Regional Science and Urban Economics 76 (2019) 74–102

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

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

Luís Silveira Santos, Isabel Proença *

ISEG-School of Economics and Management, Universidade de Lisboa, REM-Research in Economics and Mathematics, CEMAPRE, Portugal

ARTICLE INFO

Keywords:
Matrix approximation
Matrix factorization
Spatial binary choice models
Spatial lag operator inverse
Competitiveness
Environmental effects

ABSTRACT

This paper presents a new method to approximate the inverse of the spatial lag operator, used in the estimation of spatial lag models for binary dependent variables. The related matrix operations are approximated as well. Closed formulas for the elements of the approximated matrices are deduced. A GMM estimator is also presented. This estimator is a variant of Klier and McMillen's iterative GMM estimator. The approximated matrices are used in the gradients of the new iterative GMM procedure. Monte Carlo experiments suggest that the proposed approximation is accurate and allows to significantly reduce the computational complexity, and consequently the computational time, associated with the estimation of spatial binary choice models, especially for the case where the spatial weighting matrix is large and dense. Also, the simulation experiments suggest that the proposed iterative GMM estimator performs well in terms of bias and root mean square error and exhibits a minimum trade-off between computational time and unbiasedness within a class of spatial GMM estimators. Finally, the new iterative GMM estimator is applied to the analysis of competitiveness in the U.S. Metropolitan Statistical Areas. A new definition for binary competitiveness is introduced. The estimation of spatial and environmental effects are addressed as central issues.

1. Introduction

Modeling binary choice outcomes with spatial dependence has become increasingly popular in recent years. Many applications can be found in the literature, that cover, for example, the choice on the participation in environmental policies (Beron et al., 2003; Murdoch 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 (Klier and McMillen, 2008; Miyamoto et al., 2004), the decision to (re)open a business (Holloway and Lapar, 2007; LeSage et al., 2011) or the existence of high crime rates in a given neighborhood (McMillen, 1992). However, the introduction of spatial dependence in models with dichotomous dependent variables raises several complications. Considering a latent variable approach to derive a model for binary choice outcomes, Anselin (2007) shows that spatial dependence implies the presence of spatial heteroskedasticity and spatial autocorrelation, which leads to specification issues and to analytically intractable expressions for the quantities of interest. As a result, estimation becomes complex and computationally demanding.

To address the issues related to the estimation of spatial binary choice models, several approaches have been proposed. These approaches can be categorized into three major groups, according to the estimation method that they address: Maximum Likelihood (ML) methods, Bayesian methods and Generalized Method of Moments (GMM) estimators. Examples of ML based approaches are 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). The Gibbs sam-

☆ The authors gratefully acknowledge the insightful and valuable comments from the editor, Zhenlin Yang, and two anonymous referees. The authors also acknowledge the helpful comments and suggestions from the participants of the XI World Conference of the Spatial Econometrics Association (SEA). This research was partially supported by the project CEMAPRE-UID/MULTI/00491/2013, financed by the Portuguese Foundation for Science and Technology (FCT/MCTES) through national funds. Luís Silveira Santos is thankful to FCT/MCTES for the financial support through the Ph.D. scholarship SFRH/BD/92277/2013. An earlier version 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. The usual disclaimer applies.

* Corresponding author.
E-mail addresses: lsantos@iseg.ulisboa.pt (L. Silveira Santos), isabelp@iseg.ulisboa.pt (I. Proença).

https://doi.org/10.1016/j.regsciurbeco.2019.01.003
Received 20 October 2017; Received in revised form 19 October 2018; Accepted 3 January 2019
Available online 10 January 2019

pler (LeSage, 2000) and Markov Chains Monte Carlo (Smith and LeSage, 2004) are examples of Bayesian based approaches. Finally, the estimator of Pinkse and Slade (1998) and the estimator of Klier and McMillen (2008) consider the GMM framework. Nevertheless, most of these approaches become computationally burdensome in large samples.

The computational issues associated with Maximum Likelihood methods and Bayesian methods are related to the computation or simulation of high-dimensional integrals. This is a consequence of requiring the specification of the joint distribution (or, at least, some structure of the distribution) of the spatial data. Even if the high-dimensional integrals are approximated by one-dimensional integrals (Martinetti and Geniaux, 2017) or obtained by simulation algorithms (Beron and Vijverberg, 2004; Pace and LeSage, 2016), estimation can still become computationally infeasible, especially if the spatial units are influenced by many neighbors and N (the sample size) is large.

Another possibility is to consider a GMM estimation approach. Under the GMM framework, the distributional assumptions can be relaxed in such a way that no high-dimensional integration is involved. In fact, GMM only requires that a set of moment conditions is correctly specified. But, even so, estimation becomes computationally impracticable in large samples, due to N-dimensional matrix operations that have to be computed on each iteration. Nevertheless, these computational issues can be tackled through matrix approximation methods. The Taylor series approximation, the Chebyshev approximation (Pace and LeSage, 2004), the eigendecomposition of a matrix, the Cholesky decomposition (Pace and Barry, 1997a,b), the LU decomposition or the conjugate gradient method (Smirnov, 2005, 2010) are examples of approximation methods that are commonly used in spatial frameworks. However, some of these approximation methods can be computationally demanding, especially when N is large, and their accuracy depends on the nature of the approximated matrices.

The purpose of this work is twofold. Firstly, it suggests a new approximation method to deal with the computational issues related to the N-dimensional matrix operations required in the GMM estimation of spatially lagged models for binary choice outcomes. The new approximation method focuses on the approximation of the spatial lag operator inverse, since every matrix operation required in the GMM estimation procedure involves the computation of this inverse. The setup for the proposed approximation method relies on non-restrictive assumptions about the spatial weighting matrix and allows to accommodate scenarios where the spatial weighting matrix can be symmetric and non-symmetric. Considering the series expansion of the inverse and the limiting properties of the eigenstructure of normalized spatial weighting matrices, it is shown that the spatial lag operator inverse can be approximated by a sum of known matrices and a simple matrix-vector product. As a result, other related N-dimensional matrix operations can be straightforwardly approximated, as well. Also, closed formulas for the elements of the approximated matrices are available and are deduced. They are especially useful to determine the partial effects.

Secondly, it proposes a computationally simple iterative GMM estimator. This estimator is based on the iterative GMM procedure of Klier and McMillen (2008) together with the approximated matrices deduced in the first part of this paper. This approach has two important advantages. One, the moment conditions of the suggested iterative GMM estimator correspond to orthogonality conditions that use only the information in marginal distributions. Two, the approximated matrices are used in the gradients of the iterative procedure. This allows to significantly reduce the computational complexity and computational time of the suggested GMM estimator, when compared to the traditional GMM estimator. In addition, the spatial heteroskedasticity and spatial autocorrelation robust estimator of Kelejian and Prucha (2007) is used to overcome potential biases in the estimated asymptotic covariance matrix of the GMM estimator for the unknown parameter vector. Note that, the computational simplicity associated with the GMM estimation comes at a cost of accuracy, in comparison to full information methods, where the joint distribution of the spatial data is used in the estimation.

It will be shown through a detailed simulation study that the proposed approximation method fairly approximates the matrices of interest, especially when N is large and the spatial weighting matrix is dense. In addition, the proposed iterative GMM estimator proves to be accurate, especially at low and moderate levels of spatial dependence. At high levels of spatial dependence, the spatial lag parameter tends to be overestimated, which is also a feature shared by other spatial GMM procedures. Moreover, using the approximated matrices in the GMM estimation, not only allows to reduce the associated computational complexity and the overall computational time, especially when N is large and the spatial weighting matrix is dense, but also it allows to increase the precision of the proposed iterative GMM estimator in comparison to other spatial GMM estimators.

The new estimation procedure is used to assess how environmental indicators contribute to influence regional competitiveness in the U.S. Metropolitan Statistical Areas, from 2001 to 2016. A new Binary Competitiveness Indicator (BCI) is introduced. The new competitiveness indicator is based on three dimensions: labor efficiency, capital efficiency and economic growth of the corresponding area. Results show a moderately high degree of spatial dependence between the U.S. Metropolitan Statistical Areas and evidence of an “U” shaped effect of the environmental indicators on regional competitiveness. Also, the suggested estimator exhibited a good performance, in terms of computational time and goodness-of-fit.

The remainder of this paper is organized as follows. Section 2 reviews the literature on the specification and estimation of spatial lag models for binary dependent variables. Section 3 reviews the literature on methods to approximate the inverse of the spatial lag operator and related matrix functions. Section 4 introduces the new method to approximate the inverse of the spatial lag operator. Section 5 derives the new iterative GMM estimation procedure. Section 6 conducts a set of Monte Carlo experiments to assess: firstly, the accuracy and the computational time of the proposed approximation method, compared with the other existing approximation methods; secondly, the statistical properties and the computational performance of the new iterative GMM estimator, compared with the traditional GMM estimator for spatial binary choice models and the GMM estimator of the linearized spatial lag model for binary dependent variables. Section 7 presents an empirical application on the environmental impacts over a spatially lagged Binary Competitiveness Indicator (BCI), in the U.S. Metropolitan Statistical Areas. Finally, section 8 concludes. The results of the Monte Carlo experiments are summarized in Appendix A and 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 weight 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 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 $\mathbf{Y}^* = \left[Y_1^*, Y_2^*, \ldots, Y_N^*\right]^\intercal$ and $\mathbf{X} = \left[\mathbf{X}_1^\intercal, \mathbf{X}_2^\intercal, \ldots, \mathbf{X}_N^\intercal\right]^\intercal$. 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 $\boldsymbol{\xi} = \left[\xi_1, \xi_2, \ldots, \xi_N\right]^\intercal$. The $N \times N$ identity matrix is denoted by $\mathbf{I}$ and the $N \times N$ spatial weighting matrix is denoted by $\mathbf{W}$, with generic element $w_{i,j}$.

If Y_i^* is observable, the conditional expectation is given by $\mathrm{E}\left(Y_i^* \mid \mathbf{X}, \mathbf{W}\right) = \mathbf{X}_i^{\#}\boldsymbol{\beta}$, 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 binary dependent variable, which is a function of particular characteristics of Y_i^* and defined as $Y_i = 1$ if $Y_i^* \geq 0$ and $Y_i = 0$ if $Y_i^* < 0$. The conditional expectation of a spatial lag model when Y_i^* is not observable and Y_i is a binary dependent variable follows as:

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

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 $\boldsymbol{\varepsilon}$:

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

where $\mathrm{Var}(\boldsymbol{\xi} \mid \mathbf{X}, \mathbf{W})$ is fixed, to ensure identification. The scalar σ_i is strictly positive and finite, for all i, assuming that the rows and columns of the matrix $(\mathbf{I} - \alpha\mathbf{W})^{-1}$ are uniformly bounded in absolute value.

In most applications using binary response models, the conditional distribution of ξ_i is assumed to be a standard Normal distribution or a standard Logistic distribution. This implies that $\mathrm{Var}(\xi_i \mid \mathbf{X}, \mathbf{W}) = 1$ and $\mathrm{Var}(\xi_i \mid \mathbf{X}, \mathbf{W}) = \pi^2/3$, respectively, for all i. Under these two specifications, the probability distribution functions (PDFs) of the link functions are symmetric about zero, but this is generally not the case for other possible links.[2]

Note that the specification in (3) is similar to the specification of McMillen (1992) and LeSage (2000). Under this approach, only the information in the marginal distributions of ε_i conditional on $(\mathbf{X}, \mathbf{W})$ is used. The implications of this approach, regarding estimation, are discussed in section 6.

Considering a generic link function, the spatial lag model for binary dependent variables 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 $u_i = Y_i - \mathrm{E}(Y_i \mid \mathbf{X}, \mathbf{W})$ and $\varepsilon_i = Y_i^* - \mathrm{E}\left(Y_i^* \mid \mathbf{X}, \mathbf{W}\right)$. Hence, u_i is a discrete random variable assuming only two values, $1 - G(\cdot)$ and $-G(\cdot)$.

To estimate the model (5), a GMM approach is considered, based on the works of Pinkse and Slade (1998) and Klier and McMillen (2008). It is assumed that the unknown parameters $\boldsymbol{\beta}$ and α satisfy the following moment condition:

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

where $\mathbf{Z}$ is the $N \times (K + p)$ matrix of instruments, with p the number of additional instruments that are usually given by the product between the powers of $\mathbf{W}$ and the matrix of explanatory variables. The $N \times 1$ vector $\mathbf{u}_*$ correspond to the "generalized errors" (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. Klier and McMillen (2008) sets $\boldsymbol{\Xi} = (\mathbf{Z}^\intercal\mathbf{Z})^{-1}$ and the GMM estimator reduces to nonlinear two stages least squares (N2SLS). However, because the minimization problem in (8) does not have a closed formula, an iterative procedure is used to obtain a solution for the unknown parameters. 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, $\boldsymbol{\Gamma}_i^{(0)} = \left(\partial u_{*,i}/\partial\boldsymbol{\Theta}\right)|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(0)}}$, $i = 1, 2, \ldots, N$.
2. Regress $\boldsymbol{\Gamma}^{(0)}$ on $\mathbf{Z}$, in a similar fashion to (linear) 2SLS. Obtain $\widehat{\boldsymbol{\Gamma}}^{(0)}$.
3. Construct new estimates as $\boldsymbol{\Theta}^{(1)} = \boldsymbol{\Theta}^{(0)} + \left[\left(\widehat{\boldsymbol{\Gamma}}^{(0)}\right)^\intercal\left(\widehat{\boldsymbol{\Gamma}}^{(0)}\right)\right]^{-1}\left(\widehat{\boldsymbol{\Gamma}}^{(0)}\right)^\intercal \mathbf{u}_*^{(0)}$, 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 spatial heteroskedasticity and spatial autocorrelation robust covariance estimator of the (iterative) GMM estimator follows as in Kelejian and Prucha (2007):

$$\begin{aligned}\widehat{Avar\left(\widehat{\boldsymbol{\Theta}}\right)} = \left(\sum_{i=1}^{N}\widehat{\boldsymbol{\Gamma}}_i^\intercal\widehat{\boldsymbol{\Gamma}}_i\right)^{-1}\Bigg\{\sum_{i=1}^{N}\widehat{u}_i^2\widehat{\boldsymbol{\Gamma}}_i^\intercal\widehat{\boldsymbol{\Gamma}}_i + \sum_{j=1}^{n-1}\left[K\left(\frac{j}{d^*}\right)\right. \\ \left.\times\sum_{i=1}^{n-j}\widehat{u}_i\widehat{u}_{i+j}\left(\widehat{\boldsymbol{\Gamma}}_i^\intercal\widehat{\boldsymbol{\Gamma}}_{i+j} + \widehat{\boldsymbol{\Gamma}}_{i+j}^\intercal\widehat{\boldsymbol{\Gamma}}_i\right)\right]\Bigg\}\left(\sum_{i=1}^{N}\widehat{\boldsymbol{\Gamma}}_i^\intercal\widehat{\boldsymbol{\Gamma}}_i\right)^{-1}\end{aligned} \tag{9}$$

where $K(j/d^*)$ is a Kernel function with $K : \mathbb{R} \to [-1, 1]$, $K(0) = 1$, $K(j/d^*) = K(-j/d^*)$ and $K(j/d^*) = 0$, for $|j/d^*| > 1$, that satisfies $|K(j/d^*) - 1| \leq c_K|j/d^*|^{\rho_K}$, for $|j/d^*| \leq 1$, for some $\rho_K \geq 1$ and a finite positive c_K. The scalar d^* is a distance threshold.

The individual gradients for each parameter are:

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

and

$$\begin{aligned}\left(\boldsymbol{\Gamma}_\alpha\right)_i = \frac{\partial u_{*,i}}{\partial\alpha} = -u_{*,i}\left(\frac{g'\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)}{g\left(\frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{\sigma_i}\right)} - u_{*,i}\right) \\ \times\left[\frac{1}{\sigma_i}\left(\mathbf{H}_i\boldsymbol{\beta} - \frac{\mathbf{X}_i^{\#}\boldsymbol{\beta}}{2\sigma_i^2}\Upsilon_{ii}\right)\right], \quad i = 1, 2, \ldots, N\end{aligned} \tag{11}$$

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

[2] See, for example, the complementary log-log link or the Weibull link.

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 $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{X}$ and Υ_{ii} is the ith element of the diagonal of the matrix:

$$\begin{aligned}\Upsilon &= (\mathbf{I}-\alpha\mathbf{W})^{-1}\left\{\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}+\left[\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}\right]^{\intercal}\right\}\left[(\mathbf{I}-\alpha\mathbf{W})^{-1}\right]^{\intercal}\\ &= (\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}\left[(\mathbf{I}-\alpha\mathbf{W})^{-1}\right]^{\intercal}\\ &\quad+\left\{(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}\left[(\mathbf{I}-\alpha\mathbf{W})^{-1}\right]^{\intercal}\right\}^{\intercal}\end{aligned} \tag{12}$$

with the diagonal of $\boldsymbol{\Upsilon}$ equal to

$$\operatorname{diag}(\Upsilon) = 2\times\operatorname{diag}\left((\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}\left[(\mathbf{I}-\alpha\mathbf{W})^{-1}\right]^{\intercal}\right) \tag{13}$$

Having closed formulas for the gradients help to accelerate the numerical optimization process. However, because they depend on the spatial lag operator inverse, which has to be computed on each iteration, the estimation procedure becomes computationally burdensome, especially for large samples and/or dense spatial weighting matrices. To solve this issue, Klier and McMillen (2008) suggest a first order Taylor series approximation of model (5) around $\alpha = 0$. In this way, the previous iterative GMM procedure is simplified, because no large matrices need to be inverted. The major drawback of this approach is related to the poor accuracy of the estimates for the spatial lag parameter when $\alpha > 0.5$.

Another possibility is to address the previously mentioned computational issues through the approximation of the spatial lag operator inverse. Under this approach, the nonlinearity of model (5) is preserved and the estimates for the spatial lag parameter yield reasonable results for all admissible values of α. The methods that are commonly used in the literature are presented in the section below.

3. Approximation methods for the spatial lag operator inverse

To deal with the computational issues related to the inversion of the spatial lag operator, several methods have been proposed in the literature. These methods approach the matrix inversion explicitly or implicitly. For the explicit methods, the $N \times N$ spatial lag operator inverse is explicitly obtained; examples are the Taylor series approximation, the Chebyshev approximation (Pace and LeSage, 2004) and the eigendecomposition of the spatial weighting matrix. For the implicit methods, a system that involves the spatial lag operator (usually a matrix-vector product) is solved and a $N \times 1$ vector is obtained rather than a $N \times N$ matrix; examples are the Cholesky decomposition (Pace and Barry, 1997a,b), the LU decomposition and the conjugate gradient (Smirnov, 2005, 2010). Before presenting the details of the previous methods, a set of assumptions are stated. Note that these assumptions are already commonly used in the literature of spatial binary choice models (see e.g., Baltagi et al., 2016; Billé, 2013).

