The inversion of the spatial lag operator in nonlinear models: fast computation and a closed formula approximation*

Luís Silveira Santos†
ISEG–ULisboa and CEMAPRE

Isabel Proença‡
ISEG–ULisboa and CEMAPRE

May 15, 2017

Abstract

This paper implements a new method to approximate the inverse of the spatial lag operator matrix, applied to the estimation of nonlinear models with a spatially lagged dependent variable using a Generalized Method of Moments approach. This procedure drastically reduces the computational complexity of the inversion of the operator matrix, proving to be especially valuable for large spatial weights matrices. It is based on a first order truncation and high order approximation of the Taylor series expansion. It explores particular features of spatial weights matrices and, simultaneously, considers the eigensystem structure from higher order powers of normalized matrices. As result, a wide range of matrix operations can be approximated without additional computational burden. A closed formula approximation of the true elements of the inverse of the operator matrix is obtained. The procedure is implemented to the estimation of a nonlinear spatial GMM, preserving the nonlinearity of the model and significantly decreasing the amount of time needed for convergence. A Monte Carlo simulation study shows that the nonlinear spatial GMM with approximated inverse operator matrix performs well in terms of bias and mean square error. In addition, the estimator typically exhibits a minimum trade-off between time and unbiasedness within the class of the nonlinear spatial estimators. Finally, the usefulness of the estimator is illustrated with an empirical application. The authors study the spatial spillovers of competitiveness in the Metropolitan Statistical Areas of the United States of America, to assess the effects over the probability of being a major pole of economic development. These effects are also controlled for environmental quality variables.

Keywords: Spatial Econometrics, Nonlinear Models, Inverse Spatial Operator

JEL Classification: C18, C31, C35

*The authors gratefully acknowledge the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia) for the financial support through national funds: Luís Silveira Santos for the Ph.D. scholarship SFRH/BD/92277/2013 and Isabel Proença for CEMAPRE-UID/Multi/00491/2013 program. Also, the author Luís Silveira Santos would like to thank António Jorge Marques for the computer logistics, crucial for the implementation of the simulation study.

†Corresponding author. *Address:* Instituto Superior de Economia e Gestão, Rua do Quelhas n.º 6, 1200-781 Lisboa, Portugal. *E-mail:* lsantos@iseg.ulisboa.pt

‡*Address:* Instituto Superior de Economia e Gestão, Rua do Quelhas n.º 6, 1200-781 Lisboa, Portugal. *E-mail:* isabelp@iseg.ulisboa.pt

1 Introduction

Observability of the dependent variable is crucial to define the model to be estimated. Linearity and nonlinearity in parametric models is related to the ability of the researcher to fully observe the behavior of a particular dependent variable or at least some characteristics of this variable. For cross-sectional data with independent observations, the estimation of linear models and nonlinear models poses no problem, as the classic estimators and their statistical properties are, by far, extensively discussed in the literature (see the books of Amemiya, 1985; Maddala, 1986; Greene, 2003; Wooldridge, 2010).

For cross-sectional data with dependent observations, several complications arise. This is the case for spatial data. As the observational units are correlated in space, the structure of dependence should be reflected in the estimation (Paelinck, 1967). For linear models, the classic estimators need to be adjusted to account for the spatial autocorrelation and spatial heteroskedasticity induced by the spatial weights matrix (see the books of Cliff and Ord, 1973; Anselin, 1988; Cressie, 1993). Estimation relies on the eigenstructure of such matrix. For large sample sizes, estimation can be a prodigious task, due to the computational complexity associated with $N \times N$ matrix operations. For nonlinear models, this problem is amplified. The estimation procedures require the maximization of a nonlinear function, that also depend on a $N \times N$ matrix. Several approaches consider likelihood based methods (McMillen, 1992; Beron and Vijverberg, 2004, to mention a few) or Bayesian methods (LeSage, 2000; Smith and LeSage, 2004). But are also infeasible for large samples, because they rely on N dimensional integration or simulations of N-variate distributions, respectively. Distribution-free approaches, such as the Generalized Method of Moments, are more appealing on this framework (Pinkse and Slade, 1998; Fleming, 2004; Klier and McMillen, 2012). However, these methods rely on $N \times N$ matrix inversions.

Taking into account the complexity of estimating nonlinear models in spatial frameworks, the objectives of this paper can be summarized in twofold. One, to propose an approximation that drastically reduces the computational complexity and computational time of the $N \times N$ matrix inversion. Second, to implement this approximation to the estimation of a general nonlinear model with a spatially lagged binary dependent variable.

To reduce the computational complexity and computational time of the $N \times N$ inverse of the spatial lag operator, the proposed approximation is based on the eigenstructure of the spatial weights matrix. Considering the eigendecomposition for a general matrix and the Perron-Frobenius theorem, the inverse matrix can be approximated by a sum of known matrices. Also, a closed formula approximation is now available for each of the elements of the inverse matrix.

This approximation is implemented to the full GMM estimator proposed by Fleming (2004) and detailed by Klier and McMillen (2012). The estimation consists in a two stage iterative procedure, where the inverse of the spatial lag operator is replaced by the corresponding approximation. Closed formulas for the gradients are available. Thus, the estimation time is improved. The proposed estimator is compared to the full GMM estimator and to the Linearized GMM estimator (Klier and McMillen, 2012), in terms of bias, efficiency and computational time.

The simulation study suggest that the approximation of the spatial lag operator inverse is better for binary spatial weights matrices. The proposed estimator appears to perform as well as the full GMM estimator. Under increasing-domain asymptotics (Cressie, 1993), these estimators are likely to be consistent. In addition, the computational time can be drastically reduced for denser matrices.

