is well approximated by $\mathbf{W}_0^*$. Under this assumption, the eigenvectors of $\mathbf{W}_{sim}$ can be written as an approximation of orthogonal

eigenvectors. To see this, rewrite $\mathbf{W}^*_{sim}$ as a function of $\mathbf{W}_{sim}$ and eigendecompose both matrices. This yields:

$$\begin{aligned}
\mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0^*\mathbf{D}_{R^*}^{-1/2} &\approx \mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\
&\Leftrightarrow \mathbf{W}^*_{sim} \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{D}_R^{-1/2}\mathbf{W}_0\mathbf{D}_R^{-1/2}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\
&\Leftrightarrow \mathbf{W}^*_{sim} \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{W}_{sim}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\
\Leftrightarrow \mathbf{V}^*_{sim}\mathbf{\Lambda}^*\left(\mathbf{V}^*_{sim}\right)^\intercal &\approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{-1}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}
\end{aligned} \tag{27}$$

where $\mathbf{\Lambda}^*$ is the $N \times N$ diagonal matrix whose diagonal elements are the corresponding eigenvalues of $\mathbf{W}^*_{sim}$. The matrix $\mathbf{V}^*_{sim}$ is an $N \times N$ matrix whose ith column corresponds to the eigenvector associated with the ith eigenvalue of $\mathbf{W}^*_{sim}$. Note that the eigenvalues matrices $\mathbf{\Lambda}^*$ and $\mathbf{\Lambda}$ are not equal, but because $\mathbf{W}^*_{sim}$ and $\mathbf{W}_{sim}$ are similar to the corresponding row normalized matrices and, as before, assuming that the largest absolute eigenvalue of both matrices have algebraic multiplicity equal to one, in the "long run", $\lim_{h\to\infty}\left(\mathbf{\Lambda}^*\right)^h = \lim_{h\to\infty}\mathbf{\Lambda}^h = \text{diag}\left(1,0,0,\ldots,0\right)$. Therefore, the first eigenvectors of $\mathbf{W}^*_{sim}$ and $\mathbf{W}_{sim}$ are approximately related as $\text{col}\left(\mathbf{V}^*_{sim}\right)_1 \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\text{col}\left(\mathbf{V}_{sim}\right)_1$ and $\text{col}\left(\mathbf{V}^*_{sim}\right)_1^\intercal \approx \text{row}\left(\mathbf{V}^{-1}_{sim}\right)_1\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}$.

In addition, because $\mathbf{W}^*_{sim}$ is similar to a row normalized matrix, $\mathbf{W}^*$, the first eigenvector of $\mathbf{W}^*_{sim}$ can also be written as a function of a vector of ones. Hence, consider the corresponding eigendecomposition of $\mathbf{W}^*_{sim}$ and $\mathbf{W}^*$. This yields:

$$\mathbf{W}^*_{sim} = \mathbf{D}_{R^*}^{1/2}\mathbf{W}^*\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \mathbf{V}^*_{sim}\mathbf{\Lambda}^*\left(\mathbf{V}^*_{sim}\right)^\intercal = \mathbf{D}_{R^*}^{1/2}\mathbf{V}^*\mathbf{\Lambda}^*\left(\mathbf{V}^*\right)^{-1}\mathbf{D}_{R^*}^{-1/2} \tag{28}$$

with the eigenvectors related as $\mathbf{V}^*_{sim} = \mathbf{D}_{R^*}^{1/2}\mathbf{V}^*$ and $\left(\mathbf{V}^*_{sim}\right)^\intercal = \left(\mathbf{V}^*\right)^{-1}\mathbf{D}_{R^*}^{-1/2}$. But, $\left(\mathbf{V}^*_{sim}\right)^\intercal = \left(\mathbf{V}^*\right)^\intercal\mathbf{D}_{R^*}^{1/2}$. Therefore, in the "long run", $\text{col}\left(\mathbf{V}^*_{sim}\right)_1^\intercal = \boldsymbol{\iota}^\intercal\mathbf{D}_{R^*}^{1/2}$ but also $\text{col}\left(\mathbf{V}^*_{sim}\right)_1^\intercal \approx \text{row}\left(\mathbf{V}^{-1}_{sim}\right)_1\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}$. By equating these two results, $\text{row}\left(\mathbf{V}^{-1}_{sim}\right)_1$ can be expressed as:

$$\boldsymbol{\iota}^\intercal\mathbf{D}_{R^*}^{1/2} \approx \text{row}\left(\mathbf{V}^{-1}_{sim}\right)_1\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \text{row}\left(\mathbf{V}^{-1}_{sim}\right)_1 \approx \left(\boldsymbol{\iota}^\intercal\mathbf{D}_{R^*}^{1/2}\right)\mathbf{D}_{R^*}^{1/2}\mathbf{D}_R^{-1/2} \tag{29}$$

Using this approximation in the "long run" variant of equation (25) gives:

$$\begin{aligned}
&\mathbf{V}_{sim}\mathbf{\Lambda}^\infty\mathbf{V}_{sim}^{-1} = \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}^\infty\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \Leftrightarrow \\
&\Leftrightarrow \text{col}\left(\mathbf{V}_{sim}\right)_1\text{row}\left(\mathbf{V}^{-1}_{sim}\right)_1 = \mathbf{D}_R^{1/2}\text{col}\left(\mathbf{V}\right)_1\text{row}\left(\mathbf{V}^{-1}\right)_1\mathbf{D}_R^{-1/2} \Leftrightarrow \\
&\Leftrightarrow \frac{1}{\|\mathbf{D}_R^{1/2}\boldsymbol{\iota}\|_2\|\boldsymbol{\iota}^\intercal\mathbf{D}_{R^*}^{1/2}\|_2} \times \mathbf{D}_R^{-1/2}\left(\mathbf{D}_R^{1/2}\boldsymbol{\iota}\right)\left(\boldsymbol{\iota}^\intercal\mathbf{D}_{R^*}^{1/2}\right)\mathbf{D}_{R^*}^{1/2}\mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2} \approx \mathbf{W}^\infty \Leftrightarrow \\
&\Leftrightarrow \left[\sum_{i=1}^N d_{R,i}\right]^{-1/2}\left[\sum_{i=1}^N d^*_{R,i}\right]^{-1/2} \times \mathbf{J}\mathbf{D}_{R^*} \approx \mathbf{W}^\infty
\end{aligned} \tag{30}$$

where $d^*_{R,i}$ is the sum of the ith row of $\mathbf{W}^*_0$. Here, once again, the "long run" matrix, $\mathbf{W}^\infty$, is rescaled by the geometric mean of the sum of all rows of $\mathbf{W}_0$ and $\mathbf{W}^*_0$ because $\mathbf{W}$ has an

implicit decomposition into orthogonal eigenvectors. Furthermore, the matrix product $\mathbf{J}\mathbf{D}_{R^*}$ can be simplified to a matrix-vector product, because $\mathbf{D}_{R^*}$ is a diagonal matrix. The remaining results related to the closed formula of the elements of $\mathbf{W}^{\infty}$ and to the approximation of $(\mathbf{I}-\alpha\mathbf{W})^{-1}$ are similar to those from the previous subsection.

4. Estimation under low computational complexity

As mentioned earlier, the estimation of a spatial lag model with a binary dependent variable is a prodigious computational task for moderate and large sample sizes. In the context of GMM, the major computational drawback is related to the inversion of the spatial lag operator. However, this computational issue can be addressed by the approximation method discussed in the previous section, where the spatial lag operator inverse is approximated by a sum of known matrices and a simple matrix-vector product. Hence, using this approximation, the computational time and computational complexity of an estimation procedure can be significantly reduced.