3.1. Assumptions

Focusing on the properties of an initial spatial weighting matrix, $\mathbf{W}_0$, it is assumed that:

Assumption 1. *The matrix $\mathbf{W}_0$ is non-stochastic and diagonalizable.*

Assumption 2. *All of the diagonal elements of $\mathbf{W}_0$ are equal to zero.*

Assumption 3. *The matrix $(\mathbf{I}-\alpha\mathbf{W}_0)$ is non-singular for all $\alpha \in (-1/|\lambda|_{0,\max}, 1/|\lambda|_{0,\max})$, where $|\lambda|_{0,\max}$ is the largest absolute eigenvalue of $\mathbf{W}_0$. Additionally, $|\lambda|_{0,\max}$ is assumed to be bounded away from zero by some fixed constant $c_{\lambda_{0,\max}}$.*

Assumption 4. *Both row and column sums of $\mathbf{W}_0$ and $(\mathbf{I}-\alpha\mathbf{W}_0)^{-1}$ are uniformly bounded in absolute value by some constant $c_{\mathbf{W}_0}$, with $0 < c_{\mathbf{W}_0} < \infty$.*

Assumption 5. *The matrix $\mathbf{W}$ is row-normalized and equal to $\mathbf{D}_R^{-1}\mathbf{W}_0$, where $\mathbf{D}_R$ is a $N \times N$ diagonal matrix whose diagonal elements are the row sums of $\mathbf{W}_0$.*

Remark 1. *Under Assumption 5 the following properties are verified:*

(a) *$\mathbf{W}$ is non-symmetric;*
(b) *The eigenvalues of $\mathbf{W}$ are, in absolute value, less than or equal to one;*
(c) *The largest absolute eigenvalue of $\mathbf{W}$, $|\lambda|_{\max}$, is equal to one;*
(d) *The eigenvector of $\mathbf{W}$ associated with the largest absolute eigenvalue is the vector of ones, $\boldsymbol{\iota}$;*
(e) *The matrix $(\mathbf{I}-\alpha\mathbf{W})$ is non-singular for all $\alpha \in (-1,1)$ (see also Kelejian and Robinson, 1995).*

Alternatively, $\mathbf{W}_0$ can be normalized through the transformation $\mathbf{W}_{sim} = \mathbf{D}_R^{-1/2}\mathbf{W}_0\mathbf{D}_R^{-1/2}$ (Ord, 1975), where $\mathbf{W}_{sim}$ is a $N \times N$ matrix that is similar to $\mathbf{W}$. In fact, $\mathbf{W}_{sim}$ can be written as $\mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2}$. By definition, the eigenvalues of $\mathbf{W}_{sim}$ and $\mathbf{W}$ are equal, which imply that their eigenvectors are directly related. Also, in general, $\mathbf{W}_{sim}$ is non-symmetric. However, if $\mathbf{W}_0$ is symmetric, then $\mathbf{W}_{sim}$ is symmetric as well.[3] A later discussion on the eigendecomposition of $\mathbf{W}$ will recover this result.

3.2. Explicit methods

Consider the Taylor 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 $\alpha \in (-1,1)$. Following LeSage and Pace (2009), the series (14) is partitioned into a finite lower-order and an infinite higher-order series:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1} = \sum_{h=0}^{q}\alpha^h\mathbf{W}^h + \sum_{h=q+1}^{\infty}\alpha^h\mathbf{W}^h \tag{15}$$

As suggested by several authors (Arbia, 2014; Elhorst, 2014; LeSage and Pace, 2009, to name a few), for the case where α quickly converges to zero, $(\mathbf{I}-\alpha\mathbf{W})^{-1}$ can be accurately approximated through the finite lower-order series:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1} \approx \sum_{h=0}^{q}\alpha^h\mathbf{W}^h \tag{16}$$

where q is small. The expression (16) corresponds to the Taylor series approximation of the spatial lag operator inverse.

The Chebyshev approximation (Pace and LeSage, 2004) for the spatial lag operator inverse is:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1} \approx \left[\sum_{h=0}^{q}c_l(\alpha)\,T_h(\mathbf{W})\right] - \frac{1}{2}c_0(\alpha)\,\mathbf{I} \tag{17}$$

where

$$c_l(\alpha) = \frac{2}{q+1}\sum_{m=1}^{q+1}f(x_m)\cos\left(\frac{\pi\, l\left(m-\frac{1}{2}\right)}{q+1}\right) \tag{18}$$

$$x_m = \cos\left(\frac{\pi\left(m-\frac{1}{2}\right)}{q+1}\right) \tag{19}$$

$$f(x) = (1-\alpha x)^{-1} \tag{20}$$

[3] If $\mathbf{W}_0$ is symmetric, then $\mathbf{W}_0 = \mathbf{W}_0^{\intercal}$. Replacing $\mathbf{W}_0$ by $\mathbf{D}_R^{-1/2}\mathbf{W}_0\mathbf{D}_R^{-1/2}$ yields $\mathbf{W}_{sim} = \mathbf{W}_{sim}^{\intercal}$.

and $T_{k+1}(\mathbf{W}) = 2\mathbf{W}T_k(\mathbf{W}) - T_{k-1}(\mathbf{W})$, for $k \geq 1$, with $T_0(\mathbf{W}) = \mathbf{I}$ and $T_1(\mathbf{W}) = \mathbf{W}$. The scalars $c_l(\alpha)$, with $l = 0, 1, 2, \ldots, q$ are the Chebyshev coefficients and depend on the spatial lag parameter. The functions $T_k(\mathbf{W})$, with $k = 0, 1, 2, \ldots, q$ are the Chebyshev polynomials and depend only on the spatial weighting matrix.

Finally, consider the eigendecomposition of the spatial lag operator inverse:

$$\begin{aligned}(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{I}+\alpha\mathbf{W}+\alpha^2\mathbf{W}^2+\alpha^3\mathbf{W}^3+\ldots \\ &= \mathbf{I}+\alpha\mathbf{V}\Lambda\mathbf{V}^{-1}+\alpha^2\left(\mathbf{V}\Lambda\mathbf{V}^{-1}\right)^2+\alpha^3\left(\mathbf{V}\Lambda\mathbf{V}^{-1}\right)^3+\ldots \\ &= \mathbf{I}+\alpha\mathbf{V}\Lambda\mathbf{V}^{-1}+\alpha^2\mathbf{V}\Lambda^2\mathbf{V}^{-1}+\alpha^3\mathbf{V}\Lambda^3\mathbf{V}^{-1}+\ldots \\ &= \mathbf{V}\left(\mathbf{I}+\alpha\Lambda+\alpha^2\Lambda^2+\alpha^3\Lambda^3+\ldots\right)\mathbf{V}^{-1} \\ &= \mathbf{V}(\mathbf{I}-\alpha\Lambda)^{-1}\mathbf{V}^{-1}\end{aligned} \tag{21}$$

where the $N \times N$ diagonal matrix $\mathbf{\Lambda}$ contains the corresponding eigenvalues of $\mathbf{W}$ and the $N \times N$ matrix $\mathbf{V}$ contains, on each column, the ith eigenvector associated with the ith eigenvalue of $\mathbf{W}$. Contrary to the previous approximation methods, the expression in (21) is exact. Also, the inverse of the eigenvector matrix is only required to be computed once. Nevertheless, for the case where $\mathbf{W}_{sim}$ is symmetric, the eigenvectors in (21) can be expressed as orthogonal eigenvectors. In fact, replacing $\mathbf{W}_{sim}$ in (21) yields:

$$\begin{aligned}(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{D}_R^{-1/2}(\mathbf{I}-\alpha\mathbf{W}_{sim})^{-1}\mathbf{D}_R^{1/2} \\ &= \mathbf{D}_R^{-1/2}\mathbf{V}_{sim}(\mathbf{I}-\alpha\Lambda)^{-1}\mathbf{V}_{sim}^{\intercal}\mathbf{D}_R^{1/2}\end{aligned} \tag{22}$$

where the only matrix that is required to be computed is the $N \times N$ matrix $\mathbf{V}_{sim}$, that correspond to the orthogonal eigenvectors of $\mathbf{W}_{sim}$.

It is important to note that these methods can also be applied to approximate other matrix operations. In particular, they are useful to derive approximate or exact expressions for the matrix $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$, the diagonal elements of $\mathbf{\Upsilon}$ and the diagonal elements of $\mathbf{\Sigma}$, that are required in the computation of the gradients (10) and (11), on each iteration.

Focusing on the term $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$, the series expansion is given by:

$$\begin{aligned}(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{W}+2\alpha\mathbf{W}^2+3\alpha^2\mathbf{W}^3+\ldots \\ &= \sum_{h=0}^{\infty}(h+1)\,\alpha^h\mathbf{W}^{h+1}\end{aligned} \tag{23}$$

and the lower-order Taylor series approximation is:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1} \approx \sum_{h=0}^{q}(h+1)\,\alpha^h\mathbf{W}^{h+1} \tag{24}$$

Therefore, the diagonal of $\mathbf{\Upsilon}$ is approximately equal to:

$$\operatorname{diag}(\Upsilon) \approx 2 \times \operatorname{diag}\left(\sum_{j=1}^{N}\left(\left[\sum_{h=0}^{q}(h+1)\,\alpha^h\mathbf{W}^{h+1}\right]\circ\left[\sum_{h=0}^{q}\alpha^h\mathbf{W}^h\right]\right)_{i,j}\right) \tag{25}$$

where "$\circ$" is the Hadamard product operator. The previous expression implies that the diagonal elements of $\mathbf{\Upsilon}$ are approximately given by the row sums of the Hadamard product between the approximation of the matrix $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$ and the approximation of the matrix $(\mathbf{I}-\alpha\mathbf{W})^{-1}$ – the spatial lag operator inverse.

Analogously, the Chebyshev approximation of $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$ is given by (17), where the Chebyshev coefficients are replaced by the function $f(x) = x/(1-\alpha x)^2$, and the eigendecomposition of $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1}$ is:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1} = \mathbf{V}\Lambda(\mathbf{I}-\alpha\Lambda)^{-2}\mathbf{V}^{-1} \tag{26}$$

For the case where $\mathbf{W}_{sim}$ is symmetric, (26) can be written as:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I}-\alpha\mathbf{W})^{-1} = \mathbf{D}_R^{-1/2}\mathbf{V}_{sim}\Lambda(\mathbf{I}-\alpha\Lambda)^{-2}\mathbf{V}_{sim}^{\intercal}\mathbf{D}_R^{1/2} \tag{27}$$

Again, note that (26) and (27) are exact expressions.

With regard to the diagonal elements of $\mathbf{\Sigma}$, they can be obtained as the row sums of the Hadamard square of the approximated or exact expression for the spatial lag operator inverse.

Nevertheless, these approaches can still be extremely demanding if the sample size is large and/or the spatial weighting matrix is dense. This is because, for both Taylor series and Chebyshev approximation, there are as many matrix operations as the number of lower-order powers of $\mathbf{W}$, and, for the eigendecomposition, the full eigensystem is required. Also, the approximate functional form for the elements of the approximated matrices is complicated, especially for the elements of the spatial lag operator inverse.

3.3. Implicit methods

The implicit methods used to compute the inverse of the spatial lag operator are based in the solution of the following equation:

$$(\mathbf{I}-\alpha\mathbf{W})\,\boldsymbol{\varepsilon} = \boldsymbol{\xi} \tag{28}$$

for $\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ and $\boldsymbol{\xi}$ are $N \times 1$ vectors.

Consider the LU decomposition, which generalizes the Cholesky decomposition to non-symmetric matrices. Following LeSage and Pace (2009), suppose that $(\mathbf{I}-\alpha\mathbf{W}) = \mathbf{LU}$. The solution for the system $\mathbf{LU}\boldsymbol{\varepsilon} = \boldsymbol{\xi}$ is identical to the solution for $\mathbf{L}\boldsymbol{\nu} = \boldsymbol{\xi}$, where $\boldsymbol{\nu} = \mathbf{U}\boldsymbol{\varepsilon}$.

The conjugate gradient method (Smirnov, 2005, 2010) is a numerical method that minimizes the quadratic form:

$$\mathfrak{f}(\boldsymbol{\varepsilon}) = \frac{1}{2}\boldsymbol{\varepsilon}^{\intercal}(\mathbf{I}-\alpha\mathbf{W})\,\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^{\intercal}\boldsymbol{\xi} \tag{29}$$

based on orthogonal descent directions.

In contrast to the explicit methods, the implicit methods have the advantage that the $N \times N$ inverse is not explicitly computed. However, for the LU decomposition the spatial lag operator has to be decomposed into a lower triangular and upper triangular matrix, which can be computationally demanding if the sample size is large and/or the spatial weighting matrix is dense. As for the conjugate gradient method, it may not converge for certain designs of $\mathbf{W}$ because it is not accurate for matrices that are not symmetric and positive definite. Also, no functional form for the elements of the spatial lag operator inverse is available.

4. The new explicit approximation method based on known matrices

In this section a new explicit method to obtain the inverse of the spatial lag operator is proposed. Consider the series expansion of the inverse in (14). The idea is to approximate the powers $h \geq 2$ of $\mathbf{W}$ by a "long run" matrix, $\mathbf{W}^{\infty}$, which is obtained from the limiting properties of the eigenstructure of $\mathbf{W}$ and it is equal to a simple matrix-vector product. In this way, no additional matrix operations are required and a closed formula for the elements of the spatial lag operator inverse are available and deduced. The details of this new procedure are presented and discussed below.

Consider that the assumptions of Section 3.1 hold. In addition, consider the following assumption on the eigenstructure of $\mathbf{W}$:

Assumption 6. *The algebraic multiplicity of* $|\lambda|_{\max}$, $\operatorname{amult}(|\lambda|_{\max})$, *is equal to one. For a block diagonal* $\mathbf{W}$, *the largest absolute eigenvalue of each block has algebraic multiplicity equal to one.*

Note that, in general, the cases where the algebraic multiplicity of $|\lambda|_{\max}$ is greater than one are those where there are only one or two neighbors for every spatial unit. In practice, this assumption is not too restrictive, because, in most of the applications, there are more than two neighbors for every spatial unit or there are few spatial units with less than two neighbors. Nevertheless, this assumption can be relaxed,

but at a cost of computational accuracy, as it will be shown in a Monte Carlo simulation study.

Now, under Assumption 1 to Assumption 6, the approximation of the spatial lag operator inverse is given by:

$$\begin{aligned}(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{I}+\alpha\mathbf{W}+\alpha^2\mathbf{W}^2+\alpha^3\mathbf{W}^3+\cdots+\alpha^q\mathbf{W}^q+\ldots \\ &\approx \mathbf{I}+\alpha\mathbf{W}+\alpha^2\mathbf{W}^\infty+\alpha^3\mathbf{W}^\infty+\cdots+\alpha^q\mathbf{W}^\infty+\ldots \\ &= \mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^\infty\end{aligned} \tag{30}$$

which converges absolutely for all α in the parameter space (see Assumption 5 and Kelejian and Robinson, 1995). The $N \times N$ matrix $\mathbf{W}^\infty$ is the "long run" matrix and equal to $\lim_{h\to\infty}\mathbf{W}^h$. Since the eigendecomposition of $\mathbf{W}$ is available, $\mathbf{W}^\infty$ can also be written as:

$$\begin{aligned}\mathbf{W}^\infty &= \lim_{h\to\infty}\mathbf{W}^h = \mathbf{V}\left(\lim_{h\to\infty}\Lambda^h\right)\mathbf{V}^{-1} \\ &= \mathbf{V}\left[\lim_{h\to\infty}\operatorname{diag}\left(1^h,\lambda_2^h,\lambda_3^h\ldots,\lambda_N^h\right)\right]\mathbf{V}^{-1} \\ &= \mathbf{V}\begin{bmatrix}1 & 0 & \ldots & 0\\ 0 & 0 & \ldots & 0\\ \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \ldots & 0\end{bmatrix}\mathbf{V}^{-1} = \operatorname{col}(\mathbf{V})_1\operatorname{row}\left(\mathbf{V}^{-1}\right)_1\end{aligned} \tag{31}$$

because the eigenvalues $\lambda_2, \lambda_3 \ldots, \lambda_N$ are, in absolute value, less than one. Also, $\lambda_1 = |\lambda|_{\max} = 1$ (see Assumption 5 and Assumption 6). The $N \times 1$ vector $\operatorname{col}(\mathbf{V})_1$ is the first column of $\mathbf{V}$ and the $1 \times N$ vector $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$ is the first row of $\mathbf{V}^{-1}$. It is important to note that obtaining these vectors entail drastically different implications. On one hand, the expression for $\operatorname{col}(\mathbf{V})_1$ is exact and equal to $N \times 1$ vector of ones, $\boldsymbol{\iota}$ (see Assumption 5). On the other hand, to obtain $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$ the entire linear system has to be solved, which becomes computationally infeasible in large samples.

Here, the issue related to the computation of $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$ is addressed through the orthogonalization of the eigenvectors of $\mathbf{W}$, analogous to the approach presented in Section 3.2 for the eigendecomposition problem. The similar matrix, $\mathbf{W}_{sim}$, is used, because the eigenvectors are related to those of $\mathbf{W}$. However, as previously mentioned, $\mathbf{W}_{sim}$ is not necessarily symmetric, as it depends on the properties of the initial spatial weighting matrix, $\mathbf{W}_0$. Therefore, for the case where $\mathbf{W}_{sim}$ is symmetric, $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$ can be straightforwardly written as a function of an orthogonal eigenvector. For the case where $\mathbf{W}_{sim}$ is not symmetric, a "symmetrization" procedure is suggested, such that $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$ can be approximated by a function of an orthogonal eigenvector.

In the next subsections, the exact and approximated expressions for $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$ are derived, according to the symmetric and non-symmetric scenarios of $\mathbf{W}_0$. Also, it will be shown that, the expressions for $\operatorname{row}\left(\mathbf{V}^{-1}\right)_1$ are based on known quantities. In this way, the approximated expressions for the elements of the spatial lag operation inverse will be derived, as well.