The proposed estimator also contributes to the literature on regional competitiveness. The spatial spillovers and environmental effects are central issues. A binary competitiveness indicator is established for selected Metropolitan Statistical Areas of the United States, in 2009. The analysis of these subjects is important because competitiveness and environmental impacts have been progressively centering the discussions of political authorities in the U.S., since the early 2000s. A recent report (Porter, Rivkin, Desai, and Raman, 2016) showed a dilapidation of the U.S. economy during the 2000s, culminating with a severe economic crisis in the mid-2000s. Only in 2009, the U.S. economy evidenced stable signs of recovery. Similarly, due to the observable effects of global warming, the environmental impacts of economic growth are now taking part of the political agenda. Does the environmental impacts influence the probability of being competitive? Does being competitive influence the probability of neighboring areas to be competitive, as well? Results suggest that there is a negative effect of environmental indicators over the probability of

being competitive and that there is a positive statistically significant spatial effect over the probability of neighboring areas to be competitive.

The remainder of this paper is organized as follows. In section 2, a brief overview of popular spatial and non-spatial models is provided, considerin the problem of observability of the dependent variable. In section 3, a new method to approximate the spatial lag operator inverse is presented and implemented to reduce the computational time of a GMM estimation. In section 4, the results of simulation studies are discussed, showing the adequacy of the suggested approximation applied to the matrix inversion and estimation. In section 5, an empirical study proves the usefulness of the new estimator, applied to the spatial analysis of competitiveness in the U.S. Metropolitan Statistical Areas, where the environmental effects are central issues. Finally, section 6 concludes. The results of the simulation studies are shown in appendix 1. The results of the empirical study are detailed in appendix 2.

2 Observability in spatial models: specification and estimation

Consider the following model with a spatially lagged dependent variable:

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

where Y_i^* is a general dependent variable (possibly not observable) for the unit i, with $i = 1, 2, \ldots, N$ and N denotes the total number of cross-sectional units. The scalar coefficient $w_{i,j}$ correspond to the spatial weight of unit j on unit i, with $j \neq i$ and $j = 1, 2, \ldots, N$. The scalar parameter α is the corresponding spatial dependence parameter. The $1 \times K$ matrix $\mathbf{X}_i$ 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 parameter vector to be estimated. The disturbance term, ε_i, is an i.i.d. random error for the unit i. Note that the spatial weights, $w_{i,j}$, are non-negative scalars and, by convention, $w_{i,i} = 0$, for all i.

Stacking over the cross-sectional units yields:

$$\mathbf{Y}^* = \alpha \mathbf{W}_N \mathbf{Y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

where $\mathbf{Y}^*$ is now a $N \times 1$ vector of a general dependent variable, $\mathbf{X}$ is a $N \times K$ matrix of exogenous regressors and $\boldsymbol{\varepsilon}$ is a $N \times 1$ vector of i.i.d. random errors. The matrix $\mathbf{I}_N$ is the $N \times N$ identity matrix. The spatial weights, w_{ij}, are the elements of the $N \times N$ spatial weights matrix, $\mathbf{W}_N$. In addition, this matrix is non-negative, because $w_{i,j} \geq 0$.

A particular feature of the model with a spatially lagged dependent variable (or simply, spatial lag model) is the joint determination induced by the simultaneous nature of the spatially lagged term, $\mathbf{W}_N\mathbf{Y}^*$. Just as in time-series models, the spatial lags can be distributed infinitely over the right-hand side of the equation. This is illustrated by the reduced form of the equation (2):

$$\mathbf{Y}^* = \left(\mathbf{I}_N - \alpha \mathbf{W}_N\right)^{-1} \mathbf{X}\boldsymbol{\beta} + \left(\mathbf{I}_N - \alpha \mathbf{W}_N\right)^{-1} \boldsymbol{\varepsilon} \tag{3}$$

where,

$$\left(\mathbf{I}_N - \alpha \mathbf{W}_N\right)^{-1} = \mathbf{I}_N + \alpha \mathbf{W}_N + \alpha^2 \mathbf{W}_N^2 + \ldots = \sum_{k=0}^{\infty} \alpha^k \mathbf{W}_N^k \tag{4}$$

the spatial lag operator inverse, which will converge absolutely if $\|\alpha \mathbf{W}_N\| < 1$ (Bernstein, 2009). Plugging (4) in (3) shows that, for each cross-sectional unit, Y_i^* depends on unit specific explanatory variables and unit specific unobserved effects, and neighbor specific explanatory variables and neighbor specific unobserved effects, as well. In addition, the neighbor specific effects are subject to the powers of $\alpha \mathbf{W}_N$, which can be seen as a measure of the extent of the neighboring effects (Anselin, 1988).

Similarly to time-series models, one expect $|\alpha| < 1$, which implies a decay in the degree of spatial influence over unit i. However, this is a necessary but not a sufficient condition. Observe that the norm of $\alpha \mathbf{W}_N$ also depends on the spectral radius of $\mathbf{W}_N$, $\rho\left(\mathbf{W}_N\right) = |\lambda|_{\max}$, the largest absolute eigenvalue. In practice, the latter condition poses no additional problem, as $\mathbf{W}_N$ is usually normalized in such a way that the rows, or columns, or even both rows and columns, sum up to one, which ensures that $\rho\left(\mathbf{W}_N\right) = 1$. Ord (1981)

suggests to restrict the parameter space of α to $1/\lambda_{\min} < \alpha < 1$, after the normalization of $\mathbf{W}_N$, where $\lambda_{\min}$ is the smallest eigenvalue.

Now consider the following example: $\mathbf{W}_N$ is a 2×2 matrix with ones outside the main diagonal. The eigenvalues are -1 and 1. Therefore, the parameter space can be generalized to $-1 < \alpha < 1$ (Kelejian and Robinson, 1995). Again the norm condition is still verified and the spatial lag operator inverse converges.