Here, estimation is addressed by a variant of the iterative GMM estimator of Klier and McMillen (2008) and the approximated variant of the spatial lag operator inverse is used in the individual gradients of the iterative procedure (see equations 10 and 11). Under this approach, it is no longer required to compute $(\mathbf{I}-\alpha\mathbf{W})^{-1}$ on each iteration. However, the computation of the gradients still pose additional computational constraints, once they rely on matrix operations, other than the matrix inverse, whose computational complexity can be equal to that of the matrix inverse. To see this, consider the matrix $\boldsymbol{\Upsilon}$, from the individual gradient of α, where $(\mathbf{I}-\alpha\mathbf{W})^{-1}$ is replaced by the approximation of the spatial lag operator inverse:

$$
\begin{aligned}
\boldsymbol{\Upsilon} &= \left[\mathrm{Var}\left(\boldsymbol{\xi}\,|\,\mathbf{X},\mathbf{W}\right)\right]^{-2}\times\left[(\mathbf{F}\boldsymbol{\Sigma})+(\mathbf{F}\boldsymbol{\Sigma})^{\intercal}\right] \\
&\approx \left(\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)\mathbf{W}\left(\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)\left(\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)^{\intercal}+ \\
&\quad\left[\left(\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)\mathbf{W}\left(\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)\left(\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)^{\intercal}\right]^{\intercal}
\end{aligned}
\tag{31}
$$

This choice is obvious, once this matrix stands out as the most elaborate matrix to be computed. Nevertheless, it can be decomposed into two other matrices, $\mathbf{F}$ and $\boldsymbol{\Sigma}$. This way, the computationally complex operations can be addressed case by case and approximated by simpler matrix-vector operations and/or element-wise operations.

The approximated expression for the matrix $\mathbf{F}$ is given by:

$$
\mathbf{F}\approx\left(\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)\mathbf{W}=\mathbf{W}+\alpha\mathbf{W}^2+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\approx\mathbf{W}+\frac{\alpha}{1-\alpha}\mathbf{W}^{\infty} \tag{32}
$$

where $\mathbf{W}^2$ is approximated by $\mathbf{W}^{\infty}$, consistent with the approach regarding the approximation of the spatial lag operator inverse.

For the matrix $\boldsymbol{\Sigma}$, the approximated expression is given by:

$$\begin{aligned}
\boldsymbol{\Sigma} &\approx \operatorname{Var}(\boldsymbol{\xi} \,|\, \mathbf{X}, \mathbf{W}) \left(\mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)\left(\mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)^{\intercal} \\
&= \operatorname{Var}(\boldsymbol{\xi} \,|\, \mathbf{X}, \mathbf{W}) \left\{\mathbf{I} + \alpha\left(\mathbf{W} + \mathbf{W}^{\intercal}\right) + \frac{\alpha^2}{1-\alpha}\left[\mathbf{W}^{\infty} + (\mathbf{W}^{\infty})^{\intercal}\right] + \alpha^2\mathbf{W}\mathbf{W}^{\intercal}\right. \\
&\quad \left. + \frac{\alpha^3}{1-\alpha}\left[\mathbf{W}(\mathbf{W}^{\infty})^{\intercal} + \mathbf{W}^{\infty}\mathbf{W}^{\intercal}\right] + \left(\frac{\alpha^2}{1-\alpha}\right)^2 \mathbf{W}^{\infty}(\mathbf{W}^{\infty})^{\intercal}\right\} \\
&\approx \operatorname{Var}(\boldsymbol{\xi} \,|\, \mathbf{X}, \mathbf{W}) \left\{\mathbf{I} + \alpha\left(\mathbf{W} + \mathbf{W}^{\intercal}\right) + \frac{\alpha^2}{1-\alpha}\left[\mathbf{W}^{\infty} + (\mathbf{W}^{\infty})^{\intercal}\right] + \alpha^2\mathbf{W}^{\infty}\right. \\
&\quad \left. + \frac{\alpha^3}{1-\alpha}\left(\left[\operatorname{col}\left(\mathbf{W}(\mathbf{W}^{\infty})^{\intercal}\right)_1 \boldsymbol{\iota}^{\intercal}\right] + \left[\operatorname{col}\left(\mathbf{W}(\mathbf{W}^{\infty})^{\intercal}\right)_1 \boldsymbol{\iota}^{\intercal}\right]^{\intercal}\right) + \left[\frac{\alpha^4}{(1-\alpha)^2}\sum_{j=1}^{N}\left(w_{1,j}^{\infty}\right)^2\right]\mathbf{J}\right\}
\end{aligned} \tag{33}$$

where $\mathbf{W}\mathbf{W}^{\intercal}$ is approximated by $\mathbf{W}^{\infty}$, in the same manner that $\mathbf{W}^2$. In fact, $\mathbf{W}\mathbf{W}^{\intercal}$ can be approximated by $\mathbf{W}^2$, even if $\mathbf{W}$ is non-symmetric. As for the matrix products $\mathbf{W}(\mathbf{W}^{\infty})^{\intercal}$ and $\mathbf{W}^{\infty}(\mathbf{W}^{\infty})^{\intercal}$, they are simplified to, less demanding, matrix-vector products, once the row vectors of $\mathbf{W}^{\infty}$ are all equal. The first matrix product is equal to the matrix expansion of the first column vector of $\mathbf{W}(\mathbf{W}^{\infty})^{\intercal}$, that is $\operatorname{col}\left(\mathbf{W}(\mathbf{W}^{\infty})^{\intercal}\right)_1 \boldsymbol{\iota}^{\intercal}$. The second matrix product is equal to the $N \times N$ matrix of ones, $\mathbf{J}$, multiplied by a scalar constant given by the sum squared elements of the first row of $\mathbf{W}^{\infty}$.

Furthermore, a closed formula is also available for the elements of the approximated variant of the conditional covariance matrix, $\boldsymbol{\Sigma}$. In particular, the closed formula for the diagonal elements is:

$$\sigma_i^2 \approx \operatorname{Var}(\xi_i \,|\, \mathbf{X}, \mathbf{W}) \left[1 + \frac{\alpha^2(3-\alpha)}{1-\alpha}w_{i,i}^{\infty} + \frac{2\alpha^3}{1-\alpha}\sum_{j=1}^{N} w_{i,j}w_{1,j}^{\infty} + \left(\frac{\alpha^2}{1-\alpha}\right)^2 \sum_{j=1}^{N}\left(w_{1,j}^{\infty}\right)^2\right] \tag{34}$$

for all $i = 1, 2, \ldots, N$. Consequently, the rescaled regressors, $\mathbf{X}_i^{\#}/\sigma_i$, and related quantities are easier to obtain.

Finally, since only the diagonal elements of $\boldsymbol{\Upsilon}$ are required for the computation of the individual gradient of α, the approximated expression for the matrix $\boldsymbol{\Upsilon}$ is given by the Hadamard product between matrix $\mathbf{F}$ and matrix $\boldsymbol{\Sigma}$. This yields:

$$\begin{aligned}
\operatorname{diag}(\boldsymbol{\Upsilon}) &= 2 \times \left[\operatorname{Var}(\xi_i \,|\, \mathbf{X}, \mathbf{W})\right]^2 \times \operatorname{diag}(\mathbf{F}\boldsymbol{\Sigma}) \\
&= 2 \times \left[\operatorname{Var}(\xi_i \,|\, \mathbf{X}, \mathbf{W})\right]^{-2} \times \operatorname{diag}\left(\sum_{j=1}^{N}\left(\mathbf{F} \circ \boldsymbol{\Sigma}^{\intercal}\right)_{i,j}\right)
\end{aligned} \tag{35}$$

where "$\circ$" is the Hadamard product operator. The approximated expressions for the individual gradients of α and $\boldsymbol{\beta}$ follow directly from the previous results.