4.1. Case 1: symmetric $\boldsymbol{W_0}$

Consider that $\mathbf{W}_0$ is symmetric. Therefore, $\mathbf{W}_{sim}$ is also symmetric. Because $\mathbf{W}_{sim}$ can be written as a function of $\mathbf{W}$, consider the eigendecomposition for both matrices:

$$\mathbf{W}_{sim} = \mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2} \Leftrightarrow \mathbf{V}_{sim}\Lambda\mathbf{V}_{sim}^\intercal = \mathbf{D}_R^{1/2}\mathbf{V}\Lambda\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \tag{32}$$

where $\boldsymbol{\Lambda}$ is equal in both sides of the equation due to matrix similarity. The equation above implies that the eigenvectors of $\mathbf{W}_{sim}$ and the eigenvectors of $\mathbf{W}$ are related in the following way:

$$\mathbf{V}_{sim} = \mathbf{D}_R^{1/2}\mathbf{V} \text{ and } \mathbf{V}_{sim}^\intercal = \mathbf{V}^{-1}\mathbf{D}_R^{-1/2}, \text{ but also } \mathbf{V}_{sim}^\intercal = \mathbf{V}^\intercal\mathbf{D}_R^{1/2} \tag{33}$$

Consider the hth power of (32):

$$\begin{aligned}&\mathbf{V}_{sim}\Lambda^h\mathbf{V}_{sim}^\intercal = \mathbf{D}_R^{1/2}\mathbf{V}\Lambda^h\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \Leftrightarrow \\ &\Leftrightarrow \mathbf{D}_R^{-1/2}\mathbf{V}_{sim}\Lambda^h\mathbf{V}_{sim}^\intercal\mathbf{D}_R^{1/2} = \mathbf{V}\Lambda^h\mathbf{V}^{-1}\end{aligned} \tag{34}$$

Using the eigenvector relationship in (33) yields:

$$\mathbf{D}_R^{-1/2}\left(\mathbf{D}_R^{1/2}\mathbf{V}\right)\Lambda^h\left(\mathbf{V}^\intercal\mathbf{D}_R^{1/2}\right)\mathbf{D}_R^{1/2} = \mathbf{V}\Lambda^h\mathbf{V}^{-1} \tag{35}$$

Now, as $h \to \infty$:

$$\begin{aligned}\mathbf{W}^\infty &= \frac{1}{\|\mathbf{D}_R^{1/2}\operatorname{col}(\mathbf{V})_1\|\operatorname{col}(\mathbf{V})_1^\intercal\mathbf{D}_R^{1/2}\|_2} \\ &\quad\times\mathbf{D}_R^{-1/2}\left(\mathbf{D}_R^{1/2}\operatorname{col}(\mathbf{V})_1\operatorname{col}(\mathbf{V})_1^\intercal\mathbf{D}_R^{1/2}\right)\mathbf{D}_R^{1/2} \\ &= \frac{1}{\left[\sqrt{\left(d_{R,1}^{1/2}\right)^2+\left(d_{R,2}^{1/2}\right)^2+\cdots+\left(d_{R,N}^{1/2}\right)^2}\right]^2}\times\boldsymbol{\iota}^\intercal\mathbf{D}_R \\ &= \left(\sum_{i=1}^N d_{R,i}\right)^{-1}\times\mathbf{J}\mathbf{D}_R\end{aligned} \tag{36}$$

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 the first eigenvector of $\mathbf{W}$ is now orthogonal. There are two major advantages related to this expression. First, the matrix $\mathbf{W}^\infty$ is given by a simple matrix-vector product, since $\mathbf{D}_R$ is a diagonal matrix. Second, each element of the matrix $\mathbf{W}^\infty$ have an exact closed formula:

$$w_{i,j}^\infty = \left(\sum_{i=1}^N d_{R,i}\right)^{-1}\times d_{R,j} \tag{37}$$

which implies that the rows of $\mathbf{W}^\infty$ are all equal and given by the sum of the ith row of $\mathbf{W}_0$, that is $\operatorname{row}(\mathbf{W}^\infty)_i = \left(d_{R,1}, d_{R,2}, \ldots, d_{R,N}\right)$, for all i.

Plugging (36) into (30), the approximation of the spatial lag operator inverse is given by:

$$(\mathbf{I}-\alpha\mathbf{W})^{-1} \approx \mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\left(\sum_{i=1}^N d_{R,i}\right)^{-1}\mathbf{J}\mathbf{D}_R \tag{38}$$

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

$$\left((\mathbf{I}-\alpha\mathbf{W})^{-1}\right)_{i,j} \approx \mathbb{1}_{i=j}+\alpha\times w_{i,j}+\frac{\alpha^2}{1-\alpha}\times\left(\sum_{i=1}^N d_{R,i}\right)^{-1}\times d_{R,j} \tag{39}$$

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 accuracy of this approximation depends on how fast the powers of the eigenvalues $\lambda_2, \lambda_3, \ldots, \lambda_N$ converge to zero, for a given value of α. In fact, this is a special case of the approximation method proposed by Griffith (2000), for linear models.

The expressions for the approximation of the spatial lag operator inverse, in (38) and (39), allow for an interesting interpretation of the product $(\mathbf{I}-\alpha\mathbf{W})^{-1}\mathbf{X}$ which is approximately equal to $\mathbf{X}+\alpha\mathbf{W}\mathbf{X}+\alpha^2/(1-\alpha)^{-1}\left(\sum_{i=1}^N d_{R,i}\right)^{-1}\mathbf{J}\mathbf{D}_R\mathbf{X}$. This means that the previous matrix product can be decomposed into the original matrix $\mathbf{X}$, a spatial lag of the matrix $\mathbf{X}$ and the "long run" spatial lag of the matrix $\mathbf{X}$, that incorporates the bilateral effects (the combination of the neighboring effects on a given unit and the effects of a given unit on its neighbors).

Note that the previous results are valid when $\mathbf{W}$ is column-normalized or when $\mathbf{W}$ is doubly stochastic (simultaneously row- and column-normalized). For the first case, the approximation method is applied to $\mathbf{W}^\intercal$, because it is row stochastic. For the second case, considering that $\mathbf{W}_0$ is symmetric, the doubly stochastic $\mathbf{W}$ is also symmetric,

which implies that its eigenvectors are already orthogonal and the "long run" matrix simplifies to $\mathbf{W}^{\infty} = (1/n)\,\mathbf{J}$.[4]

4.2. Case 2: non-symmetric $\boldsymbol{W}_0$

Consider that $\mathbf{W}_0$ is non-symmetric. In this case, the previous result for $\mathbf{W}^{\infty}$ is not valid. To see this write $\mathbf{W}_{sim}$ as a function of $\mathbf{W}$ and consider the eigendecomposition for both matrices:

$$\mathbf{W}_{sim} = \mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2} \Leftrightarrow \mathbf{V}_{sim}\Lambda\mathbf{V}_{sim}^{-1} = \mathbf{D}_R^{1/2}\mathbf{V}\Lambda\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \tag{40}$$

with $\mathbf{V}_{sim}^{-1} \neq \mathbf{V}_{sim}^{\intercal}$ because the eigenvectors of $\mathbf{W}_{sim}$ are no longer orthogonal. Therefore, to approximate the spatial lag operator inverse 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 if unit j is a neighbor of unit i, then unit i is also a neighbor of unit j with equal weight, for all $i, j = 1, 2, \ldots, N$ and $i \neq j$. This follows as:

$$\mathbf{W}_0^* = \mathbf{W}_0 - \frac{1}{2}\left\{\mathbf{W}_0 - \mathbf{W}_0^{\intercal} - \left[\left(\mathbf{W}_0 - \mathbf{W}_0^{\intercal}\right)^{\circ 2}\right]^{\circ\frac{1}{2}}\right\} = \mathbf{W}_0 + \mathbf{A} \tag{41}$$

where $\mathbf{A}$ is the $N \times N$ "symmetrization" matrix. The operators "$\circ 2$" and "$\circ\frac{1}{2}$" are element-wise operations and correspond to the Hadamard square and to the Hadamard square root, respectively. Also, as in the previous case, a row-normalized matrix and a 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 $N \times N$ diagonal matrix whose diagonal elements are the row sums of $\mathbf{W}_0^*$, and the similar matrix is $\mathbf{W}_{sim}^* = \mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0^*\mathbf{D}_{R^*}^{-1/2}$. Note that, here, the previous assumptions (see Section 3.1) are also valid for $\mathbf{W}_0^*$, $\mathbf{W}^*$ and $\mathbf{W}_{sim}^*$.

For $\mathbf{A}$ close to the null matrix, $\mathbf{0}$, the matrix $\mathbf{W}_0$ is well approximated by $\mathbf{W}_0^*$. Then the eigenvectors of $\mathbf{W}_{sim}$ can be approximated by the orthogonal eigenvectors of $\mathbf{W}_{sim}^*$. To see this, write $\mathbf{W}_{sim}^*$ as a function of $\mathbf{W}_{sim}$ and consider the eigendecomposition for both matrices:

$$\begin{aligned}
&\mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0^*\mathbf{D}_{R^*}^{-1/2} \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\
&\quad\Leftrightarrow \mathbf{W}_{sim}^* \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{D}_R^{-1/2}\mathbf{W}_0\mathbf{D}_R^{-1/2}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\
&\quad\Leftrightarrow \mathbf{W}_{sim}^* \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{W}_{sim}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\
&\Leftrightarrow \mathbf{V}_{sim}^*\Lambda^*\left(\mathbf{V}_{sim}^*\right)^{\intercal} \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{V}_{sim}\Lambda\mathbf{V}_{sim}^{-1}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}
\end{aligned} \tag{42}$$

where the $N \times N$ diagonal matrix $\mathbf{\Lambda}^*$ contains the corresponding eigenvalues of $\mathbf{W}_{sim}^*$ and the $N \times N$ matrix $\mathbf{V}_{sim}^*$ contains, on each column, the ith eigenvector associated with the ith eigenvalue of $\mathbf{W}_{sim}^*$. Note that $\mathbf{\Lambda}^*$ and $\mathbf{\Lambda}$ are not equal, but because $\mathbf{W}_{sim}^*$ and $\mathbf{W}_{sim}$ are similar to the corresponding row-normalized matrices, $\lim_{h\to\infty}\left(\Lambda^*\right)^h = \lim_{h\to\infty}\Lambda^h$. Also, the equation above implies that the eigenvectors of $\mathbf{W}_{sim}^*$ and the eigenvectors of $\mathbf{W}_{sim}$ are approximately related as:

$$\begin{aligned}
&\mathbf{V}_{sim}^* \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{V}_{sim} \text{ and } \left(\mathbf{V}_{sim}^*\right)^{\intercal} \\
&\quad\approx \mathbf{V}_{sim}^{-1}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}, \text{ but also } \left(\mathbf{V}_{sim}^*\right)^{\intercal} \approx \mathbf{V}_{sim}^{\intercal}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}
\end{aligned} \tag{43}$$

where $\mathbf{V}_{sim}^{-1}$ can be straightforwardly approximated by $\left(\mathbf{V}_{sim}^*\right)^{\intercal}\mathbf{D}_{R^*}^{1/2}\mathbf{D}_R^{-1/2}$. Analogously to the results in (32) and (33), the eigenvectors of $\mathbf{W}_{sim}^*$ are related to the eigenvectors of $\mathbf{W}^*$ as $\mathbf{V}_{sim}^* = \mathbf{D}_{R^*}^{1/2}\mathbf{V}^*$. This implies that:

$$\mathbf{V}_{sim}^{-1} \approx \left(\mathbf{V}^*\right)^{\intercal}\mathbf{D}_{R^*}^{1/2}\mathbf{D}_{R^*}^{1/2}\mathbf{D}_R^{-1/2} \tag{44}$$

Consider the hth power of (40):

$$\begin{aligned}
&\mathbf{V}_{sim}\Lambda^h\mathbf{V}_{sim}^{-1} = \mathbf{D}_R^{1/2}\mathbf{V}\Lambda^h\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \Leftrightarrow \\
&\Leftrightarrow \mathbf{D}_R^{-1/2}\mathbf{V}_{sim}\Lambda^h\mathbf{V}_{sim}^{-1}\mathbf{D}_R^{1/2} = \mathbf{V}\Lambda^h\mathbf{V}^{-1}
\end{aligned} \tag{45}$$

Using the eigenvector relationships in (33) and (43) yields:

$$\mathbf{D}_R^{-1/2}\left(\mathbf{D}_R^{1/2}\mathbf{V}\right)\Lambda^h\left\{\left[\left(\mathbf{V}^*\right)^{\intercal}\mathbf{D}_{R^*}^{1/2}\right]\mathbf{D}_{R^*}^{1/2}\mathbf{D}_R^{-1/2}\right\}\mathbf{D}_R^{1/2} = \mathbf{V}\Lambda^h\mathbf{V}^{-1} \tag{46}$$

Note that $\mathbf{V}_{sim} = \mathbf{D}_R^{1/2}\mathbf{V}$ because no approximation is required, then the result in (33) holds. Now, as $h \to \infty$:

$$\begin{aligned}
\mathbf{W}^{\infty} &\approx \frac{1}{\|\,\mathbf{D}_R^{1/2}\mathrm{col}(\mathbf{V})_1\|_2\;\|\,\mathrm{col}\left(\mathbf{V}^*\right)_1^{\intercal}\mathbf{D}_{R^*}^{1/2}\|_2} \\
&\quad\times\mathbf{D}_R^{-1/2}\left(\mathbf{D}_R^{1/2}\mathrm{col}(\mathbf{V})_1\mathrm{col}\left(\mathbf{V}^*\right)_1^{\intercal}\mathbf{D}_{R^*}^{1/2}\right)\mathbf{D}_{R^*}^{1/2} \\
&= \left(\sum_{i=1}^{N} d_{R,i}\right)^{-1/2}\left(\sum_{i=1}^{N} d_{R,i}^*\right)^{-1/2} \times \mathbf{J}\mathbf{D}_{R^*}
\end{aligned} \tag{47}$$

where $d_{R,i}^*$ is the sum of the ith row of $\mathbf{W}_0^*$ and $\mathrm{col}\left(\mathbf{V}^*\right)_1 = \boldsymbol{\iota}$ because $\mathbf{W}^*$ is row-normalized. As before, the "long run" matrix, $\mathbf{W}^{\infty}$, is rescaled due to the orthogonalization of the first eigenvector of $\mathbf{W}$. In this case, the geometric mean between the sum of all rows of $\mathbf{W}_0$ and $\mathbf{W}_0^*$ is used. The remaining results are straightforward.

5. GMM estimation with approximated gradients

The estimation of model (5) is addressed through a variant of the iterative GMM estimator of Klier and McMillen (2008). The iterative procedure deduced in Section 2 is used and the N-dimensional matrix operations from the individual gradients (10) and (11) are approximated, considering the new method presented in Section 4. Under this approach, it is no longer required to compute the inverse of the spatial lag operator and related matrix operations on each iteration. Also, it is possible to deduce approximate closed formulas for the elements of the approximated matrices. In this way, the overall computational complexity and the computational time of the estimation is significantly reduced.

As in Section 3.2, consider the matrix $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I} - \alpha\mathbf{W})^{-1}$. Also, consider the matrices $\mathbf{\Upsilon}$ and $\mathbf{\Sigma}$, to derive the approximation for their diagonal elements.

Focusing on $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I} - \alpha\mathbf{W})^{-1}$, consider the corresponding series expansion and replace the powers $h \geq 2$ of $\mathbf{W}$ by $\mathbf{W}^{\infty}$. This yields:

$$\begin{aligned}
(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{W}(\mathbf{I} - \alpha\mathbf{W})^{-1} &= \mathbf{W} + 2\alpha\mathbf{W}^2 + 3\alpha^2\mathbf{W}^3 + \cdots \\
&\quad + (q+1)\,\alpha^q\mathbf{W}^{q+1} + \ldots \\
&\approx \mathbf{W} + 2\alpha\mathbf{W}^{\infty} + 3\alpha^2\mathbf{W}^{\infty} + \cdots \\
&\quad + (q+1)\,\alpha^q\mathbf{W}^{\infty} + \ldots \\
&= \mathbf{W} + \left(\frac{1}{(1-\alpha)^2} - 1\right)\mathbf{W}^{\infty}
\end{aligned} \tag{48}$$

In this way, the diagonal of $\mathbf{\Upsilon}$ is approximately equal to:

$$\begin{aligned}
\mathrm{diag}\,(\Upsilon) \approx 2 \times \mathrm{diag}\Bigg(\sum_{j=1}^{N}\Bigg(\Bigg[\mathbf{W} + \left(\frac{1}{(1-\alpha)^2} - 1\right)\mathbf{W}^{\infty}\Bigg] \circ \\
\left[\mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right]\Bigg)_{i,j}\Bigg)
\end{aligned} \tag{49}$$

After some algebra (49) simplifies to:

[4] Note that if (38) is multiplied by a matrix or vector with zero mean, the proposed approximation gives the same result as the linear transformation $\mathbf{I} + \alpha\mathbf{W}$.

$$\Upsilon_{ii} \approx \frac{(2-\alpha)2\alpha}{(1-\alpha)^2} w_{i,i}^{\infty} + 2\alpha \sum_{j=1}^{N} w_{i,j}^2 + \frac{(3-2\alpha)2\alpha^2}{(1-\alpha)^2} \sum_{j=1}^{N} w_{i,j} w_{1,j}^{\infty} + \frac{(2-\alpha)2\alpha^3}{(1-\alpha)^3} \sum_{j=1}^{N} (w_{i,j}^{\infty})^2 \tag{50}$$

for $i = 1, 2, \ldots, N$. Note that, because the row vectors of $\mathbf{W}^{\infty}$ are all equal, the Hadamard product between $\mathbf{W}$ and $\mathbf{W}^{\infty}$ is simplified to the element-wise product between a matrix and a row vector.

Lastly, the diagonal elements of $\boldsymbol{\Sigma}$ are equal to the row sums of the Hadamard square of the spatial lag operator inverse:

$$\operatorname{diag}(\boldsymbol{\Sigma}) \approx 2 \times \operatorname{Var}(\boldsymbol{\xi} \mid \mathbf{X}, \mathbf{W}) \times \operatorname{diag}\left(\sum_{j=1}^{N} \left(\left[\mathbf{I} + \alpha \mathbf{W} + \frac{\alpha^2}{1-\alpha} \mathbf{W}^{\infty} \right]^{\circ 2} \right)_{i,j} \right) \tag{51}$$

or simply:

$$\sigma_i^2 \approx \sigma_{\xi}^2 \left[2 + \frac{4\alpha^2}{1-\alpha} (w_{i,i}^{\infty})^2 + 2\alpha^2 \sum_{j=1}^{N} w_{i,j}^2 + \frac{4\alpha^3}{1-\alpha} \sum_{j=1}^{N} w_{i,j} w_{1,j}^{\infty} + \frac{2\alpha^4}{(1-\alpha)^2} \sum_{j=1}^{N} (w_{i,j}^{\infty})^2 \right] \tag{52}$$

where $\sigma_{\xi}^2 = \operatorname{Var}(\boldsymbol{\xi} \mid \mathbf{X}, \mathbf{W})$, for all $i = 1, 2, \ldots, N$. In this way, the approximate expression for the non-constant variances and related quantities are obtained with minimal computational requirements. Also, these quantities can now be used and interpreted in a meaningful way.

6. Monte Carlo simulations

In this section, a set of Monte Carlo experiments are presented. The explicit approximation method based on known matrices (AMBKM) is compared with the methods presented in Section 3 (the Taylor series approximation, the Chebyshev approximation, the eigendecomposition of the spatial weighting matrix, the LU decomposition and the conjugate gradient method) in terms of the accuracy to approximate the inverse of the spatial lag operator, $(\mathbf{I} - \alpha\mathbf{W})^{-1}$, the diagonal elements of the matrix $\boldsymbol{\Upsilon}$ and the matrix-vector product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{X}$. Also, 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 GMM estimator of the linearized spatial lag model for binary choice outcomes (LGMM) –, in terms of bias, root mean squared errors and computational time. A variety of simulation designs are considered, with particular interest on the adequacy of the proposed procedures to large samples frameworks.