With regard to the estimation of equation (1), if Y_i^* is observable, the model is linear and a variety of estimation methods are available, namely: Maximum Likelihood (Ord, 1975; Anselin, 1988; Anselin and Bera, 1998), Bayesian Markov Chain Monte Carlo (LeSage, 1997), instrumental variables (Anselin, 1988, 1990) or (Generalized) Method of Moments (Kelejian and Robinson, 1993; Kelejian and Prucha, 1998).

In a brief review, the (full) maximum likelihood estimator (MLE) and the Bayesian MCMC are computationally intensive estimators. This is a consequence of modeling the full dependence structure of the spatial data, which results in the computation of log-determinants and inversion of N-dimensional matrices. Even so, for a moderate sample size, they are computationally feasible. Under proper specification of the joint density of Y^* conditional on $\mathbf{X}$ both methods produce consistent estimates. Efficiency also relies on the latter assumption.

In contrast to MLE and Bayesian methods, the (Generalized) Method of Moments (GMM) and instrumental variables (IV) approaches stand as reliable alternatives. This is because they are distribution-free methods. As the estimated model results from a conjecture about the true specification, robustness to distributional misspecification is a crucial point. In addition, the estimation procedure does not require the computation of any log-determinant, nor the inversion of N-dimensional matrices, significantly reducing the computational complexity. Under the assumptions established by Kelejian and Prucha (1998), these methods produce consistent estimates. Efficiency is still a major issue for several authors. However, Wooldridge (2010) notes that the GMM with "generated" instruments can be as efficient as MLE. Hence, as the powers of $\mathbf{W}_N\mathbf{X}$ are used as instruments, they are "generated" by the nature of the data. If they are included in the estimation, the GMM in a spatial framework tends to be more efficient.

A different perspective should be taken if Y_i^* is not directly observable. On this framework, the observed dependent variable is now Y_i, which is a function of particular characteristics of Y_i^*. Depending on such characteristics, Y_i may be either a limited dependent variable or a discrete dependent variable. Furthermore, the assumption that Y_i is a linear function of the regressors, may lead to theoretically inadequate interpretations, either for a non-spatial model (Wooldridge, 2010) and for a spatial model (Beron and Vijverberg, 2004).

Case (1992) and McMillen (1992) introduced the analysis of the specification and estimation for binary dependent variables frameworks, more specifically on spatially lagged dependent variables and on spatially dependent errors. As a result of its complexity, many authors are still focused on the spatial binary dependent variable framework, even though generalizations to the spatial multichotomous dependent variable framework appear to be straightforward.

Without loss of generality, a brief discussion on the specification and estimation of non-spatial binary dependent variable models will be addressed.

2.1 Non-spatial binary dependent variable models

Along the lines of Wooldridge (2010), let Y_i^* be not observable and given by:

$$Y_i^* = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i, \qquad i = 1, 2, \ldots, N \tag{5}$$

At this point the cross-sectional units are independent and the regressors are strictly exogenous. The observed variable is Y_i and is determined as: $Y_i = 1$ if $Y_i^* \geq 0$ and $Y_i = 0$ if $Y_i^* < 0$. Therefore, Y_i follows a probabilistic model:

$$\begin{aligned} &\mathrm{E}\left(Y_i \,|\, \mathbf{X}_i\right) = 0 \times \mathrm{P}\left(Y_i = 0 \,|\, \mathbf{X}_i\right) + 1 \times \mathrm{P}\left(Y_i = 1 \,|\, \mathbf{X}_i\right) \\ &= \mathrm{P}\left(Y_i = 1 \,|\, \mathbf{X}_i\right) = \mathrm{P}\left(Y_i^* > 0 \,|\, \mathbf{X}_i\right) \\ &= \mathrm{P}\left(\mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i > 0 \,|\, \mathbf{X}_i\right) = \mathrm{P}\left(\varepsilon_i > -\mathbf{X}_i\boldsymbol{\beta} \,|\, \mathbf{X}_i\right) = G\left(\mathbf{X}_i\boldsymbol{\beta}\right) \end{aligned} \tag{6}$$

In practice, the function $G(\cdot)$ is usually considered to be a cumulative distribution function (CDF) based on the distribution of ε_i conditional on $\mathbf{X}_i$. If $G(\cdot)$ is equal to the CDF of the standard Normal distribution, Y_i will follow a Probit model. If $G(\cdot)$ is equal to the CDF of the standard Logistic distribution, Y_i will follow a Logit model.

In addition, if $\text{Var}(\varepsilon_i \,|\, \mathbf{X}_i) = \sigma_\varepsilon^2$, for all i, the probability in equation (6) must be standardized:

$$\text{P}(Y_i = 1 \,|\, \mathbf{X}_i) = \text{P}(\varepsilon_i > -\mathbf{X}_i\boldsymbol{\beta} \,|\, \mathbf{X}_i) = \text{P}\left(\frac{\varepsilon_i}{\sigma_\varepsilon} < \frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_\varepsilon} \,\middle|\, \mathbf{X}_i\right) = G\left(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right) \tag{7}$$

In terms of estimation under strict exogeneity and cross-sectional independence, there are two popular methods: maximum likelihood and nonlinear least squares. For the first method, assuming that the distribution of ε_i conditional on $\mathbf{X}_i$ is already standardized, the density of Y_i conditional on $\mathbf{X}_i$ can be defined, for a general i, as:

$$f(Y_i \,|\, \mathbf{X}_i; \boldsymbol{\beta}) = G(\mathbf{X}_i\boldsymbol{\beta})^{Y_i} + [1 - G(\mathbf{X}_i\boldsymbol{\beta})]^{(1-Y_i)} \tag{8}$$