5. Monte Carlo analysis

In this section, a set of Monte Carlo experiments are presented. The proposed iterative GMM estimator with approximated gradients (iGMMa) is compared to the estimators of Klier and McMillen (2008) – the iterative GMM estimator (iGMM) and the linearized GMM estimator (LGMM) –, in terms of bias and root mean square errors properties for small and large samples, as well as the computational time required for convergence. A variety of simulation designs are considered, with particular concern on the adequacy of these estimators to large sample frameworks.

5.1. Simulation design

The latent variable model considered in the simulations follows as:

$$\mathbf{Y}^* = (\mathbf{I} - \alpha\mathbf{W})^{-1} (\beta_0 \boldsymbol{\iota} + \beta_1 \mathbf{X}) + \boldsymbol{\varepsilon} \tag{36}$$

with $\boldsymbol{\varepsilon} = (\mathbf{I} - \alpha\mathbf{W})^{-1} \boldsymbol{\xi}$. The explanatory variable, $\mathbf{X}$, is randomly drawn, for each unit, from a $\mathcal{U}(-3, 3)$ distribution and the error $\boldsymbol{\xi}$ is randomly drawn, for each unit, from a $\mathcal{N}(0, 1)$ distribution. Following McMillen (1995), under this sampling design, the simulated model tends to produce better predictions, once the variance of $\mathbf{X}$ is much larger than the variance of $\boldsymbol{\xi}$.

Now, because Y_i^* is not observable, the observed dependent variable, Y_i is constructed as $Y_i = 1$ if $Y_i^* \geq 0$ and $Y_i = 0$ if $Y_i^* < 0$, for all i. Under this setup, the model for Y_i is a spatial lag Probit model:

$$Y_i = \Phi\left(\frac{\beta_0 x_{1i}^{\#}}{\sigma_i} + \frac{\beta_1 x_{2i}^{\#}}{\sigma_i}\right) + u_i, \qquad i = 1, 2, \ldots, N \tag{37}$$

where $x_{1i}^{\#}$ is the ith row of the matrix product $(\mathbf{I} - \alpha\mathbf{W})^{-1} \boldsymbol{\iota}$ and $x_{2i}^{\#}$ is the ith row of the matrix product $(\mathbf{I} - \alpha\mathbf{W})^{-1} \mathbf{X}$. The scalar elements σ_i are the square root diagonal elements of the conditional covariance matrix of $\boldsymbol{\varepsilon}$, equal to $[(\mathbf{I} - \alpha\mathbf{W})^{\intercal} (\mathbf{I} - \alpha\mathbf{W})]^{-1}$.

The setting for the working spatial weights matrix, $\mathbf{W}$, consists in two stages. In the first stage, the N spatial units are randomly allocated on a regular lattice and an initial spatial weights matrix, $\mathbf{W}_0$, is generated, based on a nearest neighbor structure. In the second stage, $\mathbf{W}$ is obtained from the row normalization of $\mathbf{W}_0$. Furthermore, the number of nearest neighbors is not fixed. The number of nearest neighbors is given by δN, where δ is the matrix density (the complement of sparsity), the proportion of non-zero elements in $\mathbf{W}$. This way, the large sample properties can be addressed according to the spatial statistics definitions of increasing-domain asymptotics and infill asymptotics (Cressie, 2015). The former corresponds to a sampling scenario where new spatial units are added to the edges of the lattice, but the number of neighbors, for each spatial unit, remains the same as $N \to \infty$. The latter corresponds to a scenario where new observations are added between the existing ones and a bounded area tends to get denser, hence $\delta \to 1$ as $N \to \infty$ (Anselin, 2007).

For the Monte Carlo experiments, the number of spatial units, N, vary over the set $\{100, 1000, 2000\}$. The spatial lag parameter takes on values $\alpha \in \{0, 0.2, 0.5, 0.8\}$, whereas the regression parameters are held fixed at $\beta_0 = 0$ and $\beta_1 = 1$. The matrix of instruments considered for the three GMM estimators (iGMMa, iGMM and LGMM) is $\mathbf{Z} = [\mathbf{X}\ \mathbf{WX}]$, following Kelejian and Prucha (1998). The matrix density coefficient, δ, vary over the restricted set $\{0.01, 0.1, 0.2\}$, due to a consistency condition introduced by Lee (2004), where the number of neighbors for each spatial unit cannot diverge to infinity at a rate equal or faster than the rate of the sample size (Elhorst, 2014). This yields a total of 108 experiments. For each experiment, 500 replications are used. The experiments are performed in a computer working in a quad-core Intel based processor with 3.2 GHz and 8 gigabytes of RAM, operating in a 64-bit system.

For each GMM estimator, the estimates of the regression parameters, $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\alpha}$ are reported, as well as the number of iterations and the computational time. For each set of experiments and estimator, the parameter estimates are summarized by the mean and root mean squared error (RMSE), while the computational indicators are summarized by the mean. The calculations are performed using McMillen's R package McSpatial.

5.2. Results

The results of the Monte Carlo experiments are presented in appendix A. In Table .1, the results are summarized for the case where the true value of $\alpha = 0$. Similarly, the remaining results are summarized in Table .2, Table .3 and Table .4, for the cases where the true value of $\alpha = 0.2$, $\alpha = 0.5$ and $\alpha = 0.8$, respectively.

In general, regarding the estimation of α, the iterative GMM estimators (iGMMa and iGMM) perform reasonably well, while the linearized GMM estimator exhibits an upward bias for $\alpha = 0.8$, consistent with the results of Klier and McMillen (2008). Similarly, the iterative GMM estimators produce extremely accurate estimates for β_0 and β_1, while the linearized GMM estimator appears to be missing the point, specially in terms of the ability to estimate β_1. The small sample performance of the three GMM estimators is quite cumbersome, particularly when the density of the spatial weights matrix, δ, is equal to 0.2. Also, for $\delta = 0.2$, the estimates for α exhibit a persistent downward bias, even for $N = 2000$. This unveils the severe weaknesses of the spatial estimators under infill asymptotics, as demonstrated by Lahiri (1996) and, more recently, by Lee (2004). Still, consistency can be achieved under increasing domain asymptotics. In fact, for a moderate δ ($\delta = 0.1$), the proposed iGMMa estimator appears to outperform the remaining two in terms of bias, as N becomes larger.

The RMSEs of β_0 and β_1 steadily decrease, for a fixed δ, as N grows. However, when N is fixed and δ is growing, the RMSEs of β_0 and β_1 steadily increase, which illustrates, once again, how the estimates can be distorted under infill asymptotics. With regard to the RMSE of α, it also increases as both δ and N are increasing, as well.

In terms of the computational ability of the three estimators, all of them experience an increase in the overall computational time, as N grows and the spatial weights matrix becomes denser. Moreover, depending on the value of α, the estimators may struggle to converge. In particular, as α approaches unity, the inverse of the spatial lag operator is near singularity and the computation of the gradients becomes troublesome, once they start to diverge. Nevertheless, the simulation results shows that the iGMMa estimator can reduce the overall computational time of a spatial Probit estimation by about 8 times, when compared to the iGMM estimator. Even though the iGMMa estimator requires 1 or 2 iterations more, on average, to converge, it clearly does better than the iGMM estimator. In addition, while being practically impossible to overcome the speed of the linearized GMM estimator, the potential of iGMMa estimator is proven in terms of estimation accuracy amongst the three GMM estimators, even for values of α close to unity.