6.1. Simulation design

The binary dependent variable is constructed following the setting of Klier and McMillen (2008). Consider the simplified version of the model (5) with a single explanatory variable. The explanatory variable, $\mathbf{X}$, is randomly drawn, for each unit, from a $\mathcal{U}(-1, 1)$ distribution. Under a Probit specification, the probability of success is given by:

$$P_i = \Phi\left(\frac{\beta_0 x_{1i}^{\#}}{\sigma_i} + \frac{\beta_1 x_{2i}^{\#}}{\sigma_i} \right), \qquad i = 1, 2, \ldots, N \tag{53}$$

where $\Phi(\cdot)$ is the standard Normal CDF, $x_{1,i}^{\#}$ is the ith row of the matrix product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\boldsymbol{\iota}$ and $x_{2,i}^{\#}$ is the ith row of the matrix product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{X}$. The scalars σ_i are the square root of the diagonal elements of the matrix $\left[(\mathbf{I} - \alpha\mathbf{W})^{\intercal}(\mathbf{I} - \alpha\mathbf{W})\right]^{-1}$. The observed dependent variable, Y_i, is defined as $Y_i = 1$ if $e_i \leq P_i$ and $Y_i = 0$ otherwise, where e_i is randomly drawn, for each unit, from a $\mathcal{U}(0, 1)$ distribution.

The working spatial weighting matrix, $\mathbf{W}$, is constructed according to a two stage setting. In the first stage, the N spatial units are randomly drawn points in the unit square. In the second stage, based on a distance criteria (radial distance or nearest neighbor), an initial spatial weighting matrix, $\mathbf{W}_0$, is constructed and row normalized afterwards. For the case where $\mathbf{W}_0$ is based on the radial distance criterion, the maximum distance to the closest neighbor is computed and a multiplicative factor, δ_R, is used to determine the maximum distance such that the unit j is considered to be a neighbor of unit i, for all $i, j = 1, 2, \ldots, N$. For the case where $\mathbf{W}_0$ is based on the nearest neighbor criterion, the number of nearest neighbors is given by $\delta_{NN} \times N$, where δ_{NN} is the matrix density (the complement of sparsity), the proportion of non-zero elements in $\mathbf{W}$. In this way, the large sample properties of the proposed procedures 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 and the number of neighbors, for each spatial unit, remains fixed, as $N \to \infty$. The latter corresponds to a scenario where new observations are added between the existing ones and a bounded area becomes dense (Anselin, 2007). Also, it is important to note that, under the radial distance criterion, $\mathbf{W}_0$ is symmetric, while under the nearest neighbor criterion, $\mathbf{W}_0$ is non-symmetric. Therefore, simulations are performed to assess the adequacy of the AMBKM when the assumption of symmetry is not valid.

The Monte Carlo experiments are conducted for each design of $\mathbf{W}$ and for each GMM estimator, as well. The number of spatial units, N, vary over the set $\{100, 1000, 2000\}$ and the spatial lag parameter takes on values $\alpha \in \{0, 0.2, 0.5, 0.8\}$. For the case where $\mathbf{W}$ is based on the radial distance criterion, δ_R vary over the restricted set $\{1, 2, 4\}$. For the case where $\mathbf{W}$ is based on the nearest neighbor criterion, δ_{NN} vary over the restricted set $\{0.01, 0.1, 0.2\}$. In this way, the number of neighbors is approximately equal for the two criteria. The regression parameters are held fixed at $\beta_0 = 0$ and $\beta_1 = 1$ and the matrix of instruments used in all estimation procedures is $\mathbf{Z} = [\mathbf{X}\ \mathbf{WX}\ \mathbf{W}^2\mathbf{X}\ \mathbf{W}^3\mathbf{X}]$. For each experiment, 1000 replications are used. The experiments were performed in a Linux based server, with 64 GB of RAM and composed by 24 AMD Opteron CPUs, ranging from 0.8 GHz to 2.1 GHz.

For each set of experiments per approximation method, the accuracy of the approximated spatial lag operator inverse is summarized in terms of the average relative norm difference w.r.t. the identity matrix:

$$\frac{1}{1000} \sum_{r=1}^{1000} \frac{\left\| (\mathbf{I} - \alpha\mathbf{W})(\mathbf{I} - \alpha\mathbf{W})_{approx}^{-1} - \mathbf{I} \right\|_2}{\|\mathbf{I}\|_2} \tag{54}$$

while the accuracy of the approximated diagonal elements of the matrix $\boldsymbol{\Upsilon}$ is summarized in terms of the average relative norm difference w.r.t. the true values:

$$\frac{1}{1000} \sum_{r=1}^{1000} \frac{\left\| \operatorname{diag}(\Upsilon)_{approx} - \operatorname{diag}(\Upsilon) \right\|_2}{\|\operatorname{diag}(\Upsilon)\|_2} \tag{55}$$

The accuracy of the approximation of the matrix-vector product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{X}$ is summarized by the average correlation coefficient between the approximated and the true values of the resulting vector.

For each set of experiments per GMM estimator, the estimates of the regression parameters, $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\alpha}$ are reported, as well as three computational indicators: time per loop (in seconds), number of iterations and total time (in seconds). The parameter estimates are summarized by both the mean and the root mean squared error (RMSE), while the computational indicators are summarized only by the mean. Also, for the case where $\mathbf{W}$ is based on the radial distance criterion, the number of neighbors is reported and summarized by the mean, while for the case where $\mathbf{W}$ is based on the nearest neighbor criterion, the percentage of asymmetric neighbors is reported and summarized by the mean.

The calculations were performed using R and the package McSpatial from McMillen (2013).

6.2. Results

The results of the Monte Carlo experiments are presented in Appendix A. The simulation results on the accuracy of the approximation methods are detailed in Appendix A.1 and the simulation results on the statistical and computational properties of the GMM estimators are detailed in Appendix A.2. Also, the simulation results are organized according to the criteria chosen to construct the spatial weighting matrix, $\mathbf{W}$, and according to the true values of α.

The accuracy of the approximations considerably relies on the true values of α. For $\alpha = 0$, the approximations are trivial. However, as α becomes close to unity, their accuracy worsens. In particular, the accuracy of AMBKM rapidly deteriorates for $\alpha \geq 0.5$. This highlights the fact that the weight of the infinite higher-order term that is neglected (or approximated), in the series expansion of the inverse – see (15) and (17) –, becomes larger as $\alpha \to 1$. In this way, the higher-order term is more informative to the approximations at moderate and high levels of spatial dependence.

In addition, there is a slight improvement in the accuracy of the approximations as $\mathbf{W}$ becomes dense (δ_R and δ_{NN} are increasing), for a fixed N. This happens because the magnitude of each element of the spatial lag operator inverse and related matrices, including the matrix $\boldsymbol{\Upsilon}$, is smaller for a denser $\mathbf{W}$. Also, since $\| \alpha^h \mathbf{W}^h \| \leq |\alpha|^h \| \mathbf{W} \|^h = |\alpha|^h$, the elements of the hth term of the series expansion of the inverse are bounded by $|\alpha|^h$, for any $\mathbf{W}$ satisfying Assumption 5. Therefore, as $\mathbf{W}$ becomes dense, elements with progressively smaller magnitudes are added to the series expansion of the inverse. Hence, for the case where $\mathbf{W}$ is sparse ($\delta_R = 1$ and $\delta_{NN} = 0.01$), the average relative norm differences are w.r.t. large values, while for the case where $\mathbf{W}$ is dense ($\delta_R = 4$ and $\delta_{NN} = 0.2$), the average relative norm differences are w.r.t. small values.

It should be noted that, for the reasons discussed above when $\mathbf{W}$ is sparse, using the approximated matrices in the gradients of the GMM estimation procedure, may reduce the accuracy of the estimates of α, since the term involving Υ_{ii}, in the gradient of α (see equation (11)), dominates the expression.

With regard to the approximation of the matrix-vector product $(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{X}$, the simulations show that the average correlation coefficient between the approximated and the true matrix-vector product is, in general, approximately equal to 1. However, when $\alpha = 0.8$, the average correlation coefficient deteriorates as $\mathbf{W}$ becomes sparse (δ_R and δ_{NN} are decreasing). This is particularly obvious for the Taylor series approximation and the AMBKM, where the minimum average correlation coefficient is equal to 0.987 and 0.887, respectively, corresponding to the case where $N = 100$. For the case where $N \geq 1000$, the minimum average correlation coefficient becomes equal to 0.990 and 0.937, respectively. These results emphasize, once again, the issues related to the accuracy of the approximations under the scenarios where $\mathbf{W}$ is sparse and the degree of spatial dependence is high.

In terms of computational time, for $N = 100$, all the approximation methods are fairly quick. However, as $N \to \infty$, the computational time associated with the eigendecomposition, the LU decomposition and the conjugate gradient method clearly increases, in comparison to the remaining methods, since they involve matrix operations that become computationally burdensome for large N. Considering the eigendecomposition, the full eigensystem and an N-dimensional matrix product have to be computed. For the LU decomposition, $(\mathbf{I} - \alpha\mathbf{W})$ has to be factored. For the conjugate gradient method, an N-dimensional matrix-vector has be computed on each iteration. Similarly, the computational demand associated with the Taylor series approximation and the Chebyshev approximation tends to increase, as $\mathbf{W}$ becomes dense. This is because the first four powers of $\mathbf{W}$ need to be computed. To the contrary, the computational time of the AMBKM is much less sensitive to the size and density of $\mathbf{W}$, since it involves a simple summation of known matrices.

Despite the simulation results showing that, under very specific scenarios, the AMBKM produces larger average relative norm differences and less correlated approximations w.r.t. the true operation, these effects are mitigated when regarding estimation. Furthermore, the AMBKM is the approximation method that requires minimal computational time (less than a second) to recover the quantities of interest and allows to approximate the partial effects. For these reasons, the AMBKM is particularly useful when iterative procedures have to be used to estimate spatial binary choice models with large samples and dense spatial weighting matrices.

Now focusing on the performance of the GMM estimators, the results are, in general, consistent with the previous findings in the literature (see Billé, 2013; Calabrese and Elkink, 2014; Klier and McMillen, 2008). The estimates of the regression parameters, $\widehat{\beta}_0$ and $\widehat{\beta}_1$, are extremely accurate, except for $\alpha = 0.8$. In that case, they exhibit a small bias (a downward bias for the iGMMa and the LGMM estimators and an upward bias for the iGMM estimator) that tends to vanish as $N \to \infty$ and $\mathbf{W}$ becomes dense.

The estimates of the spatial lag parameter, $\widehat{\alpha}$, are far more open to discussion, since its accuracy is simultaneously affected by the true value of the parameter, the sample size and the density of $\mathbf{W}$.

For $\alpha \leq 0.5$ and a fixed N, as $\mathbf{W}$ becomes dense, the iGMM estimator exhibits a significant growing downward bias, whereas the LGMM and iGMMa estimators are much less biased. The only exceptions are for $\alpha = 0.5$ and $\alpha = 0$, where the LGMM and iGMMa estimators exhibit a growing upwards bias, respectively. The decreased accuracy of the LGMM estimator, at moderate and high levels of spatial dependence, is expected, considering the existing simulation studies. The spurious spatial dependence estimated by the iGMMa estimator, for the case where $\alpha = 0$, and the biases displayed by the iGMM estimator evidence that, in general, the spatial GMM estimators can be severely distorted under infill asymptotics (fixed N, denser $\mathbf{W}$). This is especially obvious when $N = 100$. See Lahiri (1996) for a discussion on this matter.

For $\alpha \leq 0.5$ and a fixed density of $\mathbf{W}$, as $N \to \infty$, the iGMM estimator exhibits a downward bias that tends to decline more rapidly when $\mathbf{W}$ is sparse. The LGMM and the iGMMa estimators typically exhibit an upwards bias. For $0 \leq \alpha \leq 0.2$, this bias tends to vanish more rapidly when $\mathbf{W}$ is sparse, while for $\alpha = 0.5$, it tends to vanish more rapidly when $\mathbf{W}$ is dense. Note that, here, both infill and increasing domain asymptotics appear to operate. This implies that the rate of convergence for the various parameters can be different and possibly slower than $\sqrt{N}$, as argued by Lee (2004).

The case of $\alpha = 0.8$ is of particular relevance, since all the spatial GMM estimators exhibit a significant upward bias. Recall that, under the GMM framework, consistency relies on the validity of moment conditions, that use only the information in marginal distributions. However, other estimation methods that consider the joint dependence structure of the spatial data in the estimation, typically perform better at high levels of spatial dependence. In fact, this corroborates with the simulation results of Billé (2013) and Calabrese and Elkink (2014).

Nevertheless, there are two important results regarding the accuracy of the iGMMa estimator that should be emphasized. First, for $N \geq 1000$ and as $\mathbf{W}$ becomes dense, $\widehat{\alpha}$ is better estimated when using the iGMMa estimator, especially for the case where $\mathbf{W}$ is based on the nearest neighbor criterion.[5] This suggests that, for the iGMMa estimator, the number of neighbors for each spatial unit can diverge to infinity at a faster rate than that of the iGMM estimator, without compromising consistency (see Lee, 2004). Second, for $\alpha \geq 0.5$, the iGMMa estimator is typically less biased than the other spatial GMM estimators. The only

[5] Under the nearest neighbor criterion and using the AMBKM, $\mathbf{W}$ is based on a symmetrized version of an initial spatial weighting matrix, which implies that the number of neighbors for each spatial unit necessarily increases.

exception is for $\alpha = 0.8$ and $N = 2000$, where $\widehat{\alpha}$ is better estimated when using the iGMM estimator, but $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are better estimated when using the iGMMa estimator, especially if **W** is not sparse ($\delta_R > 1$ or $\delta_{NN} > 0.01$).

With regard to the RMSEs of the estimated parameters, the simulation results show that the RMSEs of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are substantially smaller than the RMSEs of $\widehat{\alpha}$.

For a fixed N, as **W** becomes dense, all the RMSEs increase. In particular, when $N = 100$, the RMSEs of $\widehat{\alpha}$ largely increase. These facts evidence, once again, how the estimates can be severely distorted under infill asymptotics.

For a fixed density of **W**, as $N \to \infty$, the RMSEs of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ decrease, whereas the RMSEs of $\widehat{\alpha}$ exhibit a fairly different behavior considering the criteria chosen for **W**. For the case where **W** is based on the nearest neighbor criterion, the RMSEs of $\widehat{\alpha}$ increase. For the case where **W** is based on the radial distance criterion, the RMSEs of $\widehat{\alpha}$ decrease.

Additionally note that the RMSEs of the iGMMa estimator are typically smaller than the RMSEs of the remaining spatial GMM estimators. In particular, the RMSEs of $\widehat{\alpha}$ in the iGMMa estimator are substantially smaller than in both the iGMM and LGMM estimators, even for the case where $\alpha = 0.8$.

In terms of the computational ability associated with the spatial GMM estimators, measured by the average computational time required to produce estimates for the parameters of interest, it strongly relies on the sample size and on the density of **W**. As $N \to \infty$ and **W** becomes dense, the average computational time increases. In particular, for $\alpha = 0.8$, the average computational time is even larger, since the spatial GMM estimators require, on average, 1 to 2 additional iterations to converge. This is because the inverse of the spatial lag operator is approaching singularity and the computation of the gradients becomes troublesome.

When $N = 2000$ and **W** is dense, the average computational time of the iGMMa estimator is about 3–6 times less than that of the iGMM estimator, depending whether **W** is based on the radial distance criterion or on the nearest neighbor criterion, respectively. Also, the iGMMa estimator is typically less biased than the iGMM estimator, especially for $\alpha \leq 0.5$.

The average computational time of the LGMM estimator is clearly impossible to overcome. However, the iGMMa estimator proves its ability to estimate $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\alpha}$ with more accuracy, more precision and in a reasonable amount of time, even when the true value of α is close to unity.

7. Empirical application

In this section, an empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas (U.S. MSAs) is presented. The adequacy of the previous GMM estimators to real data is assessed and compared.

The strategies to promote and/or to improve competitiveness at the regional and country level are currently centering the attention of policy makers. However, the definition of 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, 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. Then, in a broad sense, competitiveness is considered a measure for economic performance. Moreover, while promoting competitiveness, the possible environmental impacts cannot be disregarded.

The relationship between environmental degradation and economic growth has been extensively studied in the literature and hypothesized as an "U"-shaped relationship, the so-called Environmental Kuznets Curve (EKC) hypothesis (Grossman and Krueger, 1991; Panayotou, 1993; Shafik and Bandyopadhyay, 1992). However, the EKC hypothesis is not free of criticism, mainly due to the shape of the relationship and the lack of empirical evidence. Also, as Porter et al. (2015) points out, the promotion of efficient energy infrastructures and a low-carbon transition may help to improve competitiveness. In fact, this consists in an inversion of the EKC hypothesis, yet to be tested empirically. Most of the applied works focus on the analysis of competitiveness and environmental quality as separate subjects and only a few consider the 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 the EKC and on the spatial spillovers associated with the emission of air pollutants. Furthermore, none of the previous works estimate a spatial model with binary dependent variables.

Here, the analysis of the environmental effects over the competitiveness in the U.S. MSAs is addressed. A combined dataset of socioeconomic data and environmental data from the U.S. Bureau of Economic Analysis (BEA) and the U.S. Environmental Protection Agency (EPA), respectively, is used. This dataset contains information about the GDP, labor costs, price indexes, dividends, total employment and population, as well as, information about the annualized Air Quality Index (AQI) and for five main air pollutants – ground-level ozone (O_3), particle pollutants ($PM_{2.5}$ and PM_{10}), carbon monoxide (CO), sulfur dioxide (SO_2) and nitrogen dioxide (NO_2). The U.S. MSAs that are included in this analysis correspond to the continental MSAs that continuously report information for the previous variables, between 2001 and 2016 ($N = 4,848$).

As previously mentioned, there are numerous ways to define competitiveness. Because it is difficult to provide a clear interpretation or to have precise units of measurement, competitiveness can be considered a latent variable. Therefore, the many existing proxies to measure competitiveness can be used to define a new indicator. In this way, a Binary Competitiveness Indicator (BCI) is proposed. A given Metropolitan Statistical Area (MSA) is defined as competitive if, simultaneously, (1) its employment-to-population ratio is greater than the employment-to-population ratio in the combined area of the excluded MSAs and the non-MSAs; (2) its GDP per capita is greater than the GDP per capita in the combined area of the excluded MSAs and the non-MSAs; (3) its Unit Labor Costs (the cost of labor per unit of output) are less than the Unit Labor Costs in the combined area of the excluded MSAs and the non-MSAs or the Unit Capital Costs (the cost of capital per unit of output) are less than the Unit Labor Costs in the combined area of the excluded MSAs and the 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.