The log-likelihood for the whole sample is given by:

$$\boldsymbol{\ell}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \log f(Y_i \,|\, \mathbf{X}_i; \boldsymbol{\beta}) = \sum_{i=1}^{N} \{Y_i \log[G(\mathbf{X}_i\boldsymbol{\beta})] + (1 - Y_i) \log[1 - G(\mathbf{X}_i\boldsymbol{\beta})]\} \tag{9}$$

From the maximization of the log-likelihood function, the unknown parameter vector can be estimated. Consequently, the condition to be solved is $\text{E}[\mathbf{s}_i(\boldsymbol{\beta}) \,|\, \mathbf{X}_i] = 0$, where $\mathbf{s}_i(\boldsymbol{\beta})$ is the score function for the unit i.

Gourieroux, Monfort, and Trognon (1984) show that, under regularity conditions and arbitrarily distributional misspecification of the density $\mathbf{Y} \,|\, \mathbf{X}$, the maximum likelihood estimator is consistent and asymptotically normal. Robust inference is standard.

For the second method, estimation is addressed without imposing any distributional assumption. Hence, instead of modeling the latent variable, Y_i^*, and deduce the probability

based on the observed variable, Y_i, a nonlinear model is directly established for Y_i, commonly known as a single index model:

$$Y_i = G\left(\mathbf{X}_i\boldsymbol{\beta}\right) + \varepsilon_i, \qquad i = 1, 2, \ldots, N \tag{10}$$

where $G\left(\cdot\right)$ is a link function and $\mathbf{X}_i\boldsymbol{\beta}$ is the corresponding index of the function. Estimation is addressed by the nonlinear least squares (NLS) estimator and results from the minimization of:

$$\frac{1}{N}\sum_{i=1}^{N}\left[Y_i - G\left(\mathbf{X}_i\boldsymbol{\beta}\right)\right]^2 \tag{11}$$

Observe that the expression above does not have a closed form. Even so, the optimization of a nonlinear function can be done numerically. One of the most popular method for nonlinear optimization is the Gauss-Newton algorithm. It is based on a first order Taylor expansion of the function $G\left(\cdot\right)$ around an initial estimate, $\boldsymbol{\beta}^{(0)}$:

$$G\left(\mathbf{X}_i\boldsymbol{\beta}\right) \approx G\left(\mathbf{X}_i\boldsymbol{\beta}^{(0)}\right) + \left.\frac{\partial G}{\partial \boldsymbol{\beta}^{\intercal}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(0)}}\left(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}\right) \tag{12}$$

and the iterative solution is given by the ordinary least squares (OLS) estimates of the previous equation in each step. As a result, the estimates in each iteration are updated using the follow formula:

$$\begin{aligned}\boldsymbol{\beta}^{(s+1)} &= \left[\left(g^{(s)}\right)^{\intercal}\left(g^{(s)}\right)\right]^{-1}\left(g^{(s)}\right)^{\intercal}\left(G - G^{(s)} + g^{(s)}\boldsymbol{\beta}^{(s)}\right)\\ &= \boldsymbol{\beta}^{(s)} + \left[\left(g^{(s)}\right)^{\intercal}\left(g^{(s)}\right)\right]^{-1}\left(g^{(s)}\right)^{\intercal}\boldsymbol{\varepsilon}^{(s)}\end{aligned} \tag{13}$$

where $g^{(s)}$ is the first derivative of function $G^{(s)}$ and $\boldsymbol{\varepsilon}^{(s)}$ is the $N \times 1$ vector of NLS residuals, all evaluated at the estimates of step s, with $s = 0, 1, \ldots, S-1$ and S is the step where convergence is achieved.

Under classic continuity assumptions of the function $G\left(\cdot\right)$, the NLS estimator is consistent and robust to any functional form misspecification of the variance of the random error.

Note that, if $\text{Var}(\varepsilon_i \,|\, \mathbf{X}_i) = \sigma_\varepsilon^2$, the index should also be standardized. This guarantees that the NLS estimates are comparable to the maximum likelihood estimates.

Next spatial dependence will be introduced through the binary dependent variable and the available estimation methods will be discussed.

2.2 Spatial dependence in binary dependent variable models

Consider a spatially lagged binary dependent variable model. From the results discussed in the previous sections, this model follows the latent variable model from equation (3). However, the error is now given by $\mathbf{u} = (\mathbf{I}_N - \alpha \mathbf{W}_N)^{-1} \boldsymbol{\varepsilon}$, which implies that:

$$\text{Var}(\mathbf{u} \,|\, \mathbf{X}, \mathbf{W}_N) = \left[(\mathbf{I}_N - \alpha \mathbf{W}_N)^\intercal (\mathbf{I}_N - \alpha \mathbf{W}_N)\right]^{-1} = \boldsymbol{\Sigma} \tag{14}$$

assuming that $\text{Var}(\boldsymbol{\varepsilon} \,|\, \mathbf{X}, \mathbf{W}_N) = \mathbf{I}_N$. As a result, the model for the observed binary dependent variable is given by:

$$\mathbf{Y} = G\left((\mathbf{I}_N - \alpha \mathbf{W}_N)^{-1} \mathbf{X}^{\#} \boldsymbol{\beta}\right) + \mathbf{u} \tag{15}$$

where, $\mathbf{X}^{\#}$ is an $N \times K$ matrix of transformed regressors, due to the heteroskedastic nature of the error, $\mathbf{u}$. Concretely, $\mathbf{X}_i^{\#} = \mathbf{X}_i / \sigma_i$, with

$$\sigma_i^2 = \text{diag}\left(\left[(\mathbf{I}_N - \alpha \mathbf{W}_N)^\intercal (\mathbf{I}_N - \alpha \mathbf{W}_N)\right]^{-1}\right)_{i,i} \tag{16}$$

The estimation of nonlinear models under spatial dependence turns to be, in most cases, a tremendous computational task. Still, all the methods seen in the previous sections apply to this framework as well.