6. Empirical application

In this section, an empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas is presented to assess and compare the adequacy of the previous GMM estimators to real data.

Nowadays, the promotion of competitiveness is one of the main concerns for policy makers. Nevertheless, a clear definition for competitiveness is far from being consensual. In the words of Porter (1990), competitiveness is more than bilateral comparisons, it is related to the ability of the industries to innovate. Fagerberg (1988) defines competitiveness as the growth in relative unit labor costs (the cost of labour per units of output) and, eight years later, the same author considers that competitiveness can be addressed by the growth of GDP per capita or the change in research and development as a percentage of GDP (Fagerberg, 1996). More recently, in a report from the World Economic Forum, Schwab and Sala-i Martin (2010) defined 12 pillars for competitiveness, based on institutional background, physical infrastructures, macroeconomic environment, efficiency and innovation. Hence, in a broad sense, competitiveness can be considered as a measure of economic performance. Furthermore, the effects of promoting economic performance cannot be dissociated from environmental impacts. In fact, Porter et al. (2015) points out that while promoting efficient energy infrastructures and a low-carbon transition, the competitiveness may also improve. In Economics theory, these effects are addressed by the Environmental Kuznets Curve (EKC) hypothesis, originated by the works of Grossman and Krueger (1991), Shafik and Bandyopadhyay (1992) and Panayotou (1993). The EKC hypothesis states that there is an inverted "U" shaped relationship between environmental degradation and economic growth. Though economic performance is a wider concept than economic growth, the work of Porter et al. (2015) establishes the evidence that competitiveness may be affected by environmental quality, leading to an inversion of the EKC hypothesis. However, empirically, this relationship is yet to be tested. Most of the applications focus on the analysis of competitiveness and environmental quality as separate subjects and only

few works consider this analysis under a spatial framework – Rice et al. (2006) and Dudensing and Barkley (2010) on the spatial spillovers of regional competitiveness and Millimet et al. (2003) and Rupasingha et al. (2004) on the shape of EKC hypothesis and the spatial spillovers of emitting air pollutants –. In addition, none of the previous works estimate a spatial model with binary dependent variables.

In this application, the competitiveness in the U.S. MSAs is addressed using combined socioeconomic data and environmental data, collected from the U.S. Bureau of Economic Analysis (BEA) and the U.S. Environmental Protection Agency (EPA), respectively. The data set contains information on the GDP, labor costs, price index, dividends, total employment and population, as well as, information on the annualized Air Quality Index (AQI) and on three of the most representative pollutants – ground-level ozone (O_3), particle pollutants ($PM_{2.5}$) and nitrogen dioxide (NO_2) – . The MSAs included in this analysis are the continental MSAs that continuously report information for the previous variables, between 2001 and 2012 ($N = 3756$).

As mentioned earlier, there are numerous ways to define competitiveness. Hence, it is difficult to provide a clear interpretation or to have a precise unit of measurement. In practice, competitiveness can be considered a latent variable. Nevertheless, since there are so many proxies to measure competitiveness, they can be used as criteria for a new indicator. In fact, a binary competitiveness indicator can be derived such that a given MSA is considered to be competitive if, simultaneously, (1) its employment-to-population ratio is greater than or equal than the employment-to-population ratio in the combined area of the excluded MSAs and non-MSAs; (2) its GDP per capita is greater than or equal than this the GDP per capita in the combined area of the excluded MSAs and non-MSAs; (3) its Unit Labor Costs (the cost of labor per unit of output) are greater than or equal than the Unit Labor Costs in the combined area of the excluded MSAs and non-MSAs or the Unit Capital Costs (the cost of capital per unit of output) are greater than or equal than the Unit Labor Costs in the combined area of the excluded MSAs and non-MSAs, depending on whether the labor intensity ratio (the cost of labor to the cost of capital) is greater than or less than 1, respectively.

The descriptive statistics of variables included in this analysis are presented in Table .5. Considering the binary competitiveness indicator (Y), about 20% of the MSAs are labeled competitive. The variables AQI_{min} and AQI_{max} are, respectively, the minimum and maximum annual values observed for the AQI, and, as expected, AQI_{min} exhibits a low variability pattern, contrarily to AQI_{max}, that is influenced by the existence of severe outliers. The variables % days O_3, % days $PM_{2.5}$ and % days NO_2 represent, on average, a total of about 90% of days, in a year, that the observed value of the daily AQI was determined by the concentration levels of the respective pollutants.

A spatial lag Probit is applied to the pooled sample of MSAs in order to study the effects of the environmental quality indicators over the binary competitiveness and to address the intensity of spatial dependence. The spatial weights matrix $\mathbf{W}$ is considered to be block-diagonal and

given by a binary radial matrix with a distance threshold equal to 1, according to the pattern of proximity displayed in Figure .1 by the included MSAs. Under this setting, only the closest MSAs are considered to be neighbors, in order to avoid spurious interactions. The matrix of instruments is, once again, given by $\mathbf{Z} = [\mathbf{X}\ \mathbf{WX}]$. Estimation results for each of the three GMM estimators (iGMMa, iGMM and LGMM) were obtained using McMillen's R package McSpatial[4] and are shown in Table .6. The sign, magnitude and statistical significance of the estimates are coherent for the iterative GMM estimators, as opposed to the linearized GMM estimator. Even though the Probit specification is the same for the three estimators, the linearized GMM estimator appears to be missing the point. Not surprisingly, the spatial lag parameter is the most unstable estimate, ranging from $\hat{\alpha} = 0.276$ to $\hat{\alpha} = 0.578$. Still, the positive sign and statistical significance indicates that there is a positive moderate spatial effect of the neighboring areas over probability of a given MSA to be competitive. Furthermore, the signs for the estimates of AQI_{min} and AQI_{max} are particularly interesting. A unitary change in low values of the AQI has a positive impact over the probability of a given MSA to be competitive, contrarily to unitary changes in high values of the AQI. This follows along the lines of Porter et al. (2015), where the adoption of cleaner-energy and a lower-carbon transition may lead to a "win-win path", such that environmental quality comes at a "cost" of promoting new competitive areas. The global validity of these results is confirmed by the Hansen test, even though it is severely distorted for the linearized GMM estimator, due the misleading estimate of α.

The percentage of observations that are correctly predicted can be used as a measure for the overall adequacy of the GMM estimators. The iGMMa estimator correctly predicts 52.8% of the observations, compared to 45.4% for the iGMM estimator and 18.2% for the Linearized GMM estimator. Hence, the proposed iGMMa estimator performs better than the benchmark estimator (iGMM), in terms of predictive power.

In terms of computational performance, the iGMMa estimator largely outruns the iGMM estimator, only requiring 16 iterations to converge, compared to the 51 needed by the iGMM estimator to converge. Once again, the iGMMa estimator proves to be a feasible and an adequate alternative to estimate spatial binary choice models.