In Table B.1 the descriptive statistics for the variables included in this study are presented. Considering the BCI, about 15% of the U.S. MSAs are labeled as competitive. The variables AQI_{min} and AQI_{max} are, respectively, the minimum and maximum annual values 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 $\%daysO_3$, $\%daysPM_{2.5}$, $\%daysPM_{10}$, %daysCO, $\%daysSO_2$ and $\%daysNO_2$ correspond to the percentage of days that the observed value of the AQI was determined by the concentration levels of each pollutant. On average, O_3 and $PM_{2.5}$, by a large amount, the most important contributors for the observed values of the AQI, in this sample. The variables %daysAboveModerate and %daysExceptionalEvents correspond, respectively, to the percentage of days that the observed value of the AQI was above 0.51 and to the percentage of days that the observed value of the AQI was affected by "exceptional events" (wildfires or other natural disasters).

A spatial lag Probit is applied to the pooled sample of the U.S. MSAs to study the effects of the environmental quality indicators over the spatially lagged BCI. The spatial weighting matrix, **W**, is block-diagonal and based on the radial distance criterion with a distance threshold equal to 1, according to the pattern of proximity displayed in Fig. B.1. Under this specification for **W**, the spurious spatial interactions are controlled, because only the closest U.S. MSAs are considered to be neighbors. Two models are estimated: the unrestricted model and the restricted model. The first model considers a general specification, based on the available information on the pollutants and air quality and assuming that there is a quadratic relationship between the AQI and the BCI. The second model is a restricted version of the first model, focusing on statistically significant effects.

In Table B.2 the estimation results for the previous models are presented, considering the three GMM estimators (iGMMa, iGMM and LGMM). The instruments $\mathbf{Z} = [\mathbf{X}\ \mathbf{WX}\ \mathbf{W}^2\mathbf{X}\ \mathbf{W}^3\mathbf{X}]$ were used in all estimation procedures. Also, time effects were added and the Mundlak (1978) approach was used to filter the eventual dependence between the unobserved regional effects and the explanatory variables. The estimation routines[6] are based on the R package McSpatial of McMillen (2013).

In general, the estimates for the unrestricted model are quite poor in terms of statistical significance, except for the linear and quadratic effects of AQI_{min} and for the estimate for the spatial lag parameter. In fact, the estimated signs for the coefficients of AQI_{min} are of particular interest, due to the statistical evidence towards the existence of an "U"-shaped effect. Also, the estimates for the spatial lag parameter reveal that there may be a high degree of spatial dependence in the sample. However, because the Wald test rejects the null of overall significance, the robustness of the previous results to the exclusion of several statistically insignificant variables should be checked. In fact, the test for exclusion restrictions allowed to considerably simplify the initial specification to a more parsimonious one. In the new specification, only the linear and quadratic AQI_{min}, $\%daysO_3$ and $\%daysPM_{2.5}$ remained. Interestingly, these variables are also used in the applied literature (Millimet et al., 2003; Rupasingha et al., 2004).

The estimates for the restricted model are now individually and jointly statistically significant (except for the variable $\%daysPM_{2.5}$, in the iGMM estimation). Most importantly, the magnitude of the estimates do not change much, in comparison to the unrestricted model. Therefore, based on the previously noted "U"-shaped effect of AQI_{min}, there is evidence towards the idea that the implementation of environmental-friendly policies may initially involve substantial conversion costs, penalizing regional competitiveness, but, at some point, those costs can be transformed into development opportunities based on new services or products, with large benefits to the economy as a whole. This follows along the lines of Porter et al. (2015) and it is referred as a "win-win path". Nevertheless, some ambiguity may arise concerning the positive estimated signs for the variables $\%daysO_3$ and $\%daysPM_{2.5}$. However, note that, for the case where environmental-friendly policies are implemented and the air quality is actually improved, the observed values for the AQI can still be determined by the concentration levels of the previous pollutants. Recall that O_3 and $PM_{2.5}$ largely contribute to the observed values of the AQI. Lastly, having estimates for the spatial lag parameter above 0.7 is evidence towards the importance of the spillover effects over regional competitiveness. This emphasizes the idea that regional policies do benefit the neighboring areas, regarding their economic efficiency.

From the estimation of the previous models, both the iGMMa and iGMM estimators exhibit a quite similar performance, based on Hansen tests and on three measures of goodness-of-fit: the McFadden R^2, the squared correlation coefficient between the observed and the predicted values – $\rho^2(Y, \widehat{Y})$ – and the percentage of the correctly predicted observations – $\%(\widehat{Y} = Y)$. The adequacy of the moment conditions is not rejected and the predictive power is quite noticeable. This contrasts with the performance of the LGMM estimator, where the Hansen tests reject the null of correct moment conditions and the McFadden R^2 is persistently negative, displaying a very poor fit to the data. In terms of computational time, the iGMMa estimator clearly outruns the iGMM estimator. In this way, the iGMMa estimator proves to be a feasible and an adequate alternative to estimate spatial binary choice models using real data.

[6] Available upon request.

8. Conclusions

In this paper a new approximation method based on known matrices (AMBKM) is proposed. It addresses the computational issues related to the GMM estimation of spatially lagged models for binary dependent variables. Focusing on the inversion of the spatial lag operator, a simple and intuitive approximation is deduced and applied to approximate other related *N*-dimensional matrix operations. It is demonstrated that, these matrices are approximated by known matrices and simple matrix-vector operations. Furthermore, it is demonstrated that closed formulas for the elements of the approximated matrices can be easily deduced.

The proposed AMBKM is based on a set of non-restrictive assumptions that allow to accommodate several frameworks for the spatial weighting matrix. This method is computationally feasible in large samples, because the resulting approximations are based on known matrices, up to an estimated parameter. This is important to note, since it avoids the *N*-dimensional matrix operations required in the alternative approximation methods, which turns them computationally infeasible in large samples. Moreover, it allows to obtain a closed formula to approximate the partial effects, that can be decomposed into three separate effects (regardless a scale factor): the pure direct effects (from **I**), the first order neighboring effects (from **W**) and the "global" effects (from $\mathbf{W}^{\infty}$), which combines the "long run" direct and indirect effects.

This paper also proposes a new GMM estimator based on a modification of the iterative GMM estimator of Klier and McMillen (2008). Aiming at the reduction of the overall computational complexity and the computational time, the approximated matrices are used in the gradients of the new estimation procedure.

Simulations show that the proposed approximation method yields reasonably accurate approximations for the spatial lag operator inverse and related matrices, especially when the spatial weighting matrix is large and dense. Also, the computational time required to obtain these approximations is minimal, regardless the computational complexity of the true operation and the dimension of the spatial weighting matrix.

In addition, the Monte Carlo experiments show that the proposed estimator – the iterative GMM with approximated gradients (iGMMa) –, performs reasonably well in terms of the estimation of the parameters, except for the case where α is close to unity. Nevertheless, for $\alpha \leq 0.5$, the existing biases are attenuated as the spatial weighting matrix becomes large and dense. Also, the iGMMa estimator proved to be surprisingly accurate, for the case where the spatial weighting matrix was based on the nearest neighbor criterion, with a moderate to large number of neighbors. Furthermore, the iGMMa estimator outperformed the benchmark iterative GMM (iGMM) estimator in terms of computational time, accuracy and precision, and outperformed the GMM estimator of the linearized spatial lag model for binary choice outcomes (LGMM) in terms of accuracy and precision. In fact, the iGMMa estimator stood as most precise estimator, even for the case where α is close to unity.

The usefulness of the proposed iGMMa estimator is illustrated in an empirical application that measures the impact of environmental indicators over the competitiveness of the U.S. Metropolitan Statistical Areas. A new Binary Competitiveness Indicator (BCI) is introduced and a spatial lag Probit is estimated, addressing the level of spatial dependence in regional competitiveness. The iGMMa estimator proved to perform as well as the benchmark iGMM estimator, in terms of

predictive power, and outperformed the LGMM estimator. Moreover, in this example, where a large data set is used and several explanatory variables are included in the estimation, the iGMMa estimator proved to be computationally superior to the other spatial GMM estimators.

The performance and attractiveness of the proposed iGMMa estimator in estimating models with spatially lagged binary dependent variables lead to obvious extensions, especially the estimation of models with spatially lagged errors and with higher order spatial lags. The estimation of spatial models for other discrete and censored dependent variables can be addressed by GMM, using the approximated matrices, as well.

All the algorithms used in this paper, the proposed approximation method and the estimation procedures, can be easily implemented using the R package McSpatial from McMillen (2013).

Appendix A. Simulation results

Appendix A.1. Approximation methods for the spatial lag operator inverse

Table A.1.1
Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0$ and $\mathbf{W}$ based on the radial distance criterion.

δ_R		1			2			4		
N		$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.059]	[0.003]	[0.001]	[0.013]	[0.002]	≈[0.000]	[0.016]	[0.004]	≈[0.000]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.063]	[0.056]	≈[0.000]	[0.036]	[0.008]	≈[0.000]	[0.038]	[0.008]	≈[0.000]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.072]	[0.088]	≈[0.000]	[0.029]	[0.033]	≈[0.000]	[0.040]	[0.044]	≈[0.000]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.106]	[0.085]	≈[0.000]	[0.053]	[0.060]	≈[0.000]	[0.068]	[0.075]	≈[0.000]
	Eigen	0.000	–	1.000	0.000	–	1.000	0.000	–	1.000
		[0.010]	[0.005]	≈[0.000]	[0.009]	[0.003]	≈[0.000]	[0.011]	[0.003]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.008]			[0.006]			[0.004]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.073]			[0.009]			[0.010]
1000	True	[0.023]	[0.004]	[0.001]	[0.077]	[0.008]	[0.001]	[0.848]	[0.022]	[0.001]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.309]	[0.151]	[0.001]	[0.216]	[0.156]	[0.001]	[0.412]	[0.165]	[0.002]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.068]	[0.133]	[0.001]	[0.309]	[0.537]	[0.003]	[1.738]	[2.889]	[0.007]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.184]	[0.152]	[0.001]	[0.402]	[0.553]	[0.003]	[2.033]	[2.291]	[0.007]
	Eigen	0.000	–	1.000	0.000	–	1.000	0.000	–	1.000
		[2.189]	[1.102]	[0.005]	[2.241]	[1.100]	[0.005]	[2.164]	[1.099]	[0.005]
	LU			1.000			1.000			1.000
				[0.264]			[0.261]			[0.329]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.428]			[0.428]			[0.429]
2000	True	[0.043]	[0.007]	[0.001]	[0.306]	[0.017]	[0.001]	[3.894]	[0.052]	[0.001]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.675]	[0.606]	[0.001]	[0.796]	[0.620]	[0.001]	[0.943]	[0.615]	[0.003]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.126]	[0.270]	[0.002]	[0.765]	[1.828]	[0.007]	[4.505]	[8.716]	[0.020]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.225]	[0.285]	[0.002]	[1.163]	[1.504]	[0.008]	[5.646]	[6.497]	[0.020]
	Eigen	0.000	–	1.000	0.000	–	1.000	0.000	–	1.000
		[15.669]	[8.442]	[0.018]	[15.534]	[8.442]	[0.018]	[15.797]	[8.430]	[0.018]
	LU			1.000			1.000			1.000
				[1.712]			[1.721]			[1.713]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.048]			[3.048]			[3.048]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag(Υ) are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

Table A.1.2
Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.2$ and $\mathbf{W}$ based on the radial distance criterion.

δ_R		1			2			4		
N		$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.014]	[0.013]	[0.001]	[0.018]	[0.019]	[0.001]	[0.078]	[0.020]	[0.001]
	AMBKM	0.042	0.665	≈1.000	0.028	0.501	≈1.000	0.010	0.236	≈1.000
		[0.026]	[0.010]	[0.001]	[0.027]	[0.009]	≈[0.000]	[0.031]	[0.008]	≈[0.000]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.003	≈1.000
		[0.021]	[0.020]	≈[0.000]	[0.032]	[0.033]	≈[0.000]	[0.039]	[0.046]	≈[0.000]
	Cheb4	≈0.000	0.001	≈1.000	≈0.000	0.003	≈1.000	≈0.000	0.002	≈1.000
		[0.045]	[0.047]	≈[0.000]	[0.115]	[0.064]	≈[0.000]	[0.133]	[0.076]	≈[0.000]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[0.011]	[0.003]	≈[0.000]	[0.010]	[0.003]	≈[0.000]	[0.010]	[0.004]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.004]			[0.004]			[0.005]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.009]			[0.009]			[0.010]
1000	True	[0.267]	[14.103]	[0.008]	[1.003]	[14.350]	[0.008]	[3.221]	[14.903]	[0.008]
	AMBKM	0.041	0.706	≈1.000	0.039	0.703	≈1.000	0.036	0.647	≈1.000
		[0.153]	[0.129]	[0.003]	[0.179]	[0.129]	[0.003]	[0.242]	[0.133]	[0.003]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.052]	[0.093]	[0.001]	[0.425]	[0.779]	[0.004]	[1.541]	[2.663]	[0.007]
	Cheb4	≈0.000	0.002	≈1.000	≈0.000	0.003	≈1.000	≈0.000	0.003	≈1.000
		[0.152]	[0.118]	[0.001]	[0.671]	[0.686]	[0.004]	[1.957]	[2.125]	[0.007]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[2.226]	[1.098]	[0.005]	[2.226]	[1.152]	[0.005]	[2.150]	[1.097]	[0.005]
	LU			1.000			1.000			1.000
				[0.257]			[0.257]			[0.373]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.443]			[0.445]			[0.440]
2000	True	[1.390]	[114.155]	[0.031]	[4.264]	[115.969]	[0.031]	[21.177]	[118.638]	[0.031]
	AMBKM	0.041	0.712	≈1.000	0.040	0.724	≈1.000	0.038	0.685	≈1.000
		[0.547]	[0.528]	[0.012]	[0.601]	[0.584]	[0.013]	[0.785]	[0.532]	[0.013]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.133]	[0.291]	[0.002]	[0.602]	[1.404]	[0.006]	[4.295]	[8.630]	[0.020]
	Cheb4	≈0.000	0.002	≈1.000	≈0.000	0.003	≈1.000	≈0.000	0.003	≈1.000
		[0.288]	[0.299]	[0.002]	[0.938]	[1.218]	[0.006]	[5.628]	[6.445]	[0.020]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[17.567]	[8.442]	[0.018]	[15.495]	[8.437]	[0.018]	[15.782]	[8.440]	[0.018]
	LU			1.000			1.000			1.000
				[1.704]			[1.730]			[1.710]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.116]			[3.117]			[3.100]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag(Υ) are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

Table A.1.3
Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.5$ and $\mathbf{W}$ based on the radial distance criterion.

δ_R		1			2			4		
N		$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.014]	[0.014]	[0.001]	[0.017]	[0.017]	[0.001]	[0.020]	[0.020]	[0.001]
	AMBKM	0.258	0.786	0.995	0.208	0.630	0.999	0.108	0.260	≈1.000
		[0.026]	[0.009]	≈[0.000]	[0.027]	[0.008]	≈[0.000]	[0.086]	[0.008]	≈[0.000]
	Taylor4	0.038	0.110	≈1.000	0.033	0.089	≈1.000	0.032	0.118	≈1.000
		[0.023]	[0.023]	≈[0.000]	[0.028]	[0.034]	≈[0.000]	[0.037]	[0.041]	≈[0.000]
	Cheb4	0.006	0.018	≈1.000	0.004	0.032	≈1.000	0.003	0.014	≈1.000
		[0.046]	[0.049]	≈[0.000]	[0.054]	[0.059]	≈[0.000]	[0.066]	[0.072]	≈[0.000]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[0.010]	[0.003]	≈[0.000]	[0.011]	[0.060]	≈[0.000]	[0.009]	[0.003]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.004]			[0.005]			[0.004]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.009]			[0.009]			[0.009]
1000	True	[0.283]	[14.053]	[0.008]	[0.897]	[14.280]	[0.008]	[3.143]	[14.939]	[0.008]
	AMBKM	0.257	0.847	0.996	0.247	0.818	0.999	0.228	0.720	≈1.000
		[0.154]	[0.130]	[0.003]	[0.234]	[0.128]	[0.003]	[0.239]	[0.134]	[0.003]
	Taylor4	0.032	0.095	≈1.000	0.032	0.083	≈1.000	0.032	0.087	≈1.000
		[0.056]	[0.103]	[0.001]	[0.285]	[0.606]	[0.003]	[1.330]	[2.575]	[0.007]
	Cheb4	0.006	0.023	≈1.000	0.004	0.035	≈1.000	0.003	0.037	≈1.000
		[0.107]	[0.128]	[0.001]	[0.449]	[0.552]	[0.003]	[1.790]	[2.135]	[0.007]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[2.184]	[1.098]	[0.005]	[2.178]	[1.097]	[0.005]	[2.151]	[1.096]	[0.005]
	LU			1.000			1.000			1.000
				[0.257]			[0.260]			[0.259]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.461]			[0.451]			[0.453]
2000	True	[1.302]	[114.794]	[0.031]	[3.984]	[116.021]	[0.031]	[12.962]	[117.668]	[0.031]
	AMBKM	0.259	0.853	0.996	0.250	0.842	0.999	0.242	0.795	≈1.000
		[0.559]	[0.581]	[0.013]	[0.593]	[0.583]	[0.013]	[0.726]	[0.533]	[0.013]
	Taylor4	0.033	0.092	≈1.000	0.032	0.081	≈1.000	0.032	0.083	≈1.000
		[0.115]	[0.307]	[0.002]	[0.601]	[1.413]	[0.006]	[2.938]	[6.074]	[0.016]
	Cheb4	0.007	0.024	≈1.000	0.004	0.037	≈1.000	0.003	0.038	≈1.000
		[0.204]	[0.259]	[0.002]	[0.929]	[1.217]	[0.006]	[3.965]	[4.699]	[0.016]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[17.265]	[8.441]	[0.018]	[15.486]	[8.438]	[0.018]	[15.746]	[8.436]	[0.018]
	LU			1.000			1.000			1.000
				[1.709]			[1.719]			[1.722]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.191]			[3.153]			[3.154]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag(Υ) are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

Table A.1.4
Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.8$ and $\mathbf{W}$ based on the radial distance criterion.