Concretely, applying (full) maximum likelihood to estimate a nonlinear model with a spatially lagged binary dependent variable implies that the distribution of $\mathbf{u}$ conditional on $\mathbf{X}$ follows a multivariate distribution. Here the log-likelihood for the whole sample is given by:

$$\boldsymbol{\ell}(\boldsymbol{\beta}, \alpha) = \int_{\mathcal{A}_N} \int_{\mathcal{A}_{N-1}} \dots \int_{\mathcal{A}_2} \int_{\mathcal{A}_1} g(\mathbf{u}) \, \mathrm{d}\mathbf{u} \tag{17}$$

where $g(\mathbf{u})$ is the multivariate PDF and $\mathcal{A}_i$, $i = 1, 2, \dots, N$, is an interval such that:

$$\mathcal{A}_i = \begin{cases} (-\infty, \mathbf{X}_i^{\#\#}\boldsymbol{\beta}), & \text{if } Y_i = 1 \\ (\mathbf{X}_i^{\#\#}\boldsymbol{\beta}, \infty), & \text{if } Y_i = 0 \end{cases} \tag{18}$$

with $\mathbf{X}_i^{\#\#}$ the ith row of the matrix $(\mathbf{I}_N - \alpha\mathbf{W}_N)^{-1}\mathbf{X}^{\#}$.

Under correct distributional specification, the major drawback of this method is related to the computation of the N-dimensional integral, which requires extraordinary computer capability. Still, several authors suggest estimation procedures based on maximum likelihood methods. Case (1992) estimates a spatial Probit and assumes a block diagonal structure for the spatial weights matrix, where each block has $N_b - 1$ neighbors, with $b = 1, 2, \dots, B$ and B is the number of blocks. This translates into a generic element $w_{ij} = 1/(N_g - 1)$ for the block diagonal row-standardized spatial weights matrix and allows the simplification of $\mathbf{I}_N - \alpha\mathbf{W}_N$. Under these assumptions, the model of interest follows a Chamberlain-Mundlak mean augmented regression, with spatially dependent errors. The off-diagonal information of $\boldsymbol{\Sigma}$ is neglected and estimation is simplified to an i.i.d. standardized Probit. McMillen (1992) uses the expectation-maximization algorithm (EM algorithm) to obtain estimates for a Probit with spatial dependence (in the dependent variable and in the error term). Only the diagonal information of $\boldsymbol{\Sigma}$ is considered in the estimation, therefore the expectations are derived from an heteroskedastic Probit. The off-diagonal information of $\boldsymbol{\Sigma}$ is obtained after the estimation. Beron and Vijverberg (2004) and Pace and LeSage (2016) consider the RIS simulator and the GHK simulator, respectively, to evaluate the N-dimensional integral. Bhat (2011) and Wang, Iglesias, and Wooldridge (2013) consider the pairs of closest observations and derive a partial maximum likelihood method to estimate a Probit with spatially dependent errors (the first work also generalizes the estimation to the spatially lagged dependent variable framework). Martinetti and Geniaux (2017) consider the Mendell-Elston approximation method to compute the multivariate Normal probabilities as the product of univariate conditional probabilities. However, the previous methods either fail to account for the

full dependence structure of the spatial data or they are computationally burdensome even for moderate sample sizes.

Just as maximum likelihood, the Bayesian approach is also widely used in the estimation of spatial nonlinear models. The methodology is based on the works of LeSage (2000) and Smith and LeSage (2004). In the first, the probabilities are simulated by the Gibbs sampler, whereas in the second it is based on Markov Chains Monte Carlo (MCMC). Again, these methods are computationally burdensome.

Again distribution-free methods stand as reliable alternatives due to their robustness to distributional misspecification. Pinkse and Slade (1998) consider the GMM estimation of an heteroskedastic Probit with spatially dependent errors. Only the diagonal information of $\mathbf{\Sigma}$ is considered, thus one obtain estimates for the univariate conditional probability. In general, estimators that use only the diagonal information of $\mathbf{\Sigma}$ produce consistent estimates, but they are extremely inefficient compared to full MLE approaches (Wang et al., 2013). However, even if it was computationaly feasible to estimate the multivariate probability, there are several problems concerning the interpretation of such quantity (Greene, 2003).

Fleming (2004) suggested a full GMM estimation approach of a Probit with spatial dependence (in both the dependent variable and in the error term), to overcome such issues. In particular, the structural form of a Probit with spatially lagged dependent variable is considered, similar to the specification of equation (15). Because the spatial lag operator inverse, $(\mathbf{I}_N - \alpha \mathbf{W}_N)^{-1}$, is introduced in the objective function, the full structure of spatial dependence is considered in the estimation. The objective function follows as:

$$\mathcal{Q}(\boldsymbol{\beta}, \alpha) = \mathbf{u}_*^{\intercal} \mathbf{Z} \mathbf{\Xi} \mathbf{Z}^{\intercal} \mathbf{u}_* \tag{19}$$

where $\mathbf{Z}$ is the $N \times (k+p)$ matrix of instruments, with $\mathbf{X}$ and powers (p) of $\mathbf{W}_N\mathbf{X}$ as instrumental variables (it is usual to consider, $p = 2$). The $(k+p) \times (k+p)$ matrix $\mathbf{\Xi}$ is a symmetric positive definite matrix and the $N \times 1$ vector $\mathbf{u}_*$ is the generalized residuals vector. Gourieroux, Monfort, Renault, and Trognon (1987) demonstrate that the generalized residuals have desirable properties in comparison to those obtained from the pure NLS estimation. The generalized residuals are given by:

$$u_{*,i} = \frac{[Y_i - G(\cdot)]\, g(\cdot)}{G(\cdot)\,[1 - G(\cdot)]}, \qquad i = 1, 2, \ldots, N \tag{20}$$

where function $g(\cdot)$ is the first derivative of $G(\cdot)$ and the indexes of these functions are $(\mathbf{I}_N - \alpha \mathbf{W}_N)^{-1} \mathbf{X}\boldsymbol{\beta}$ instead of $(\mathbf{I}_N - \alpha \mathbf{W}_N)^{-1} \mathbf{X}^{\#}\boldsymbol{\beta}$. This last consideration is the major drawback of this approach, as the index is not rescaled by the diagonal elements of $\boldsymbol{\Sigma}$. As a consequence, the estimates of the parameter vector will be biased. Klier and McMillen (2012) presents two estimation approaches, based on equation (15), which allows to solve the drawback of the latter work. In addition, it generalizes the estimator of Pinkse and Slade to a Probit with a spatially lagged dependent variable.

Along the lines of Klier and McMillen (2012), the objective function for the Probit with spatially lagged dependent variable equal to the objective function of Fleming estimator, but $\boldsymbol{\Xi} = (\mathbf{Z}^{\intercal}\mathbf{Z})^{-1}$. This implies that the full GMM estimator reduces to a nonlinear two stages least squares (N2SLS) estimator. Again, as the objective function in (19) does not have a closed form, an iterative procedure is used. Similarly to the Gauss-Newton algorithm, in equation (13), 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 $\boldsymbol{\Gamma}_i = (\partial u_{*,i}/\partial \boldsymbol{\Theta})|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{(0)}}$, $i = 1, 2, \ldots, N$.
2. Regress $\boldsymbol{\Gamma}$ on $\mathbf{Z}$, in a similar fashion to (linear) 2SLS. Obtain $\hat{\boldsymbol{\Gamma}}$.
3. Construct new estimates as $\boldsymbol{\Theta}^{(s+1)} = \boldsymbol{\Theta}^{(s)} + \left[\left(\hat{\boldsymbol{\Gamma}}^{(s)}\right)^{\intercal}\left(\hat{\boldsymbol{\Gamma}}^{(s)}\right)\right]^{-1}\left(\hat{\boldsymbol{\Gamma}}^{(s)}\right)^{\intercal}\mathbf{u}_*^{(s)}$, where $\mathbf{u}_*^{(s)}$ are the generalized residuals evaluated at the estimates of step s.
4. Repeat steps 1. to 3., using the estimates from the last iteration, until the algorithm converges.

The estimated asymptotic variance of the N2SLS estimator is:

$$\widehat{Avar}\left(\hat{\boldsymbol{\Theta}}\right) = \left(\hat{\boldsymbol{\Gamma}}^{\intercal}\hat{\boldsymbol{\Gamma}}\right)^{-1}\left[\sum_{i=1}^{N} \hat{u}_i^2 \hat{\boldsymbol{\Gamma}}_i^{\intercal}\hat{\boldsymbol{\Gamma}}_i\right]\left(\hat{\boldsymbol{\Gamma}}^{\intercal}\hat{\boldsymbol{\Gamma}}\right)^{-1} \tag{21}$$

For a general function $G(\cdot)$, the individual gradients for each parameter are given by:

$$(\mathbf{\Gamma}_{\boldsymbol{\beta}})_i = \frac{\partial u_{*,i}}{\partial \boldsymbol{\beta}^{\intercal}} = -u_{*,i}\left(\frac{\nabla g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right)}{g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right)} - u_{*,i}\right)\mathbf{X}_i^{\#\#}, \qquad i = 1, 2, \ldots, N \tag{22}$$

and

$$(\mathbf{\Gamma}_{\alpha})_i = \frac{\partial u_{*,i}}{\partial \alpha} = -u_{*,i}\left(\frac{\nabla g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right)}{g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right)} - u_{*,i}\right)\left[\mathbf{H}_i\boldsymbol{\beta} - \frac{\mathbf{X}_i^{\#\#}\boldsymbol{\beta}}{2\sigma_i^2}\Upsilon_{ii}\right], \qquad i = 1, 2, \ldots, N \tag{23}$$

where $\nabla g\left(\cdot\right)$ is the second derivative of the function $G\left(\cdot\right)$, $\mathbf{H}_i$ is the ith row of the matrix $\left(\mathbf{I}_N - \alpha\mathbf{W}_N\right)^{-1}\mathbf{W}_N\mathbf{X}^{\#\#}$ and Υ_{ii} is the ith element of the diagonal of the matrix:

$$\mathbf{\Upsilon} = \left(\mathbf{I}_N - \alpha\mathbf{W}_N\right)^{-1}\left\{\mathbf{W}_N\left(\mathbf{I}_N - \alpha\mathbf{W}_N\right)^{-1} + \left[\mathbf{W}_N\left(\mathbf{I}_N - \alpha\mathbf{W}_N\right)^{-1}\right]^{\intercal}\right\}\left[\left(\mathbf{I}_N - \alpha\mathbf{W}_N\right)^{-1}\right]^{\intercal} \tag{24}$$

For a Probit with a spatially lagged dependent variable, the term $\nabla g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right)/g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right) = -\mathbf{X}_i^{\#\#}$.