7. Conclusion

In this paper the estimation of spatial binary choice models was addresed. These models became particularly popular in applications that dealt with spatially correlated dependent variables reflecting a dichotomous phenomena. Examples of such dependent variables are: the implementation of a policy, the decision between two opposing alternatives or a binary status related to an unobserved variable. Under this framework, the observed outcome of given unit is determined by

[4] The programs used are available upon request.

the observed outcomes of neighboring units. Hence if spatial dependence is deliberately omitted from the estimation, the resulting estimates are likely to be inconsistent (McMillen, 1992). However, the estimation of spatial binary choice models is known to be a prodigious computational task. Either because it is required to compute a high dimensional integral or it is required to compute complex N dimensional matrix operations. To solve the issue of high dimensional integration, several approximation methods were proposed (Beron and Vijverberg, 2004, Pace and LeSage, 2016, Martinetti and Geniaux, 2017), but the computational difficulties remained, specially for highly connected spatial data. As for the complex N dimensional matrix operations issue, the solution involved a linearization of the model around an initial value of the spatial parameter, such that these matrix operations could be avoided (Klier and McMillen, 2008), but consistency turned out to be a major concern.

Focusing on spatially lagged models with a binary dependent variable, a simple and intuitive method to approximate the complex N dimensional matrix operations was presented. Specifically, the most complex matrix operation, the spatial lag operator inverse, was approximated by a sum of known matrices and a simple matrix-vector product, based on the eigenstructure characteristics of normalized spatial weights matrices and the limiting properties of their high order powers. The approximated variant of the spatial lag operator inverse was then applied to the gradients of the iterative GMM procedure of Klier and McMillen (2008), allowing to simplify the computation of the gradients and to reduce the overall computational complexity of the estimation procedure. In addition, the approximation method allowed to obtain closed formulas for the elements of the approximated matrices.

In a Monte Carlo analysis, the small and large sample properties and the computational ability of the new estimator – the iterative GMM with approximated gradients (iGMMa) –, were assessed and compared to the GMM estimators of Klier and McMillen (2008) – the iterative GMM (iGMM) and the linearized GMM (LGMM) –. The results showed that both iGMMa and iGMM estimators performed well in terms of estimation of the parameters, accurately recovering their true value, as opposed to LGMM. The iGMMa estimator outperformed the iGMM estimator in terms of computational time, specially when N was large. Surprisingly, the iGMMa proved to be extremely valuable for the case where the true value of the spatial lag parameter was close to unity, with a moderate number of neighbors in the spatial weights matrix, only exhibiting a small downward bias for the estimate of the spatial lag parameter. These results were confirmed in an empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas, where the iGMMa estimator outperformed the remaining two GMM estimators in terms of predictive power and the computational ability of handling large data sets.

As a consequence of the impressive performance of the proposed iGMMa estimator, immediate extensions are available, specially for models with spatially lagged errors and for models with further lags of the dependent variable, as well as for spatial models with discrete and cen-

sored variables. In addition, the approximated variant of the spatial lag operator inverse can be straightforwardly applied to spatial linear models, addressing, in particular, the computation of the log-determinant.

All the functions used in this paper and the proposed approximation method can be easily implemented using McMillen's R package McSpatial.

Appendix A

Table .1: Performance of the iterative GMM estimator with approximated gradients compared to other GMM estimators for the spatial Probit model with $\alpha = 0$

δ		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.001	0.001	0.001	0.083	0.040	0.060	0.132	0.065	0.096
		(0.110)	(0.109)	(0.081)	(0.331)	(0.280)	(0.236)	(0.413)	(0.340)	(0.288)
	$\hat{\beta}_0$	−0.017	−0.018	−0.008	−0.008	−0.008	−0.004	−0.002	0.002	0.001
		(0.200)	(0.200)	(0.131)	(0.217)	(0.210)	(0.149)	(0.411)	(0.202)	(0.157)
	$\hat{\beta}_1$	1.097	1.103	0.642	1.104	1.106	0.631	1.093	1.083	0.617
		(0.222)	(0.226)	(0.386)	(0.241)	(0.238)	(0.393)	(0.214)	(0.208)	(0.404)
	Time:									
	Loop	0.088	0.079		0.108	0.105		0.113	0.102	
	#iter	6	6		6	6		7	6	
	Total	0.577	0.532	0.058	0.781	0.700	0.052	0.853	0.686	0.054
1000	$\hat{\alpha}$	−0.003	−0.009	0.001	0.084	0.028	0.003	0.143	0.064	−0.003
		(0.110)	(0.110)	(0.073)	(0.338)	(0.283)	(0.250)	(0.417)	(0.337)	(0.409)
	$\hat{\beta}_0$	−0.003	−0.003	0.000	0.004	0.003	0.002	0.003	0.003	0.000
		(0.063)	(0.063)	(0.036)	(0.063)	(0.061)	(0.051)	(0.071)	(0.062)	(0.068)
	$\hat{\beta}_1$	1.009	1.010	0.547	1.010	1.009	0.547	1.010	1.007	0.547
		(0.056)	(0.057)	(0.454)	(0.056)	(0.056)	(0.454)	(0.056)	(0.053)	(0.454)
	Time:									
	Loop	1.640	3.755		1.413	4.622		1.475	6.530	
	#iter	6	6		7	6		7	6	
	Total	9.934	22.586	0.057	9.597	27.765	0.061	10.253	39.221	0.070
2000	$\hat{\alpha}$	0.000	−0.006	−0.002	0.084	0.032	0.004	0.121	−0.194	−0.012
		(0.104)	(0.105)	(0.076)	(0.321)	(0.271)	(0.243)	(0.398)	(0.735)	(0.434)
	$\hat{\beta}_0$	−0.001	−0.001	0.000	−0.001	−0.001	−0.002	−0.003	−0.002	0.000
		(0.044)	(0.043)	(0.025)	(0.043)	(0.042)	(0.033)	(0.048)	(0.065)	(0.054)
	$\hat{\beta}_1$	1.003	1.003	0.541	1.003	1.003	0.541	1.007	1.007	0.542
		(0.038)	(0.038)	(0.459)	(0.038)	(0.038)	(0.459)	(0.041)	(0.041)	(0.458)
	Time:									
	Loop	6.069	25.667		5.945	27.860		4.441	41.777	
	#iter	6	6		7	6		7	6	
	Total	36.531	154.046	0.055	39.643	167.207	0.100	30.749	251.474	0.150

NOTE: Numbers in brackets are root mean square errors (RMSE). Computational time in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table .2: Performance of the iterative GMM estimator with approximated gradients compared to other GMM estimators for the spatial Probit model with $\alpha = 0.2$