δ_R		1			2			4		
N		$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.014]	[0.014]	≈[0.000]	[0.016]	[0.017]	[0.001]	[0.020]	[0.020]	[0.001]
	AMBKM	0.663	0.861	0.933	0.550	0.469	0.977	0.206	0.041	0.999
		[0.026]	[0.008]	≈[0.000]	[0.027]	[0.009]	≈[0.000]	[0.030]	[0.008]	≈[0.000]
	Taylor4	0.354	0.696	0.990	0.340	0.636	0.997	0.340	0.747	≈1.000
		[0.021]	[0.020]	≈[0.000]	[0.027]	[0.030]	≈[0.000]	[0.039]	[0.044]	≈[0.000]
	Cheb4	0.128	0.139	0.999	0.094	0.091	0.999	0.069	0.205	≈1.000
		[0.045]	[0.047]	≈[0.000]	[0.054]	[0.113]	≈[0.000]	[0.067]	[0.128]	≈[0.000]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[0.068]	[0.003]	≈[0.000]	[0.009]	[0.005]	≈[0.000]	[0.064]	[0.003]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.004]			[0.004]			[0.005]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.010]			[0.009]			[0.009]
1000	True	[0.290]	[14.017]	[0.008]	[0.790]	[14.391]	[0.008]	[2.992]	[14.871]	[0.008]
	AMBKM	0.659	0.952	0.937	0.636	0.873	0.970	0.579	0.639	0.992
		[0.152]	[0.131]	[0.003]	[0.168]	[0.129]	[0.003]	[0.233]	[0.128]	[0.003]
	Taylor4	0.340	0.645	0.990	0.334	0.602	0.996	0.335	0.613	0.999
		[0.059]	[0.110]	[0.001]	[0.225]	[0.488]	[0.003]	[1.335]	[2.489]	[0.007]
	Cheb4	0.144	0.121	0.999	0.086	0.118	≈1.000	0.081	0.109	≈1.000
		[0.110]	[0.136]	[0.001]	[0.366]	[0.510]	[0.003]	[1.844]	[1.997]	[0.007]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[2.180]	[1.097]	[0.005]	[2.178]	[1.099]	[0.005]	[2.146]	[1.151]	[0.005]
	LU			1.000			1.000			1.000
				[0.258]			[0.255]			[0.260]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.497]			[0.480]			[0.472]
2000	True	[1.188]	[115.265]	[0.031]	[4.058]	[115.905]	[0.030]	[23.791]	[118.997]	[0.031]
	AMBKM	0.661	0.962	0.938	0.641	0.912	0.972	0.605	0.739	0.991
		[0.544]	[0.529]	[0.012]	[0.654]	[0.524]	[0.013]	[0.900]	[0.533]	[0.013]
	Taylor4	0.343	0.633	0.990	0.333	0.593	0.996	0.334	0.602	0.999
		[0.108]	[0.226]	[0.002]	[0.584]	[1.356]	[0.006]	[4.739]	[9.435]	[0.020]
	Cheb4	0.142	0.113	0.999	0.084	0.126	≈1.000	0.077	0.112	≈1.000
		[0.191]	[0.242]	[0.002]	[0.908]	[1.177]	[0.006]	[5.975]	[6.938]	[0.020]
	Eigen	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000
		[18.701]	[8.439]	[0.018]	[15.560]	[8.438]	[0.018]	[15.782]	[8.438]	[0.018]
	LU			1.000			1.000			1.000
				[1.713]			[1.711]			[1.725]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.334]			[3.267]			[3.228]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag(Υ) are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

Table A.1.5
Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.060]	[0.071]	[0.001]	[0.012]	[0.002]	≈[0.000]	[0.013]	[0.002]	≈[0.000]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.062]	[0.056]	≈[0.000]	[0.034]	[0.011]	≈[0.000]	[0.034]	[0.010]	≈[0.000]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.071]	[0.026]	≈[0.000]	[0.025]	[0.027]	≈[0.000]	[0.090]	[0.033]	≈[0.000]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.101]	[0.081]	≈[0.000]	[0.048]	[0.052]	≈[0.000]	[0.111]	[0.118]	≈[0.000]
	Eigen	1.000	–	0.971	0.743	–	0.987	0.725	–	0.991
		[0.014]	[0.002]	≈[0.000]	[0.099]	[0.004]	≈[0.000]	[0.043]	[0.005]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.009]			[0.004]			[0.006]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.074]			[0.067]			[0.010]
1000	True	[0.025]	[0.004]	[0.001]	[0.413]	[0.018]	[0.001]	[1.935]	[0.033]	[0.001]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.261]	[0.154]	[0.001]	[0.286]	[0.216]	[0.001]	[0.381]	[0.160]	[0.002]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.068]	[0.133]	[0.001]	[1.142]	[2.176]	[0.006]	[2.958]	[4.869]	[0.008]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.128]	[0.154]	[0.001]	[1.562]	[1.872]	[0.006]	[3.554]	[3.817]	[0.008]
	Eigen	2.175	–	0.995	1.163	–	0.998	1.206	–	0.999
		[22.666]	[2.106]	[0.012]	[22.468]	[2.104]	[0.012]	[22.258]	[2.167]	[0.012]
	LU			1.000			1.000			1.000
				[0.258]			[0.258]			[0.313]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.425]			[0.426]			[0.426]
2000	True	[0.125]	[0.010]	[0.001]	[11.273]	[0.076]	[0.001]	[17.259]	[0.151]	[0.001]
	AMBKM	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.743]	[0.599]	[0.001]	[1.142]	[0.622]	[0.004]	[1.537]	[0.648]	[0.007]
	Taylor4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.334]	[0.653]	[0.004]	[7.081]	[13.673]	[0.025]	[21.942]	[36.299]	[0.031]
	Cheb4	≈0.000	–	≈1.000	≈0.000	–	≈1.000	≈0.000	–	≈1.000
		[0.470]	[0.610]	[0.004]	[9.009]	[9.972]	[0.025]	[23.981]	[25.574]	[0.031]
	Eigen	2.793	–	0.995	1.275	–	≈1.000	1.306	–	0.999
		[165.543]	[16.423]	[0.045]	[164.393]	[16.438]	[0.046]	[164.121]	[16.440]	[0.046]
	LU			1.000			1.000			1.000
				[1.700]			[1.726]			[1.732]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.038]			[3.042]			[3.051]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag(Υ) are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

Table A.1.6
Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.2$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.012]	[0.002]	≈[0.000]	[0.015]	[0.017]	[0.001]	[0.017]	[0.017]	[0.001]
	AMBKM	0.070	0.674	≈1.000	0.041	0.612	≈1.000	0.035	0.513	≈1.000
		[0.028]	[0.009]	[0.001]	[0.026]	[0.008]	≈[0.000]	[0.028]	[0.009]	≈[0.000]
	Taylor4	0.001	0.004	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.019]	[0.016]	≈[0.000]	[0.024]	[0.025]	≈[0.000]	[0.027]	[0.088]	≈[0.000]
	Cheb4	≈0.000	0.001	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.003	≈1.000
		[0.041]	[0.101]	≈[0.000]	[0.049]	[0.052]	≈[0.000]	[0.055]	[0.059]	≈[0.000]
	Eigen	1.031	0.147	0.961	0.454	0.045	0.995	0.949	0.094	0.991
		[0.011]	[0.002]	≈[0.000]	[0.104]	[0.005]	≈[0.000]	[0.042]	[0.004]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.004]			[0.006]			[0.004]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.009]			[0.009]			[0.009]
1000	True	[0.385]	[14.318]	[0.008]	[2.376]	[14.864]	[0.008]	[5.806]	[15.566]	[0.008]
	AMBKM	0.044	0.714	≈1.000	0.038	0.633	≈1.000	0.032	0.527	≈1.000
		[0.149]	[0.124]	[0.004]	[0.225]	[0.125]	[0.004]	[0.321]	[0.128]	[0.003]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.066]	[0.130]	[0.001]	[1.133]	[2.098]	[0.006]	[3.005]	[4.911]	[0.008]
	Cheb4	≈0.000	0.003	≈1.000	≈0.000	0.004	≈1.000	≈0.000	0.003	≈1.000
		[0.124]	[0.152]	[0.001]	[1.438]	[1.680]	[0.006]	[3.540]	[3.850]	[0.008]
	Eigen	2.111	0.074	0.993	1.389	0.036	0.997	1.169	0.029	0.999
		[22.391]	[2.103]	[0.012]	[22.338]	[2.158]	[0.012]	[22.256]	[2.161]	[0.012]
	LU			1.000			1.000			1.000
				[0.256]			[0.255]			[0.258]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.442]			[0.438]			[0.435]
2000	True	[2.986]	[116.255]	[0.031]	[39.494]	[121.062]	[0.031]	[50.895]	[127.362]	[0.031]
	AMBKM	0.042	0.721	≈1.000	0.037	0.634	≈1.000	0.032	0.526	≈1.000
		[0.567]	[0.512]	[0.013]	[0.905]	[0.528]	[0.012]	[1.322]	[0.581]	[0.012]
	Taylor4	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000	≈0.000	0.002	≈1.000
		[0.332]	[0.651]	[0.004]	[7.047]	[13.587]	[0.025]	[21.848]	[36.129]	[0.031]
	Cheb4	≈0.000	0.003	≈1.000	≈0.000	0.004	≈1.000	≈0.000	0.004	≈1.000
		[0.465]	[0.604]	[0.004]	[8.925]	[9.940]	[0.025]	[24.132]	[25.575]	[0.031]
	Eigen	2.597	0.046	0.996	1.318	0.020	0.999	1.316	0.021	0.999
		[165.286]	[16.469]	[0.045]	[164.427]	[16.479]	[0.046]	[164.123]	[16.473]	[0.046]
	LU			1.000			1.000			1.000
				[1.695]			[1.781]			[1.708]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.118]			[3.151]			[3.096]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag(Υ) are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

Table A.1.7
Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.5$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.012]	[0.002]	≈[0.000]	[0.015]	[0.017]	[0.001]	[0.016]	[0.017]	[0.001]
	AMBKM	0.498	0.831	0.981	0.252	0.690	0.997	0.212	0.525	0.999
		[0.025]	[0.010]	≈[0.000]	[0.026]	[0.009]	≈[0.000]	[0.028]	[0.008]	≈[0.000]
	Taylor4	0.059	0.154	≈1.000	0.035	0.101	≈1.000	0.034	0.101	≈1.000
		[0.019]	[0.016]	≈[0.000]	[0.022]	[0.025]	≈[0.000]	[0.027]	[0.031]	≈[0.000]
	Cheb4	0.021	0.022	≈1.000	0.005	0.026	≈1.000	0.004	0.030	≈1.000
		[0.041]	[0.043]	≈[0.000]	[0.047]	[0.053]	≈[0.000]	[0.054]	[0.062]	≈[0.000]
	Eigen	1.360	0.144	0.969	0.712	0.015	0.996	0.837	0.014	0.992
		[0.014]	[0.058]	≈[0.000]	[0.098]	[0.004]	≈[0.000]	[0.043]	[0.004]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.005]			[0.005]			[0.004]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.019]			[0.009]			[0.009]
1000	True	[0.374]	[14.235]	[0.008]	[2.472]	[14.851]	[0.008]	[5.844]	[15.611]	[0.008]
	AMBKM	0.273	0.851	0.996	0.236	0.704	≈1.000	0.201	0.541	≈1.000
		[0.150]	[0.127]	[0.004]	[0.228]	[0.184]	[0.003]	[0.321]	[0.181]	[0.004]
	Taylor4	0.036	0.097	≈1.000	0.033	0.090	≈1.000	0.034	0.097	≈1.000
		[0.067]	[0.130]	[0.001]	[1.076]	[2.146]	[0.006]	[3.000]	[4.861]	[0.008]
	Cheb4	0.005	0.028	≈1.000	0.004	0.038	≈1.000	0.004	0.036	≈1.000
		[0.178]	[0.152]	[0.001]	[1.492]	[1.682]	[0.006]	[3.542]	[3.860]	[0.008]
	Eigen	1.977	0.013	0.991	1.115	0.005	0.999	0.885	0.006	0.999
		[22.299]	[2.101]	[0.012]	[22.355]	[2.104]	[0.012]	[22.282]	[2.156]	[0.012]
	LU			1.000			1.000			1.000
				[0.253]			[0.255]			[0.253]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.461]			[0.447]			[0.447]
2000	True	[3.133]	[116.241]	[0.032]	[39.723]	[121.320]	[0.031]	[50.680]	[127.814]	[0.031]
	AMBKM	0.262	0.851	0.998	0.235	0.704	≈1.000	0.202	0.541	≈1.000
		[0.627]	[0.521]	[0.012]	[0.905]	[0.520]	[0.012]	[1.308]	[0.532]	[0.012]
	Taylor4	0.034	0.088	≈1.000	0.033	0.089	≈1.000	0.034	0.096	≈1.000
		[0.278]	[0.659]	[0.004]	[7.169]	[13.616]	[0.025]	[21.948]	[36.176]	[0.031]
	Cheb4	0.005	0.035	≈1.000	0.004	0.039	≈1.000	0.004	0.037	≈1.000
		[0.522]	[0.605]	[0.004]	[8.836]	[9.995]	[0.025]	[23.965]	[25.363]	[0.031]
	Eigen	2.631	0.011	0.994	1.269	0.003	0.999	1.002	0.003	0.999
		[165.227]	[16.518]	[0.045]	[164.365]	[16.474]	[0.045]	[164.075]	[16.475]	[0.046]
	LU			1.000			1.000			1.000
				[1.699]			[1.760]			[1.722]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.168]			[3.265]			[3.136]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag(Υ) are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Computational times in seconds. Averages based on 1000 replications.

Table A.1.8
Simulation results for the approximated spatial lag operator inverse and related matrix functions, considering the new approximation method based on known matrices (AMBKM), fourth-order Taylor series approximation (Taylor4), fourth-order Chebyshev approximation (Cheb4), the Eigendecomposition (Eigen), the LU decomposition (LU) and the Conjugate Gradient method (CGrad), with $\alpha = 0.8$ and $\mathbf{W}$ based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$	$\mathbf{S}^{-1}$	diag(Υ)	$\mathbf{S}^{-1}\mathbf{X}$
100	True	[0.012]	[0.002]	≈[0.000]	[0.015]	[0.017]	[0.001]	[0.016]	[0.017]	[0.001]
	AMBKM	2.551	0.955	0.887	0.873	0.661	0.936	0.748	0.376	0.984
		[0.027]	[0.008]	≈[0.000]	[0.026]	[0.008]	≈[0.000]	[0.027]	[0.008]	≈[0.000]
	Taylor4	0.665	0.781	0.987	0.373	0.677	0.989	0.358	0.675	0.999
		[0.019]	[0.016]	≈[0.000]	[0.022]	[0.026]	≈[0.000]	[0.027]	[0.033]	≈[0.000]
	Cheb4	0.863	0.241	0.999	0.119	0.074	0.999	0.088	0.069	≈1.000
		[0.042]	[0.043]	≈[0.000]	[0.050]	[0.052]	≈[0.000]	[0.054]	[0.060]	≈[0.000]
	Eigen	3.641	0.136	0.970	0.865	0.004	0.995	0.822	0.004	0.992
		[0.012]	[0.002]	≈[0.000]	[0.043]	[0.004]	≈[0.000]	[0.043]	[0.004]	≈[0.000]
	LU			1.000			1.000			1.000
				[0.005]			[0.005]			[0.005]
	CGrad			0.603			≈1.000			≈1.000
				[0.016]			[0.010]			[0.010]
1000	True	[0.375]	[14.218]	[0.008]	[2.381]	[14.871]	[0.008]	[5.800]	[15.617]	[0.008]
	AMBKM	0.735	0.943	0.939	0.612	0.628	0.994	0.680	0.344	0.998
		[0.263]	[0.125]	[0.004]	[0.284]	[0.128]	[0.004]	[0.323]	[0.129]	[0.004]
	Taylor4	0.377	0.651	0.990	0.346	0.631	0.999	0.356	0.659	≈1.000
		[0.066]	[0.128]	[0.001]	[1.188]	[2.096]	[0.006]	[2.963]	[4.872]	[0.008]
	Cheb4	0.129	0.094	0.999	0.088	0.082	≈1.000	0.090	0.064	≈1.000
		[0.180]	[0.150]	[0.001]	[1.437]	[1.686]	[0.006]	[3.604]	[3.818]	[0.008]
	Eigen	1.840	0.003	0.995	0.982	0.001	0.999	1.056	0.001	0.999
		[22.331]	[2.102]	[0.012]	[22.302]	[2.102]	[0.012]	[22.282]	[2.161]	[0.012]
	LU			1.000			1.000			1.000
				[0.252]			[0.256]			[0.256]
	CGrad			≈1.000			≈1.000			≈1.000
				[0.498]			[0.475]			[0.498]
2000	True	[3.090]	[116.212]	[0.031]	[39.872]	[121.396]	[0.031]	[50.982]	[127.827]	[0.031]
	AMBKM	0.671	0.937	0.957	0.605	0.621	0.997	0.682	0.343	0.999
		[0.560]	[0.580]	[0.012]	[0.976]	[0.524]	[0.012]	[1.429]	[0.594]	[0.013]
	Taylor4	0.350	0.616	0.993	0.345	0.627	≈1.000	0.354	0.658	≈1.000
		[0.277]	[0.652]	[0.004]	[7.097]	[13.523]	[0.025]	[21.872]	[36.160]	[0.031]
	Cheb4	0.099	0.110	0.999	0.088	0.084	≈1.000	0.086	0.064	≈1.000
		[0.470]	[0.610]	[0.004]	[8.781]	[9.921]	[0.025]	[24.093]	[25.557]	[0.031]
	Eigen	2.170	0.002	0.998	1.304	≈0.000	0.999	1.128	≈0.000	0.999
		[165.791]	[16.481]	[0.046]	[164.499]	[16.476]	[0.046]	[164.210]	[16.478]	[0.046]
	LU			1.000			1.000			1.000
				[1.716]			[1.721]			[1.705]
	CGrad			≈1.000			≈1.000			≈1.000
				[3.299]			[3.244]			[3.261]

NOTE: The matrix $\mathbf{S} = \mathbf{I} - \alpha\mathbf{W}$. The values for the column $\mathbf{S}^{-1}$ are the average norm differences w.r.t. the identity matrix. The values for the column diag(Υ) are the average absolute deviations w.r.t. the true values. The values for the column $\mathbf{S}^{-1}\mathbf{X}$ are the average correlation coefficient between the approximated and the true operation. Numbers in brackets are average computational times. Averages are based on 1000 replications.

Appendix A.2. GMM estimation

Table A.2.1
Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0$ and **W** based on the radial distance criterion.