At this point, the estimation of the Probit with a spatially lagged dependent variable by N2SLS appears to be a feasible solution. The closed forms for the gradient matrix help to accelerate the optimization process, because there is no need to compute them numerically. However, on each iteration, one has to compute the inverse $\left(\mathbf{I}_N - \alpha\mathbf{W}_N\right)^{-1}$, which is a time-consuming operation, especially for a moderate sample size.

As a result, Klier and McMillen (2012) suggest a procedure that allows for a simplification of the estimation procedure and to reduce the computational burden. It consists in a linearization of the model around $\alpha = 0$. This choice is obvious: no matrices need to be inverted and none of the gradients are equal to zero. In addition, the error term, $\mathbf{u}$, now has a constant variance, which implies that the regressors no longer have to be rescaled, $\mathbf{X}_i^{\#\#} = \mathbf{X}_i$, and one can obtain initial estimates of $\boldsymbol{\beta}$ by standard Probit estimation. Furthermore, for a general function $G\left(\cdot\right)$, the gradients simplify to:

$$(\mathbf{\Gamma}_{\boldsymbol{\beta}})_i = \frac{\partial u_{*,i}}{\partial \boldsymbol{\beta}^{\intercal}} = -u_{*,i}\left(\frac{\nabla g\left(\mathbf{X}_i\boldsymbol{\beta}\right)}{g\left(\mathbf{X}_i\boldsymbol{\beta}\right)} - u_{*,i}\right)\mathbf{X}_i, \qquad i = 1, 2, \ldots, N \tag{25}$$

and

$$(\mathbf{\Gamma}_\alpha)_i = \frac{\partial u_{*,i}}{\partial \alpha} = -u_{*,i}\left(\frac{\nabla g\left(\mathbf{X}_i\boldsymbol{\beta}\right)}{g\left(\mathbf{X}_i\boldsymbol{\beta}\right)} - u_{*,i}\right)\mathbf{H}_i\boldsymbol{\beta}, \qquad i = 1, 2, \ldots, N \tag{26}$$

where, now $\mathbf{H}_i$ is the ith row of the matrix $\mathbf{W}_N\mathbf{X}$. For a Probit with a spatially lagged dependent variable, the term $\nabla g\left(\mathbf{X}_i\boldsymbol{\beta}\right)/g\left(\mathbf{X}_i\boldsymbol{\beta}\right) = -\mathbf{X}_i$.

In terms of estimation, it is similar to the iterative procedure of the full GMM. However, only one step is required to obtain estimates for the parameter vector $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \alpha)^\intercal$, because $\boldsymbol{\beta}$ can be consistently estimated by standard Probit and the gradients do not depend on α. Therefore, in step 3., there is no need to update the estimates. Simply regress $\mathbf{u}_*^{(0)} + (\mathbf{\Gamma}_{\boldsymbol{\beta}})^\intercal \boldsymbol{\beta}^{(0)}$ on $\mathbf{\Gamma}_{\boldsymbol{\beta}}$ and $\mathbf{\Gamma}_\alpha$. The corresponding coefficients are the estimated values of $\boldsymbol{\beta}$ and α.

The estimated asymptotic variance of the Linearized GMM estimator is:

$$\widehat{Avar}\left(\hat{\boldsymbol{\Theta}}\right) = \left(\sum_{i=1}^{N}\hat{u}_i^2\right)\left(\hat{\mathbf{\Gamma}}^\intercal\hat{\mathbf{\Gamma}}\right)^{-1} \tag{27}$$

Observe that the Linearized GMM only requires a single iteration to achieve convergence and produce estimates. Thus, it outstands all of the remaining estimation methods, in terms of computational time. In addition, no N-dimensional matrices need to be inverted, largely reducing the computational complexity of this method. However, as the linearization is made around $\alpha = 0$, this method only provide good estimates for the spatial lag parameter if $|\alpha| \leq 0.5$.

At this point, either the full GMM or the Linearized GMM have disadvantages, whether due to excessive computational complexity, or to excessive simplification that may lead to inconsistency. A middle-ground solution may rely on an approximation procedure applied to the most complex matrix operation, the inversion of $(\mathbf{I}_N - \alpha\mathbf{W}_N)$, that is able to turn the estimation a computationally feasible operation for large samples. Below, the approach for this approximation procedure is presented.

3 Approximation of the spatial lag operator inverse

Consider the spatial lag operator inverse, $(\mathbf{I}_N - \alpha \mathbf{W}_N)^{-1}$. For general matrices, the computational complexity of the matrix inversion is $O(N^3)$. As a consequence, even for a moderate size sample, this is an extremely time-demanding operation. Recognizing the restriction that such kind of issue poses to a practitioner, the main goal is to approximate the inverse by less complex matrix operations that take into account the characteristics of the eigenstructure of $\mathbf{W}$ (for notational ease, the subscripts N are dropped).

Recall that $\mathbf{W}$ is a non-negative normalized matrix. This normalization is based on an initial matrix $\mathbf{W}_0$, that may be non-symmetric. Let $\mathbf{D}_R$ be a diagonal matrix, where the row sums of $\mathbf{W}_0$ are the diagonal elements. The matrix $\mathbf{W} = \mathbf{D}_R^{-1}\mathbf{W}_0$ has row vectors that sum up to one, hence it is row stochastic. The transpose of a row stochastic matrix, $\mathbf{W}^\intercal = \mathbf{W}_0^\intercal \mathbf{D}_R^{-1}$ has column vectors that sum up to one, hence it is a column stochastic matrix. To obtain a doubly stochastic matrix an iterative procedure is required, based on $\mathbf{W}^{(s+1)} = \left(\mathbf{D}_R^{(s)}\right)^{-1/2} \mathbf{W}^{(s)} \left(\mathbf{D}_R^{(s)}\right)^{-1/2}$, with $s = 1, 2, \ldots, S$. Only at the convergence step, S, both row and column vectors sum up to one. Notice that, at step $s = 0$, $\mathbf{W}^{(1)} = \mathbf{D}_R^{-1/2}\mathbf{W}_0\mathbf{D}_R^{-1/2}$. The latter matrix is similar to $\mathbf{W}$ and to $\mathbf{W}^\intercal$. In other words, $\mathbf{W}^{(1)} = \mathbf{W}_{sim}$ has the same eigenvalues that $\mathbf{W}$ and $\mathbf{W}^\intercal$.