δ		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.200	0.193	0.142	0.211	0.137	0.144	0.223	0.135	0.151
		(0.116)	(0.108)	(0.102)	(0.349)	(0.286)	(0.251)	(0.399)	(0.337)	(0.287)
	$\hat{\beta}_0$	−0.018	−0.018	−0.009	−0.015	−0.013	−0.010	−0.015	−0.011	−0.001
		(0.185)	(0.189)	(0.116)	(0.213)	(0.206)	(0.150)	(0.240)	(0.211)	(0.161)
	$\hat{\beta}_1$	1.079	1.115	0.681	1.118	1.110	0.647	1.104	1.084	0.619
		(0.232)	(0.257)	(0.357)	(0.234)	(0.247)	(0.385)	(0.236)	(0.218)	(0.402)
	Time:									
	Loop	0.108	0.083		0.114	0.109		0.109	0.105	
	#iter	6	6		7	6		7	6	
	Total	0.703	0.547	0.056	0.833	0.724	0.045	0.834	0.697	0.049
1000	$\hat{\alpha}$	0.223	0.193	0.148	0.243	0.148	0.151	0.251	0.142	0.174
		(0.116)	(0.091)	(0.091)	(0.347)	(0.273)	(0.273)	(0.428)	(0.340)	(0.417)
	$\hat{\beta}_0$	0.000	0.000	0.000	0.000	0.000	−0.003	0.005	0.003	0.002
		(0.053)	(0.053)	(0.026)	(0.058)	(0.057)	(0.049)	(0.069)	(0.065)	(0.072)
	$\hat{\beta}_1$	1.009	1.010	0.547	1.010	1.009	0.547	1.010	1.007	0.547
		(0.057)	(0.058)	(0.447)	(0.053)	(0.053)	(0.454)	(0.055)	(0.059)	(0.454)
	Time:									
	Loop	1.511	3.449		1.408	4.975		1.477	6.578	
	#iter	6	6		7	6		7	6	
	Total	9.196	20.724	0.057	9.650	29.917	0.052	10.516	39.557	0.061
2000	$\hat{\alpha}$	0.225	0.191	0.153	0.273	0.173	0.139	0.220	0.030	0.190
		(0.118)	(0.090)	(0.093)	(0.343)	(0.262)	(0.263)	(0.426)	(0.668)	(0.429)
	$\hat{\beta}_0$	−0.001	0.000	−0.001	0.002	0.002	0.001	0.004	−0.002	0.001
		(0.037)	(0.037)	(0.019)	(0.043)	(0.040)	(0.037)	(0.056)	(0.065)	(0.052)
	$\hat{\beta}_1$	1.001	1.002	0.549	1.007	1.006	0.543	1.007	1.011	0.543
		(0.038)	(0.038)	(0.452)	(0.038)	(0.038)	(0.458)	(0.038)	(0.079)	(0.457)
	Time:									
	Loop	5.535	24.863		5.418	34.427		4.449	41.755	
	#iter	6	6		7	6		7	6	
	Total	33.733	149.220	0.064	37.795	206.701	0.095	31.544	252.729	0.134

NOTE: Numbers in brackets are root mean square errors (RMSE). Computational time in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table .3: Performance of the iterative GMM estimator with approximated gradients compared to other GMM estimators for the spatial Probit model with $\alpha = 0.5$

δ		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.508	0.459	0.360	0.506	0.344	0.334	0.375	0.255	0.253
		(0.098)	(0.084)	(0.158)	(0.322)	(0.277)	(0.287)	(0.397)	(0.396)	(0.376)
	$\hat{\beta}_0$	0.002	0.002	0.003	−0.016	−0.010	−0.007	−0.021	−0.024	−0.006
		(0.166)	(0.178)	(0.100)	(0.265)	(0.240)	(0.184)	(0.283)	(0.259)	(0.207)
	$\hat{\beta}_1$	0.905	1.075	0.653	1.146	1.105	0.670	1.094	1.080	0.620
		(0.213)	(0.265)	(0.365)	(0.312)	(0.258)	(0.368)	(0.219)	(0.219)	(0.401)
	Time:									
	Loop	0.115	0.081		0.108	0.105		0.117	0.102	
	#iter	7	6		7	6		7	6	
	Total	0.930	0.556	0.047	0.858	0.698	0.049	0.917	0.686	0.050
1000	$\hat{\alpha}$	0.713	0.492	0.481	0.521	0.349	0.500	0.434	0.280	0.488
		(0.242)	(0.057)	(0.088)	(0.289)	(0.249)	(0.298)	(0.384)	(0.358)	(0.432)
	$\hat{\beta}_0$	0.003	0.002	0.001	−0.007	−0.007	0.000	−0.001	−0.001	−0.003
		(0.047)	(0.042)	(0.062)	(0.066)	(0.063)	(0.111)	(0.075)	(0.074)	(0.132)
	$\hat{\beta}_1$	0.988	1.008	0.614	1.013	1.009	0.558	1.015	1.008	0.552
		(0.061)	(0.060)	(0.388)	(0.059)	(0.055)	(0.444)	(0.065)	(0.058)	(0.449)
	Time:									
	Loop	1.400	3.663		1.429	4.545		1.474	6.850	
	#iter	7	6		7	6		7	6	
	Total	10.566	21.944	0.058	10.674	27.340	0.068	10.887	41.240	0.059
2000	$\hat{\alpha}$	0.726	0.489	0.496	0.553	0.381	0.502	0.436	0.375	0.465
		(0.250)	(0.054)	(0.084)	(0.284)	(0.226)	(0.292)	(0.411)	(0.624)	(0.458)
	$\hat{\beta}_0$	0.001	0.001	0.002	−0.001	−0.001	0.000	−0.004	−0.001	0.003
		(0.033)	(0.030)	(0.051)	(0.043)	(0.041)	(0.077)	(0.051)	(0.084)	(0.087)
	$\hat{\beta}_1$	0.994	1.004	0.586	1.007	1.004	0.548	1.006	1.027	0.546
		(0.042)	(0.041)	(0.415)	(0.042)	(0.039)	(0.453)	(0.057)	(0.111)	(0.454)
	Time:									
	Loop	5.834	25.038		4.852	35.765		4.475	41.923	
	#iter	8	6		7	6		7	6	
	Total	46.646	150.268	0.079	36.788	215.525	0.094	33.050	260.884	0.123

NOTE: Numbers in brackets are root mean square errors (RMSE). Computational time in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table .4: Performance of the iterative GMM estimator with approximated gradients compared to other GMM estimators for the spatial Probit model with $\alpha = 0.8$

δ		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\hat{\alpha}$	0.661	0.576	0.464	0.582	0.413	0.398	0.367	0.259	0.234
		(0.259)	(0.246)	(0.358)	(0.388)	(0.463)	(0.470)	(0.586)	(0.622)	(0.645)
	$\hat{\beta}_0$	0.163	0.149	0.010	−0.001	−0.017	0.027	0.041	0.029	0.006
		(0.257)	(0.249)	(0.070)	(0.529)	(0.469)	(0.295)	(0.812)	(0.711)	(0.464)
	$\hat{\beta}_1$	0.551	0.736	0.420	0.940	0.911	0.581	1.020	1.019	0.595
		(0.473)	(0.320)	(0.583)	(0.228)	(0.223)	(0.444)	(0.195)	(0.226)	(0.423)
	Time:									
	Loop	0.123	0.097		0.108	0.108		0.118	0.107	
	#iter	8	6		8	6		7	6	
	Total	1.102	0.689	0.039	0.893	0.718	0.045	0.953	0.734	0.051
1000	$\hat{\alpha}$	1.266	0.711	1.246	0.660	0.406	1.278	0.458	0.302	1.258
		(0.467)	(0.091)	(0.459)	(0.401)	(0.484)	(0.668)	(0.566)	(0.607)	(0.809)
	$\hat{\beta}_0$	0.009	0.013	−0.049	−0.004	−0.007	−0.079	0.007	0.011	0.009
		(0.037)	(0.031)	(0.170)	(0.162)	(0.155)	(0.578)	(0.204)	(0.193)	(0.618)
	$\hat{\beta}_1$	0.793	0.894	0.593	0.997	0.985	0.594	1.003	1.003	0.577
		(0.211)	(0.115)	(0.408)	(0.075)	(0.060)	(0.409)	(0.055)	(0.063)	(0.425)
	Time:									
	Loop	1.431	3.453		1.450	4.554		1.468	6.845	
	#iter	9	6		8	6		7	6	
	Total	12.912	21.321	0.054	12.034	27.726	0.067	10.947	41.646	0.060
2000	$\hat{\alpha}$	1.862	0.721	1.271	0.729	0.606	1.341	0.558	0.624	1.190
		(1.071)	(0.080)	(0.479)	(0.313)	(0.299)	(0.687)	(0.553)	(0.691)	(0.858)
	$\hat{\beta}_0$	0.017	0.007	0.050	−0.012	−0.002	−0.070	−0.001	−0.035	0.069
		(0.040)	(0.038)	(0.155)	(0.129)	(0.068)	(0.477)	(0.154)	(0.203)	(0.449)
	$\hat{\beta}_1$	0.858	0.950	0.637	0.996	1.102	0.573	1.006	1.061	0.556
		(0.148)	(0.056)	(0.364)	(0.037)	(0.451)	(0.428)	(0.047)	(0.169)	(0.445)
	Time:									
	Loop	6.099	16.914		4.384	47.335		5.088	47.882	
	#iter	13	7		8	7		7	6	
	Total	81.116	122.998	0.070	35.781	319.595	0.094	38.483	294.524	0.123