δ_R		1			2			4		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\widehat{\alpha}$	0.074	−0.033	−0.006	0.295	−0.498	−0.020	0.672	−0.894	−0.098
		(0.377)	(0.429)	(0.576)	(0.720)	(1.295)	(1.319)	(2.168)	(3.360)	(5.309)
	$\widehat{\beta}_0$	0.005	0.001	0.001	0.010	0.023	0.012	−0.015	0.072	−0.039
		(0.141)	(0.172)	(0.159)	(0.149)	(0.352)	(0.241)	(0.274)	(0.754)	(0.944)
	$\widehat{\beta}_1$	1.084	1.097	1.024	1.094	1.145	1.019	1.092	1.143	1.020
		(0.263)	(0.268)	(0.256)	(0.272)	(0.367)	(0.250)	(0.247)	(0.316)	(0.267)
	Time:									
	Loop	0.062	0.034		0.066	0.035		0.069	0.042	
	# Iterations	4	4		5	5		5	5	
	Total	0.305	0.201	0.068	0.371	0.233	0.064	0.418	0.288	0.076
	# Neighbors	6	6	6	21	21	21	60	60	60
1000	$\widehat{\alpha}$	0.005	−0.014	0.001	0.056	−0.057	−0.005	0.188	−0.336	−0.055
		(0.201)	(0.202)	(0.203)	(0.351)	(0.435)	(0.418)	(0.572)	(1.154)	(0.906)
	$\widehat{\beta}_0$	0.001	0.001	0.001	0.003	0.001	0.001	0.000	0.000	0.001
		(0.044)	(0.046)	(0.045)	(0.044)	(0.049)	(0.047)	(0.041)	(0.073)	(0.057)
	$\widehat{\beta}_1$	1.002	1.005	1.002	1.003	1.006	0.999	1.012	1.030	1.005
		(0.074)	(0.075)	(0.074)	(0.078)	(0.078)	(0.078)	(0.081)	(0.142)	(0.076)
	Time:									
	Loop	1.867	0.504		2.115	1.091		1.953	3.989	
	# Iterations	3	4		4	4		5	5	
	Total	6.819	2.517	0.824	9.517	5.697	0.881	10.572	21.991	1.447
	# Neighbors	9	9	9	33	33	33	118	118	118
2000	$\widehat{\alpha}$	0.000	−0.010	0.000	0.017	−0.044	−0.007	0.129	−0.129	−0.015
		(0.148)	(0.147)	(0.148)	(0.285)	(0.326)	(0.306)	(0.462)	(0.698)	(0.601)
	$\widehat{\beta}_0$	0.000	0.000	0.000	−0.002	−0.001	−0.001	0.000	−0.002	0.000
		(0.030)	(0.030)	(0.030)	(0.031)	(0.034)	(0.033)	(0.030)	(0.040)	(0.035)
	$\widehat{\beta}_1$	1.003	1.005	1.003	1.003	1.005	1.002	1.001	1.007	0.999
		(0.054)	(0.054)	(0.054)	(0.053)	(0.054)	(0.054)	(0.056)	(0.083)	(0.055)
	Time:									
	Loop	6.935	2.651		7.304	7.372		6.839	20.889	
	# Iterations	3	3		4	4		4	5	
	Total	23.064	11.109	2.464	30.193	33.943	2.489	34.917	107.516	4.050
	# Neighbors	9	9	9	37	37	37	134	134	134

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table A.2.2
Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.2$ and **W** based on the radial distance criterion.

δ_R		1			2			4		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\widehat{\alpha}$	0.189	0.132	0.228	0.279	−0.147	0.221	0.827	−0.099	0.143
		(0.388)	(0.389)	(0.589)	(0.655)	(1.184)	(1.370)	(1.866)	(3.875)	(5.414)
	$\widehat{\beta}_0$	0.006	0.009	0.004	−0.007	−0.004	−0.004	0.011	−0.064	0.020
		(0.132)	(0.142)	(0.148)	(0.141)	(0.246)	(0.224)	(0.254)	(0.689)	(0.778)
	$\widehat{\beta}_1$	1.083	1.109	1.022	1.105	1.104	1.017	1.135	1.144	1.020
		(0.275)	(0.278)	(0.254)	(0.272)	(0.349)	(0.259)	(0.301)	(0.343)	(0.263)
	Time:									
	Loop	0.063	0.033		0.065	0.036		0.070	0.041	
	# Iterations	4	5		5	5		5	6	
	Total	0.319	0.198	0.063	0.361	0.230	0.067	0.439	0.283	0.077
	# Neighbors	6	6	6	20	20	20	58	58	58
1000	$\widehat{\alpha}$	0.225	0.186	0.223	0.238	0.130	0.237	0.284	−0.028	0.218
		(0.205)	(0.165)	(0.202)	(0.380)	(0.355)	(0.428)	(0.518)	(1.070)	(0.889)
	$\widehat{\beta}_0$	−0.001	−0.001	−0.001	0.000	0.001	0.000	−0.003	−0.006	−0.004
		(0.033)	(0.035)	(0.033)	(0.036)	(0.039)	(0.037)	(0.036)	(0.072)	(0.053)
	$\widehat{\beta}_1$	1.001	1.007	0.999	1.011	1.010	1.004	1.011	1.033	1.001
		(0.077)	(0.077)	(0.077)	(0.084)	(0.080)	(0.080)	(0.083)	(0.139)	(0.078)
	Time:									
	Loop	1.489	0.441		1.596	0.933		1.670	3.549	
	# Iterations	3	4		4	5		5	5	
	Total	5.708	2.185	0.629	7.547	4.965	0.684	9.273	19.786	1.311
	# Neighbors	9	9	9	33	33	33	119	119	119
2000	$\widehat{\alpha}$	0.222	0.189	0.220	0.234	0.166	0.232	0.270	0.064	0.243
		(0.156)	(0.125)	(0.153)	(0.292)	(0.251)	(0.300)	(0.463)	(0.610)	(0.628)
	$\widehat{\beta}_0$	0.000	0.000	0.000	−0.001	−0.001	−0.001	0.000	−0.001	0.000
		(0.023)	(0.024)	(0.023)	(0.026)	(0.027)	(0.026)	(0.027)	(0.037)	(0.031)
	$\widehat{\beta}_1$	0.997	1.002	0.996	1.001	1.003	1.001	1.001	1.021	0.998
		(0.054)	(0.054)	(0.054)	(0.056)	(0.056)	(0.055)	(0.053)	(0.156)	(0.054)
	Time:									
	Loop	6.381	2.596		5.949	6.381		6.356	21.036	
	# Iterations	3	4		4	4		5	5	
	Total	22.786	11.184	2.015	25.671	29.407	2.021	33.666	108.958	3.964
	# Neighbors	9	9	9	37	37	37	133	133	133

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table A.2.3
Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.5$ and **W** based on the radial distance criterion.

δ_R		1			2			4		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\widehat{\alpha}$	0.409	0.328	0.684	0.496	−0.110	0.627	0.837	−0.620	0.739
		(0.451)	(0.389)	(0.686)	(0.733)	(1.330)	(1.528)	(1.773)	(4.238)	(4.292)
	$\widehat{\beta}_0$	−0.002	0.002	0.001	0.007	−0.010	0.005	−0.030	−0.031	−0.049
		(0.110)	(0.118)	(0.119)	(0.160)	(0.509)	(0.282)	(0.300)	(1.054)	(0.738)
	$\widehat{\beta}_1$	1.051	1.082	0.947	1.121	1.109	1.003	1.066	1.139	1.007
		(0.265)	(0.268)	(0.250)	(0.318)	(0.326)	(0.259)	(0.268)	(0.371)	(0.264)
	Time:									
	Loop	0.063	0.033		0.066	0.035		0.069	0.042	
	# Iterations	5	5		5	5		5	6	
	Total	0.335	0.203	0.064	0.379	0.223	0.067	0.436	0.300	0.078
	# Neighbors	6	6	6	20	20	20	59	59	59
1000	$\widehat{\alpha}$	0.669	0.477	0.678	0.615	0.376	0.718	0.474	0.274	0.696
		(0.267)	(0.106)	(0.279)	(0.414)	(0.247)	(0.483)	(0.554)	(1.010)	(0.944)
	$\widehat{\beta}_0$	0.000	0.000	0.000	0.002	0.002	0.002	0.001	0.001	−0.001
		(0.019)	(0.025)	(0.019)	(0.027)	(0.029)	(0.027)	(0.041)	(0.078)	(0.054)
	$\widehat{\beta}_1$	0.985	1.004	0.961	1.009	1.002	0.990	1.020	1.075	1.003
		(0.093)	(0.076)	(0.084)	(0.105)	(0.076)	(0.076)	(0.105)	(0.296)	(0.078)
	Time:									
	Loop	1.557	0.457		1.635	0.959		1.645	3.483	
	# Iterations	5	5		5	5		5	5	
	Total	8.076	2.621	0.561	8.765	5.337	0.674	9.473	19.264	1.302
	# Neighbors	9	9	9	34	34	34	120	120	120
2000	$\widehat{\alpha}$	0.705	0.492	0.688	0.698	0.443	0.738	0.553	0.422	0.721
		(0.256)	(0.070)	(0.245)	(0.357)	(0.141)	(0.386)	(0.494)	(0.544)	(0.670)
	$\widehat{\beta}_0$	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	−0.001
		(0.012)	(0.016)	(0.012)	(0.015)	(0.018)	(0.015)	(0.024)	(0.030)	(0.026)
	$\widehat{\beta}_1$	0.979	1.004	0.966	1.008	1.005	0.993	1.007	1.050	0.997
		(0.065)	(0.056)	(0.064)	(0.062)	(0.052)	(0.053)	(0.058)	(0.209)	(0.056)
	Time:									
	Loop	8.178	3.004		7.848	7.641		6.861	21.501	
	# Iterations	5	5		5	5		5	5	
	Total	42.737	16.435	2.237	42.574	41.329	2.379	39.126	116.620	4.037
	# Neighbors	10	10	10	36	36	36	136	136	136

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table A.2.4
Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.8$ and **W** based on the radial distance criterion.

δ_R		1			2			4		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\widehat{\alpha}$	0.797	0.575	1.582	0.852	0.397	1.732	1.392	0.004	1.160
		(0.589)	(0.509)	(1.418)	(0.985)	(1.460)	(2.137)	(2.318)	(4.172)	(5.572)
	$\widehat{\beta}_0$	−0.002	−0.010	−0.007	0.016	0.024	0.010	−0.032	−0.124	0.113
		(0.123)	(0.140)	(0.305)	(0.285)	(0.484)	(0.604)	(0.696)	(1.847)	(2.455)
	$\widehat{\beta}_1$	0.933	0.995	0.748	0.969	1.000	0.873	1.013	1.040	0.912
		(0.274)	(0.338)	(0.349)	(0.277)	(0.272)	(0.284)	(0.260)	(0.284)	(0.275)
	Time:									
	Loop	0.066	0.032		0.067	0.036		0.071	0.042	
	# Iterations	5	5		5	5		6	6	
	Total	0.389	0.211	0.062	0.410	0.242	0.068	0.467	0.291	0.076
	# Neighbors	6	6	6	21	21	21	60	60	60
1000	$\widehat{\alpha}$	1.403	0.690	1.584	1.418	1.009	1.853	1.173	1.330	1.824
		(0.641)	(0.117)	(0.836)	(0.796)	(0.711)	(1.157)	(0.991)	(1.257)	(1.398)
	$\widehat{\beta}_0$	0.002	−0.004	0.001	−0.007	−0.001	−0.003	0.010	−0.018	−0.005
		(0.044)	(0.014)	(0.063)	(0.060)	(0.096)	(0.100)	(0.113)	(0.163)	(0.131)
	$\widehat{\beta}_1$	0.880	0.956	0.799	0.941	1.114	0.916	0.988	1.110	0.973
		(0.171)	(0.080)	(0.215)	(0.109)	(0.350)	(0.114)	(0.076)	(0.258)	(0.081)
	Time:									
	Loop	1.673	0.449		1.581	1.046		1.653	3.683	
	# Iterations	6	6		6	6		5	6	
	Total	11.297	3.034	0.521	10.185	6.592	0.569	10.449	22.218	1.276
	# Neighbors	8	8	8	39	39	39	125	125	125
2000	$\widehat{\alpha}$	1.255	0.650	1.616	1.522	0.825	1.880	1.795	1.491	1.876
		(0.455)	(0.150)	(0.844)	(0.798)	(0.363)	(1.141)	(1.159)	(1.087)	(1.283)
	$\widehat{\beta}_0$	−0.007	−0.019	0.002	−0.003	0.009	−0.001	0.001	0.007	0.001
		(0.007)	(0.019)	(0.044)	(0.060)	(0.040)	(0.071)	(0.079)	(0.115)	(0.086)
	$\widehat{\beta}_1$	0.890	0.986	0.813	0.953	1.129	0.926	0.983	1.264	0.975
		(0.110)	(0.014)	(0.196)	(0.091)	(0.272)	(0.091)	(0.060)	(0.420)	(0.060)
	Time:									
	Loop	6.707	2.594		5.965	7.410		5.962	22.797	
	# Iterations	7	5		6	6		5	6	
	Total	48.554	14.623	1.728	37.195	47.109	1.916	35.994	135.114	3.671
	# Neighbors	8	8	8	44	44	44	143	143	143

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table A.2.5
Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0$ and **W** based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\widehat{\alpha}$	−0.005	−0.001	0.004	0.198	−0.215	0.017	0.249	−0.499	−0.024
		(0.225)	(0.215)	(0.250)	(0.570)	(0.847)	(0.953)	(0.776)	(1.401)	(1.445)
	$\widehat{\beta}_0$	0.003	0.005	0.002	−0.004	0.019	−0.001	0.004	0.004	0.003
		(0.142)	(0.152)	(0.141)	(0.132)	(0.246)	(0.183)	(0.146)	(0.377)	(0.266)
	$\widehat{\beta}_1$	1.050	1.083	1.024	1.106	1.097	1.018	1.133	1.140	1.033
		(0.256)	(0.267)	(0.257)	(0.295)	(0.279)	(0.257)	(0.289)	(0.305)	(0.260)
	Time:									
	Loop	0.061	0.032		0.063	0.035		0.066	0.037	
	# Iterations	4	4		5	5		5	5	
	Total	0.288	0.181	0.061	0.353	0.224	0.066	0.376	0.242	0.064
	%Asymmetry	0.274	0.274	0.274	0.151	0.151	0.151	0.146	0.146	0.146
1000	$\widehat{\alpha}$	0.012	−0.015	0.004	0.184	−0.260	0.028	0.264	−0.654	−0.025
		(0.227)	(0.229)	(0.230)	(0.561)	(1.102)	(0.812)	(0.645)	(1.959)	(1.345)
	$\widehat{\beta}_0$	0.001	0.002	0.002	0.000	0.002	0.003	0.001	0.005	0.001
		(0.044)	(0.045)	(0.044)	(0.040)	(0.076)	(0.053)	(0.041)	(0.127)	(0.074)
	$\widehat{\beta}_1$	1.003	1.006	1.002	1.016	1.028	1.003	1.015	1.031	1.006
		(0.077)	(0.077)	(0.078)	(0.082)	(0.174)	(0.077)	(0.087)	(0.138)	(0.078)
	Time:									
	Loop	1.382	0.495		1.568	2.324		1.575	5.631	
	# Iterations	3	4		5	5		5	5	
	Total	5.420	2.430	0.549	8.526	12.931	0.969	9.796	31.838	1.970
	% Asymmetry	0.130	0.130	0.130	0.106	0.106	0.106	0.125	0.125	0.125
2000	$\widehat{\alpha}$	0.006	−0.021	0.000	0.176	−0.309	−0.030	0.303	−0.426	0.032
		(0.232)	(0.242)	(0.232)	(0.583)	(1.104)	(0.838)	(0.708)	(1.903)	(1.258)
	$\widehat{\beta}_0$	−0.001	−0.001	−0.001	−0.001	0.003	0.000	0.000	−0.001	0.002
		(0.029)	(0.031)	(0.030)	(0.030)	(0.052)	(0.040)	(0.028)	(0.070)	(0.049)
	$\widehat{\beta}_1$	1.000	1.001	0.999	1.005	1.019	1.000	1.007	1.020	1.001
		(0.053)	(0.053)	(0.053)	(0.055)	(0.119)	(0.054)	(0.056)	(0.113)	(0.055)
	Time:									
	Loop	6.846	4.995		6.435	38.929		6.486	49.120	
	# Iterations	3	4		5	5		5	5	
	Total	26.624	21.849	2.248	36.496	205.426	5.626	47.796	279.975	14.784
	% Asymmetry	0.102	0.102	0.102	0.101	0.101	0.101	0.123	0.123	0.123

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table A.2.6
Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.2$ and **W** based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\widehat{\alpha}$	0.164	0.163	0.214	0.325	−0.121	0.250	0.293	−0.380	0.203
		(0.214)	(0.218)	(0.268)	(0.556)	(0.828)	(0.900)	(0.702)	(1.394)	(1.440)
	$\widehat{\beta}_0$	0.002	0.001	0.001	−0.006	−0.002	0.006	−0.009	0.002	0.014
		(0.119)	(0.131)	(0.115)	(0.137)	(0.273)	(0.184)	(0.142)	(0.305)	(0.231)
	$\widehat{\beta}_1$	1.026	1.071	0.990	1.100	1.103	1.010	1.098	1.148	1.013
		(0.251)	(0.271)	(0.254)	(0.291)	(0.270)	(0.258)	(0.274)	(0.392)	(0.258)
	Time:									
	Loop	0.062	0.033		0.064	0.035		0.064	0.037	
	# Iterations	4	4		5	5		5	5	
	Total	0.303	0.189	0.061	0.346	0.221	0.064	0.365	0.242	0.067
	%Asymmetry	0.273	0.273	0.273	0.152	0.152	0.152	0.147	0.147	0.147
1000	$\widehat{\alpha}$	0.237	0.185	0.236	0.257	−0.087	0.226	0.366	−0.344	0.196
		(0.241)	(0.188)	(0.239)	(0.518)	(0.933)	(0.814)	(0.614)	(1.827)	(1.333)
	$\widehat{\beta}_0$	0.000	0.000	0.000	0.003	0.000	0.001	0.000	0.004	−0.002
		(0.033)	(0.035)	(0.033)	(0.038)	(0.065)	(0.050)	(0.040)	(0.105)	(0.074)
	$\widehat{\beta}_1$	1.002	1.006	0.999	1.010	1.036	1.001	1.016	1.033	1.006
		(0.082)	(0.080)	(0.080)	(0.076)	(0.164)	(0.075)	(0.085)	(0.135)	(0.079)
	Time:									
	Loop	1.300	0.488		1.474	2.163		1.546	5.685	
	# Iterations	4	4		5	5		5	5	
	Total	5.288	2.398	0.502	7.908	11.903	0.943	9.723	32.795	1.964
	% Asymmetry	0.130	0.130	0.130	0.106	0.106	0.106	0.125	0.125	0.125
2000	$\widehat{\alpha}$	0.227	0.174	0.224	0.300	−0.081	0.221	0.386	−0.340	0.190
		(0.238)	(0.194)	(0.237)	(0.538)	(0.992)	(0.831)	(0.647)	(1.730)	(1.249)
	$\widehat{\beta}_0$	−0.001	−0.001	−0.001	0.000	0.000	0.001	−0.001	0.001	−0.002
		(0.023)	(0.024)	(0.023)	(0.026)	(0.045)	(0.035)	(0.029)	(0.072)	(0.047)
	$\widehat{\beta}_1$	1.000	1.003	0.999	1.009	1.029	1.003	1.007	1.022	1.000
		(0.053)	(0.053)	(0.053)	(0.067)	(0.125)	(0.052)	(0.061)	(0.098)	(0.056)
	Time:									
	Loop	5.646	4.368		6.266	38.941		6.251	48.360	
	# Iterations	4	4		5	5		5	5	
	Total	22.771	19.424	1.836	36.008	203.903	5.603	46.456	271.254	14.558
	% Asymmetry	0.103	0.103	0.103	0.101	0.101	0.101	0.123	0.123	0.123

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table A.2.7
Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.5$ and **W** based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\widehat{\alpha}$	0.375	0.393	0.581	0.528	0.132	0.795	0.537	−0.091	0.739
		(0.260)	(0.238)	(0.397)	(0.595)	(0.852)	(1.233)	(0.660)	(1.539)	(1.476)
	$\widehat{\beta}_0$	0.004	0.005	0.000	−0.009	0.032	0.007	−0.002	−0.015	−0.005
		(0.090)	(0.106)	(0.083)	(0.124)	(0.264)	(0.282)	(0.124)	(0.456)	(0.295)
	$\widehat{\beta}_1$	1.006	1.092	0.870	1.099	1.114	0.990	1.109	1.162	1.002
		(0.241)	(0.301)	(0.285)	(0.305)	(0.324)	(0.253)	(0.314)	(0.407)	(0.265)
	Time:									
	Loop	0.064	0.033		0.065	0.035		0.066	0.037	
	# Iterations	5	5		5	5		5	5	
	Total	0.354	0.206	0.062	0.375	0.227	0.063	0.385	0.246	0.066
	%Asymmetry	0.274	0.274	0.274	0.151	0.151	0.151	0.147	0.147	0.147
1000	$\widehat{\alpha}$	0.683	0.459	0.706	0.509	0.316	0.721	0.463	0.073	0.744
		(0.294)	(0.111)	(0.315)	(0.569)	(0.855)	(0.853)	(0.672)	(1.682)	(1.274)
	$\widehat{\beta}_0$	0.000	0.000	0.000	0.000	0.004	0.001	−0.002	−0.004	−0.003
		(0.020)	(0.025)	(0.020)	(0.037)	(0.066)	(0.048)	(0.041)	(0.114)	(0.072)
	$\widehat{\beta}_1$	0.999	1.006	0.972	1.010	1.060	1.000	1.018	1.053	0.999
		(0.093)	(0.075)	(0.080)	(0.078)	(0.199)	(0.075)	(0.099)	(0.211)	(0.082)
	Time:									
	Loop	1.469	0.521		1.499	2.183		1.562	5.623	
	# Iterations	5	5		5	5		5	5	
	Total	7.873	2.937	0.517	8.560	12.478	0.938	10.004	32.083	1.964
	% Asymmetry	0.131	0.131	0.131	0.106	0.106	0.106	0.126	0.126	0.126
2000	$\widehat{\alpha}$	0.702	0.462	0.721	0.547	0.390	0.731	0.485	0.293	0.698
		(0.301)	(0.104)	(0.321)	(0.557)	(0.814)	(0.832)	(0.692)	(1.647)	(1.338)
	$\widehat{\beta}_0$	−0.001	−0.001	−0.001	0.002	0.002	0.002	0.001	0.005	0.003
		(0.014)	(0.018)	(0.014)	(0.025)	(0.041)	(0.032)	(0.030)	(0.065)	(0.056)
	$\widehat{\beta}_1$	0.998	1.003	0.984	1.009	1.058	1.000	1.011	1.033	0.999
		(0.057)	(0.056)	(0.058)	(0.055)	(0.190)	(0.053)	(0.067)	(0.176)	(0.052)
	Time:									
	Loop	7.442	5.147		6.422	38.466		6.764	48.674	
	# Iterations	5	5		5	5		5	5	
	Total	40.022	27.738	1.980	38.652	208.172	5.454	48.511	274.855	14.525
	% Asymmetry	0.102	0.102	0.102	0.101	0.101	0.101	0.123	0.123	0.123

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Table A.2.8
Simulation results for the Spatial Probit model considering the iterative GMM estimator with approximated gradients (iGMMa), the iterative GMM estimator (iGMM) and the GMM estimator for the linearized model (LGMM), with $\alpha = 0.8$ and **W** based on the nearest neighbors criterion.

δ_{NN}		0.01			0.1			0.2		
N		iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM	iGMMa	iGMM	LGMM
100	$\widehat{\alpha}$	0.377	0.637	1.609	0.839	0.411	1.928	0.767	0.166	1.897
		(0.472)	(0.249)	(15.527)	(0.699)	(0.698)	(1.920)	(0.885)	(1.562)	(2.322)
	$\widehat{\beta}_0$	0.012	0.003	0.171	0.004	−0.004	−0.008	0.023	0.070	0.031
		(0.079)	(0.092)	(5.020)	(0.174)	(0.197)	(0.442)	(0.238)	(0.633)	(0.730)
	$\widehat{\beta}_1$	0.805	0.984	0.577	0.954	0.998	0.798	1.002	1.048	0.860
		(0.296)	(0.359)	(0.478)	(0.255)	(0.255)	(0.310)	(0.279)	(0.341)	(0.284)
	Time:									
	Loop	0.064	0.033		0.065	0.034		0.065	0.037	
	# Iterations	5	6		6	5		6	5	
	Total	0.356	0.234	0.065	0.409	0.221	0.063	0.422	0.252	0.066
	%Asymmetry	0.272	0.272	0.272	0.152	0.152	0.152	0.145	0.145	0.145
1000	$\widehat{\alpha}$	1.523	0.627	1.773	1.049	1.329	1.942	0.877	1.190	1.720
		(0.792)	(0.176)	(1.028)	(0.824)	(1.283)	(1.461)	(0.854)	(1.752)	(1.618)
	$\widehat{\beta}_0$	0.016	0.000	−0.002	−0.005	0.005	0.003	−0.001	0.008	−0.001
		(0.046)	(0.019)	(0.078)	(0.053)	(0.176)	(0.139)	(0.066)	(0.192)	(0.172)
	$\widehat{\beta}_1$	0.881	0.979	0.817	0.997	1.150	0.968	0.994	1.074	0.980
		(0.139)	(0.090)	(0.197)	(0.107)	(0.307)	(0.085)	(0.086)	(0.270)	(0.080)
	Time:									
	Loop	1.464	0.541		1.456	2.188		1.512	5.654	
	# Iterations	6	6		6	6		5	6	
	Total	9.445	3.432	0.421	8.984	13.301	0.860	10.190	33.624	1.894
	% Asymmetry	0.128	0.128	0.128	0.105	0.105	0.105	0.125	0.125	0.125
2000	$\widehat{\alpha}$	1.786	0.621	1.850	1.759	1.477	1.900	1.462	1.130	1.701
		(1.034)	(0.179)	(1.089)	(1.236)	(1.190)	(1.391)	(1.227)	(1.747)	(1.590)
	$\widehat{\beta}_0$	−0.001	0.010	−0.001	−0.002	−0.001	−0.005	0.002	−0.018	−0.007
		(0.031)	(0.010)	(0.065)	(0.065)	(0.133)	(0.096)	(0.065)	(0.139)	(0.119)
	$\widehat{\beta}_1$	0.859	0.975	0.886	0.994	1.176	0.987	0.998	1.056	0.990
		(0.150)	(0.025)	(0.125)	(0.061)	(0.348)	(0.057)	(0.059)	(0.190)	(0.054)
	Time:									
	Loop	6.170	4.801		5.842	39.105		5.834	47.706	
	# Iterations	6	6		5	6		5	6	
	Total	38.760	30.506	1.885	36.167	227.035	5.528	43.368	282.436	14.560
	% Asymmetry	0.100	0.100	0.100	0.101	0.101	0.101	0.123	0.123	0.123

NOTE: Simulations based on 1000 replications. Numbers are mean values and numbers in parentheses are root mean square errors (RMSEs). Computational times in seconds. True values of the regressions parameters fixed at $\beta_0 = 0$ and $\beta_1 = 1$.

Appendix B. Empirical application

Table B.1
Descriptive statistics for the variables included in the empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas.

	Mean	Std. Dev.	Min	Q_1	Median	Q_3	Max	N
BCI	0.147	0.354	0.000	0.000	0.000	0.000	1.000	4848
AQI_{min}	0.131	0.086	0.000	0.060	0.120	0.190	0.430	4848
AQI_{max}	1.514	0.931	0.380	1.120	1.430	1.710	22.120	4848
% days O_3	0.480	0.276	0.000	0.312	0.468	0.682	1.000	4848
% days $PM_{2.5}$	0.406	0.273	0.000	0.192	0.386	0.584	1.000	4848
% days PM_{10}	0.030	0.101	0.000	0.000	0.000	0.011	1.000	4848
% days CO	0.008	0.048	0.000	0.000	0.000	0.000	0.738	4848
% days SO_2	0.052	0.126	0.000	0.000	0.000	0.019	0.962	4848
% days NO_2	0.024	0.056	0.000	0.000	0.000	0.019	0.499	4848
% days Above Moderate	0.370	0.213	0.000	0.202	0.344	0.504	0.966	4848
% days Exceptional Events	0.024	0.092	0.000	0.000	0.000	0.000	0.940	4848

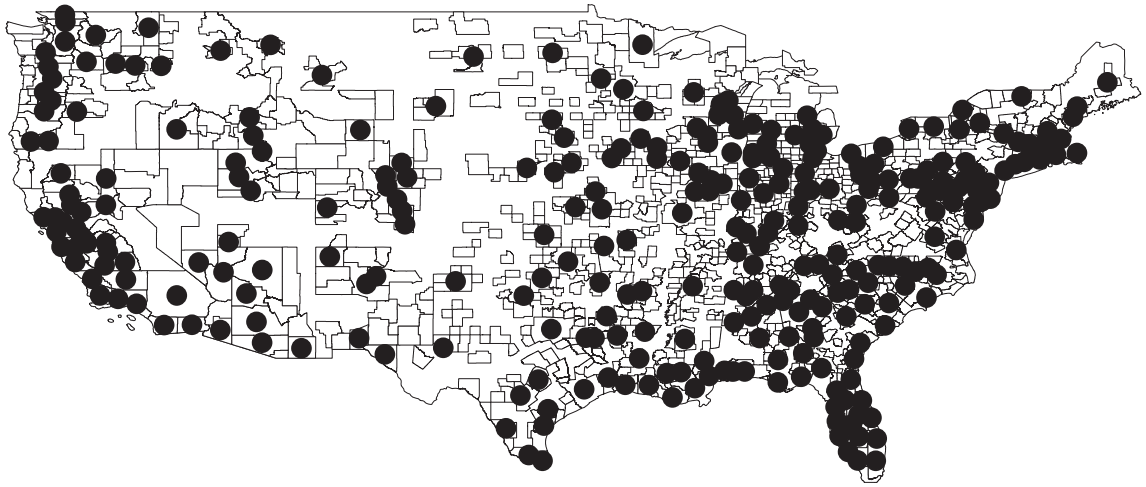

Fig. B.1 Centroids of the U.S. Metropolitan Statistical Areas included in the empirical application.

Table B.2
Spatial lag Probit estimation results for the empirical application on the competitiveness in the U.S. Metropolitan Statistical Areas.

	Dependent variable: BCI					
	UNRESTRICTED MODEL			RESTRICTED MODEL		
	(iGMMa)	(iGMM)	(LGMM)	(iGMMa)	(iGMM)	(LGMM)
Intercept	0.046	0.178	−0.423	−0.059	0.194	−0.328
	(0.806)	(0.932)	(1.136)	(0.796)	(0.900)	(1.144)
AQI_{min}	−2.370**	−2.589**	−3.022***	−2.154**	−2.339**	−2.749***
	(1.039)	(1.061)	(0.990)	(1.025)	(1.047)	(0.989)
AQI^2_{min}	6.630**	8.078**	8.850***	6.213*	7.423**	8.419**
	(3.354)	(3.319)	(3.261)	(3.336)	(3.299)	(3.280)
AQI_{max}	−0.022	−0.033	0.024			
	(0.085)	(0.090)	(0.094)			
AQI^2_{max}	0.004	0.004	−0.003			
	(0.005)	(0.006)	(0.008)			
% days O_3	0.776	0.780	0.674	0.703***	0.640**	0.643***
	(0.589)	(0.622)	(0.553)	(0.247)	(0.250)	(0.236)
% days $PM_{2.5}$	0.396	0.366	0.349	0.445*	0.325	0.435**
	(0.586)	(0.611)	(0.573)	(0.232)	(0.231)	(0.215)
% days PM_{10}	0.965	0.967	0.884			
	(0.717)	(0.750)	(0.632)			
% days SO_2	−0.717	−0.524	−0.917			
	(0.676)	(0.698)	(0.662)			
% days NO_2	0.400	0.395	0.626			
	(0.942)	(0.964)	(0.897)			
% days Above Moderate	0.334	0.231	0.421			
	(0.352)	(0.375)	(0.327)			
% days Exceptional Events	0.116	0.241	0.056			
	(0.359)	(0.371)	(0.373)			
Spatial Lag (α)	0.771***	0.686***	0.954**	0.768***	0.711***	1.051***
	(0.051)	(0.056)	(0.375)	(0.052)	(0.055)	(0.388)
Observations	4848	4848	4848	4848	4848	4848
# Neighbors (average)	16	16	16	16	16	16
# Iterations	11	14	–	15	18	–
Total Time (in seconds)	33.999	123.689	7.102	34.928	154.002	6.891
# Instruments	141	141	141	113	113	113
Wald test (overall sig.)	24.155	23.004	30.584	11.123	12.490	15.435
(p-value)	(0.012)	(0.018)	(0.001)	(0.025)	(0.014)	(0.004)
Wald test (excl. restr.)[1]	–	–	–	12.783	9.999	15.781
(p-value)	(−)	(−)	(−)	(0.078)	(0.189)	(0.027)
Hansen's J test	85.178	49.565	1124.975	77.899	44.497	1530.180
(p-value)	(0.911)	≈(1.000)	≈(0.000)	(0.638)	≈(1.000)	≈(0.000)
McFadden R^2	0.038	0.061	−0.428	0.037	0.058	−0.662
$\rho^2(\hat{Y}, Y)$	0.056	0.061	0.033	0.055	0.059	0.003
$\%(\hat{Y} = Y)$	0.861	0.856	0.679	0.859	0.856	0.559

NOTE: Robust standard errors in parentheses, based on Kelejian and Prucha (2007). Time effects and Mundlak variables were added. Significance at the 1%, 5% and 10% levels indicated by ***, ** and *, respectively. Wald test for exclusion restrictions. Under the null hypothesis, the coefficients for the variables AQI_{max}, AQI^2_{max}, % days PM_{10}, % days SO_2, % days NO_2, % days Above Moderate and % days Exceptional Events are jointly equal to zero.

References

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

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

Baltagi, B.H., LeSage, J.P., Pace, R.K., 2016. Spatial econometrics: qualitative and limited dependent variables. Emerald Group Publishing.

Beron, K.J., Murdoch, J.C., Vijverberg, W.P.M., 2003. Why Cooperate? Public Goods, Economic Power, and the Montreal Protocol. Rev. Econ. Stat. 85 (2), 286–297. 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. 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. Transp. Res. Part B Methodol. 45 (7), 923–939. https://dx.doi.org/10.1016/j.trb.2011.04.005.

Billé, A.G., 2013. Computational issues in the estimation of the spatial probit model: A comparison of various estimators. Rev. Reg. Stud. 43 (2, 3), 131–154. https://journal.srsa.org/ojs/index.php/RRS/article/view/43.23.3.

Calabrese, R., Elkink, J.A., 2014. Estimators of binary spatial autoregressive models: A Monte Carlo study. J. Reg. Sci. 54 (4), 664–687. https://dx.doi.org/10.1111/jors.12116.

Case, A., 1992. Neighborhood influence and technological change. Reg. Sci. Urban Econ. 22 (3), 491–508. 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. Rev. Reg. Stud. 40 (2), 197–226. 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. Econ. J. 98 (391), 355–374. http://www.jstor.org/stable/2233372.

Fagerberg, J., 1996. Technology and competitiveness. Oxf. Rev. Econ. Pol. 12 (3), 39–51. https://doi.org/10.1093/oxrep/12.3.39.

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

Gourieroux, C., Monfort, A., Renault, E., Trognon, A., 1987. Generalised residuals. J. Econom. 34 (1), 5–32. 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. Lin. Algebra Appl. 321 (1), 95–112. 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, 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. J. Agric. Econ. 58 (1), 37–60. 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. Agric. Econ. 27 (3), 383–402. 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., 2007. HAC estimation in a spatial framework. J. Econom. 140 (1), 131–154. https://dx.doi.org/10.1016/j.jeconom.2006.09.005.

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. 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. J. Bus. Econ. Stat. 26 (4), 460–471. https://dx.doi.org/10.1198/073500107000000188.

Lahiri, S.N., 1996. On Inconsistency of Estimators Based on Spatial Data under Infill Asymptotics. Sankhya: Ind. J. Statist. Ser. A. 58 (3), 403–417. 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. 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. Geogr. Anal. 32 (1), 19–35, https://doi.org/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. J. Roy. Stat. Soc. 174 (4), 1007–1027. 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. Reg. Sci. Urban Econ. 64, 30–45. https://dx.doi.org/10.1016/j.regsciurbeco.2017.02.002.

McMillen, D.P., 1992. Probit with spatial autocorrelation. J. Reg. Sci. 32 (3), 335–348. https://dx.doi.org/10.1111/j.1467-9787.1992.tb00190.x.

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? Rev. Econ. Stat. 85 (4), 1038–1047. 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. Transport. Res. Rec.: J. Transport. Res. Board. 1898, 183–190. https://dx.doi.org/10.3141/1898-22.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica. 69–85. https://dx.doi.org/10.2307/1913646.

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. J. Publ. Econ. 87 (2), 337–362. https://dx.doi.org/10.1016/S0047-2727(01)00152-9.

Ord, K., 1975. Estimation methods for models of spatial interaction. J. Am. Stat. Assoc. 70 (349), 120–126. https://dx.doi.org/10.2307/2285387.

Pace, R.K., Barry, R., 1997a. Quick computation of spatial autoregressive estimators. Geogr. Anal. 29 (3), 232–247. https://dx.doi.org/10.1111/j.1538-4632.1997.tb00959.x.

Pace, R.K., Barry, R., 1997b. Sparse spatial autoregressions. Stat. Probab. Lett. 33 (3), 291–297. https://dx.doi.org/10.1016/S0167-7152(96)00140-X.

Pace, R.K., LeSage, J.P., 2004. Chebyshev approximation of log-determinants of spatial weight matrices. Comput. Stat. Data Anal. 45 (2), 179–196. https://dx.doi.org/10.1016/S0167-9473(02)00321-3.

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. 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. J. Econom. 85 (1), 125–154. https://dx.doi.org/10.1016/S0304-4076(97)00097-3.

Porter, M.E., 1990. The Competitive Advantage of Nations. Harv. Bus. Rev. 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. Reg. Sci. Urban Econ. 36 (6), 727–752. 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. Pap. Reg. Sci. 83 (2), 407–424. 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, http://documents.worldbank.org/curated/en/833431468739515725/Economic-growth-and-environmental-quality-time-series-and-cross-country-evidence.

Smirnov, O.A., 2005. Computation of the information matrix for models with spatial interaction on a lattice. J. Comput. Graph Stat. 14 (4), 910–927. https://dx.doi.org/10.1198/106186005X78585.

Smirnov, O.A., 2010. Modeling spatial discrete choice. Reg. Sci. Urban Econ. 40 (5), 292–298. https://dx.doi.org/10.1016/j.regsciurbeco.2009.09.004.

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. 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. J. Econom. 172 (1), 77–89. 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. Ecol. Econ. 97, 120–128. https://dx.doi.org/10.1016/j.ecolecon.2013.11.010.