NOTE: Numbers in brackets are root mean square errors (RMSE). Computational time in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Appendix B

Table .5: Descriptive statistics for the variables included in the empirical application on competitiveness

	Y	AQI_{min}	AQI_{max}	% days O_3	% days $PM_{2.5}$	% days NO_2
Mean	0.202	0.124	1.422	0.422	0.432	0.035
Std. Dev.	0.401	0.080	0.908	0.290	0.294	0.073
Min	0.000	0.000	0.380	0.000	0.000	0.000
Q_1	0.000	0.060	1.090	0.215	0.191	0.000
Median	0.000	0.110	1.330	0.385	0.398	0.000
Q_3	0.000	0.170	1.600	0.649	0.648	0.033
Max	1.000	0.430	22.120	1.000	1.000	0.653
N	3,756	3,756	3,756	3,756	3,756	3,756

Figure .1: Centroids of the U.S. Metropolitan Statistical Areas included in the empirical application on competitiveness

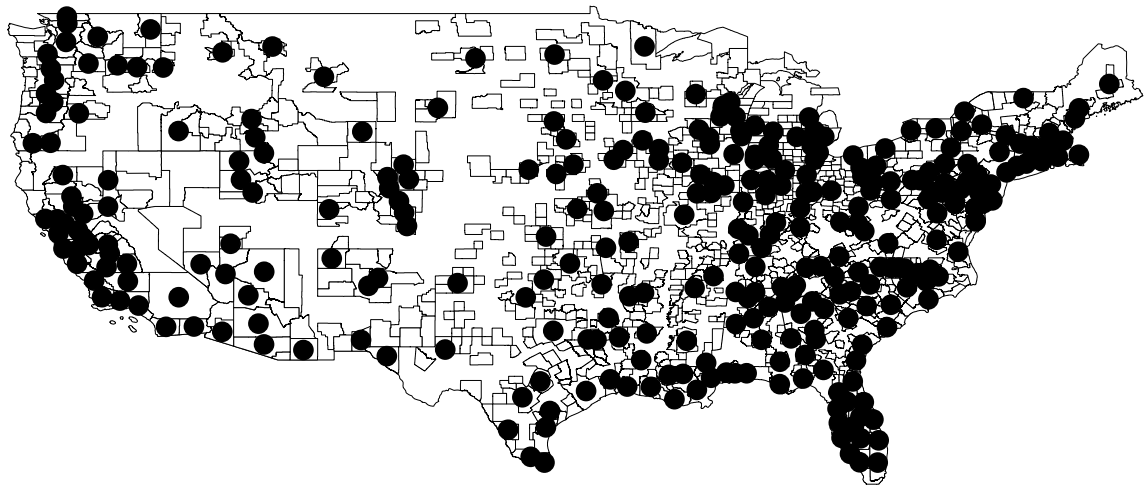

Table .6: Spatial lag Probit estimation results for the empirical application on competitiveness

	Dependent variable:		
	Y		
	(iGMMa)	(iGMM)	(LGMM)
Intercept	-0.535^{*}	-0.568^{***}	-0.389^{**}
	(0.279)	(0.168)	(0.158)
AQI_{min}	1.629^{***}	1.783^{***}	0.939^{***}
	(0.308)	(0.344)	(0.305)
AQI_{max}	-0.118^{**}	-0.137^{**}	0.017
	(0.059)	(0.066)	(0.059)
% days O_3	-0.156	-0.161	-0.156
	(0.120)	(0.132)	(0.125)
% days $PM_{2.5}$	0.052	0.049	-0.063
	(0.118)	(0.131)	(0.128)
% days NO_2	0.150	0.220	-0.090
	(0.306)	(0.323)	(0.326)
Spatial Lag	0.578^{**}	0.422^{**}	0.276^{***}
	(0.278)	(0.175)	(0.075)
Observations	3,756	3,756	3,756
# Iterations	16	51	
Hansen's J test	5.224	5.375	37.991
(p-value)	(0.485)	(0.503)	(1.000)

NOTE: Standard errors in parentheses. Significance at the 1%, 5% and 10% levels indicated by ***, ** and *, respectively.

References

Anselin, L., 2007. Spatial Econometrics. In: Baltagi, B. (Ed.), A Companion to Theoretical Econometrics. Blackwell Publishing Ltd, pp. 310–330.
URL https://dx.doi.org/10.1002/9780470996249.ch15

Arbia, G., 2014. A Primer for Spatial Econometrics: With Applications in R. Springer.

Beron, K. J., Murdoch, J. C., Vijverberg, W. P. M., 2003. Why Cooperate? Public Goods, Economic Power, and the Montreal Protocol. The Review of Economics and Statistics 85 (2), 286–297.
URL https://dx.doi.org/10.1162/003465303765299819

Beron, K. J., Vijverberg, W. P. M., 2004. Probit in a Spatial Context: A Monte Carlo Analysis. In: Luc Anselin, R. J. G. M. F., Rey, S. J. (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 169–195.
URL https://dx.doi.org/10.1007/978-3-662-05617-2_8

Bhat, C. R., 2011. The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Transportation Research Part B: Methodological 45 (7), 923–939.
URL https://dx.doi.org/10.1016/j.trb.2011.04.005

Case, A., 1992. Neighborhood influence and technological change. Regional Science and Urban Economics 22 (3), 491–508.
URL https://dx.doi.org/10.1016/0166-0462(92)90041-X

Cressie, N. A. C., 2015. Statistics for Spatial Data. Wiley, New York.

Dudensing, R. M., Barkley, D. L., 2010. Competitiveness of southern metropolitan areas: The role of new economy policies. The Review of Regional Studies 40 (2), 197–226.
URL http://journal.srsa.org/ojs/index.php/RRS/article/view/208

Elhorst, J. P., 2014. Spatial Econometrics: From Cross-Sectional Data to Spatial Panels. Springer.

Fagerberg, J., 1988. International Competitiveness. The Economic Journal 98 (391), 355–374.
URL http://www.jstor.org/stable/2233372

Fagerberg, J., 1996. Technology and competitiveness. Oxford Review of Economic Policy 12 (3), 39–51.
URL http://dx.doi.org/10.1093/oxrep/12.3.39

Fiva, J. H., Rattsø, J., 2007. Local choice of property taxation: evidence from Norway. Public Choice 132 (3), 457–470.
URL https://dx.doi.org/10.1007/s11127-007-9171-z

Gourieroux, C., Monfort, A., Renault, E., Trognon, A., 1987. Generalised residuals. Journal of Econometrics 34 (1), 5–32.
URL https://dx.doi.org/10.1016/0304-4076(87)90065-0

Griffith, D. A., 2000. Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses. Linear Algebra and its Applications 321 (1), 95–112.
URL https://dx.doi.org/10.1016/S0024-3795(00)00031-8

Grossman, G. M., Krueger, A. B., 1991. Environmental Impacts of a North American Free Trade Agreement. Working Paper 3914, National Bureau of Economic Research.
URL http://www.nber.org/papers/w3914

Härdle, W., Müller, M., Sperlich, S., Werwatz, A., 2004. Nonparametric and semiparametric models. Springer Science & Business Media.

Holloway, G., Lapar, M. L. A., 2007. How Big is Your Neighbourhood? Spatial Implications of Market Participation Among Filipino Smallholders. Journal of Agricultural Economics 58 (1), 37–60.
URL https://dx.doi.org/10.1111/j.1477-9552.2007.00077.x

Holloway, G., Shankar, B., Rahman, S., 2002. Bayesian spatial probit estimation: a primer and an application to HYV rice adoption. Agricultural Economics 27 (3), 383–402.
URL https://dx.doi.org/10.1016/S0169-5150(02)00070-1

Horowitz, J. L., 2009. Semiparametric and nonparametric methods in econometrics. Springer.

Kelejian, H. H., Prucha, I. R., 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. The Journal of Real Estate Finance and Economics 17 (1), 99–121.
URL https://dx.doi.org/10.1023/A:1007707430416

Kelejian, H. H., Robinson, D. P., 1995. Spatial Correlation: A Suggested Alternative to the Autoregressive Model. In: Anselin, L., Florax, R. J. G. M. (Eds.), New Directions in Spatial Econometrics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 75–95.
URL https://dx.doi.org/10.1007/978-3-642-79877-1_3

Klier, T., McMillen, D. P., 2008. Clustering of Auto Supplier Plants in the United States. Journal of Business & Economic Statistics 26 (4), 460–471.
URL https://dx.doi.org/10.1198/073500107000000188

Lahiri, S. N., 1996. On inconsistency of estimators based on spatial data under infill asymptotics. Sankhyā: The Indian Journal of Statistics, Series A 58 (3), 403–417.
URL http://www.jstor.org/stable/25051119

Lee, L.-F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72 (6), 1899–1925.
URL https://dx.doi.org/10.1111/j.1468-0262.2004.00558.x

LeSage, J. P., 2000. Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. Geographical Analysis 32 (1), 19–35.
URL 10.1111/j.1538-4632.2000.tb00413.x

LeSage, J. P., Kelley Pace, R., Lam, N., Campanella, R., Liu, X., 2011. New Orleans business recovery in the aftermath of Hurricane Katrina. Journal of the Royal Statistical Society: Series A 174 (4), 1007–1027.
URL https://dx.doi.org/10.1111/j.1467-985X.2011.00712.x

LeSage, J. P., Pace, R., 2009. Introduction to Spatial Econometrics. Statistics: A Series of Textbooks and Monographs. CRC Press.

Martinetti, D., Geniaux, G., 2017. Approximate likelihood estimation of spatial probit models. Regional Science and Urban Economics 64, 30–45.
URL https://dx.doi.org/10.1016/j.regsciurbeco.2017.02.002

McMillen, D. P., 1992. Probit with spatial autocorrelation. Journal of Regional Science 32 (3), 335–348.
URL https://dx.doi.org/10.1111/j.1467-9787.1992.tb00190.x

McMillen, D. P., 1995. Spatial Effects in Probit Models: A Monte Carlo Investigation. In: Anselin, L., Florax, R. J. G. M. (Eds.), New Directions in Spatial Econometrics. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 189–228.
URL https://dx.doi.org/10.1007/978-3-642-79877-1_9

McMillen, D. P., 2013. McSpatial: Nonparametric spatial data analysis. R package version 2.

Millimet, D. L., List, J. A., Stengos, T., 2003. The environmental kuznets curve: Real progress or misspecified models? The Review of Economics and Statistics 85 (4), 1038–1047.
URL https://dx.doi.org/10.1162/003465303772815916

Miyamoto, K., Vichiensan, V., Shimomura, N., Páez, A., 2004. Discrete Choice Model with Structuralized Spatial Effects for Location Analysis. Transportation Research Record: Journal of the Transportation Research Board 1898, 183–190.
URL https://dx.doi.org/10.3141/1898-22

Murdoch, J. C., Sandler, T., Vijverberg, W. P., 2003. The participation decision versus the level of participation in an environmental treaty: a spatial probit analysis. Journal of Public Economics 87 (2), 337–362.
URL https://dx.doi.org/10.1016/S0047-2727(01)00152-9

Pace, R. K., LeSage, J. P., 2016. Fast Simulated Maximum Likelihood Estimation of the Spatial Probit Model Capable of Handling Large Samples. In: Fomby, T. B., Hill, R. C., Jeliazkov, I., Escanciano, J. C., Hillebrand, E. (Eds.), Spatial Econometrics: Qualitative and Limited Dependent Variables. pp. 3–34.
URL https://dx.doi.org/10.1108/S0731-905320160000037008

Panayotou, T., 1993. Empirical tests and policy analysis of environmental degradation at different stages of economic development. World Employment Programme research working paper 292778, Geneva.

Pinkse, J., Slade, M. E., 1998. Contracting in space: An application of spatial statistics to discrete-choice models. Journal of Econometrics 85 (1), 125–154.
URL https://dx.doi.org/10.1016/S0304-4076(97)00097-3

Porter, M. E., 1990. The Competitive Advantage of Nations. Harvard Business Review 68 (2), 73–93.

Porter, M. E., Gee, D. S., Pope, G. J., 2015. America's Unconventional Energy Opportunity: A win-win plan for the economy, the environment, and a lower-carbon, cleaner-energy future. Tech. rep., Harvard Business School and The Boston Consulting Group.

Rice, P., Venables, A. J., Patacchini, E., 2006. Spatial determinants of productivity: Analysis for the regions of Great Britain. Regional Science and Urban Economics 36 (6), 727–752.
URL https://dx.doi.org/10.1016/j.regsciurbeco.2006.03.006

Rupasingha, A., Goetz, S. J., Debertin, D. L., Pagoulatos, A., 2004. The environmental kuznets curve for us counties: A spatial econometric analysis with extensions. Papers in Regional Science 83 (2), 407–424.
URL https://dx.doi.org/10.1111/j.1435-5597.2004.tb01915.x

Schwab, K., Sala-i Martin, X., 2010. The Global Competitiveness Report 2010-2011. World Economic Forum, Geneva.

Shafik, N., Bandyopadhyay, S., 1992. Economic growth and environmental quality:time series and cross-country evidence. World development report WPS 904, World Bank.
URL http://documents.worldbank.org/curated/en/833431468739515725/Economic-growth-and-environmental-quality-time-series-and-cross-country-evidence

Smith, T. E., LeSage, J. P., 2004. A Bayesian probit model with spatial dependencies. In: Fomby, T. B., Hill, R. C., Jeliazkov, I., Escanciano, J. C., Hillebrand, E. (Eds.), Spatial and Spatiotemporal Econometrics. pp. 127–160.
URL https://dx.doi.org/10.1016/S0731-9053(04)18004-3

Wang, H., Iglesias, E. M., Wooldridge, J. M., 2013. Partial maximum likelihood estimation of spatial probit models. Journal of Econometrics 172 (1), 77–89.
URL https://dx.doi.org/10.1016/j.jeconom.2012.08.005

Wollni, M., Andersson, C., 2014. Spatial patterns of organic agriculture adoption: Evidence from Honduras. Ecological Economics 97, 120–128.
URL https://dx.doi.org/10.1016/j.ecolecon.2013.11.010