On this framework, matrix similarity and the symmetry of $\mathbf{W}_0$ play a major role on the approximation procedure, as it will be shown. For now, because $\mathbf{W}$ and $\mathbf{W}^\intercal$ are always non-symmetric matrices, the following theorem establish the result on the eigendecomposition of a general matrix:

Theorem 1.1 (Bernstein, 2009, p. 339)**.** *Let $\mathbf{W} \in \mathbb{C}^{n\times n}$, assume that $\mathbf{W}$ is diagonalizable over $\mathbb{C}$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$, and let $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ are linearly independent eigenvectors of $\mathbf{W}$ associated with $\lambda_1, \lambda_2, \ldots, \lambda_N$, then $\mathbf{W} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, where $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \ldots \ \mathbf{v}_N]$. Conversely, if $\mathbf{V} \in \mathbb{C}^{n\times n}$ is nonsingular and $\mathbf{W} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, then the columns of $\mathbf{V}$ are the associated eigenvectors.*

Also, as $\mathbf{W}$ is non-negative, the next theorem provide the main result on the eigenstructure of such matrices. This theorem is also known as the Perron-Frobenius theorem:

Theorem 1.2 (Perron-Frobenius)**.** *Let $\mathbf{W} \in \mathbb{C}^{n\times n}$, where $n \geq 2$, and assume that $\mathbf{W}$ is non-negative. Then,*

i) The spectral radius of $\mathbf{W}$*,* $\rho(\mathbf{W})$*, is an eigenvalue of* $\mathbf{W}$*;*

ii) There exists a nonzero nonnegative vector $\mathbf{v} \in \mathbb{C}^n$ *such that* $\mathbf{W}\mathbf{v} = \rho(\mathbf{W})\,\mathbf{v}$

For a row stochastic matrix, $\mathbf{W}$, its spectral radius is equal to one, then it must be true that $\mathbf{W}\boldsymbol{\iota} = \boldsymbol{\iota}$, where $\boldsymbol{\iota}$ is a vector of ones (Bernstein, 2009, p. 275). Therefore, the largest eigenvector of $\mathbf{W}$ associated with the largest eigenvalue is $\mathbf{v}_1 = \boldsymbol{\iota}$.

Now, consider the Taylor series expansion of the inverse, as in equation (4). Under the result of theorem 1.1, the following holds:

$$\begin{aligned}(\mathbf{I} - \alpha\mathbf{W})^{-1} &= \mathbf{I} + \alpha\mathbf{W} + \alpha^2\mathbf{W}^2 + \alpha^3\mathbf{W}^3 + \dots \\ &= \mathbf{I} + \alpha\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} + \alpha^2\left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\right)^2 + \alpha^3\left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\right)^3 + \dots \\ &= \mathbf{I} + \alpha\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} + \alpha^2\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} + \alpha^3\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} + \dots \\ &= \mathbf{I} + \alpha\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} + \alpha^2\mathbf{V}\boldsymbol{\Lambda}^2\mathbf{V}^{-1} + \alpha^3\mathbf{V}\boldsymbol{\Lambda}^3\mathbf{V}^{-1} + \dots\end{aligned} \tag{28}$$

and $(\mathbf{I} - \alpha\mathbf{W})^{-1}$ can be rewritten as $\mathbf{V}(\mathbf{I} - \alpha\boldsymbol{\Lambda})^{-1}\mathbf{V}^{-1}$. However, this would yield no computational gains. All the eigenvalues and eigenvectors must be computed and a matrix still needs to be inverted. In practice, one would be worse off.

Nevertheless, from the previous equation, it is possible to conclude that, as $p \to \infty$, the powers of the Taylor expansion converge to:

$$\lim_{p\to\infty} \mathbf{W}^p = \lim_{p\to\infty} \left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}\right)^p = \mathbf{V}\left(\lim_{p\to\infty} \boldsymbol{\Lambda}^p\right)\mathbf{V}^{-1} \tag{29}$$

If the largest absolute eigenvalue of $\mathbf{W}$ is assumed to have algebraic multiplicity of one[1], then $\lim_{p\to\infty} \boldsymbol{\Lambda}^p = \lim_{p\to\infty} \operatorname{diag}(1^p, \lambda_2^p, \dots, \lambda_N^p)$. But the remaining eigenvalues are, in absolute value, less than one, so $\boldsymbol{\Lambda}^\infty = \operatorname{diag}(1, 0, \dots, 0)$. In other words, the "long run" matrix, $\mathbf{W}^\infty$, is equal to $\operatorname{col}(\mathbf{V})_1 \operatorname{row}(\mathbf{V}^{-1})_1$, with $\operatorname{col}(\mathbf{V})_1 = \boldsymbol{\iota}$.

[1]This assumption can be relaxed at a cost of approximation accuracy.

To fully identify $\mathbf{W}^{\infty}$, the first row of the inverse eigenvector matrix needs to be computed. This poses an additional problem, because, to obtain row $(\mathbf{V}^{-1})_1$, one has to solve the entire linear system.

Nevertheless, recall that the matrices $\mathbf{W}$ and $\mathbf{W}_{sim}$ are similar, therefore the eigenvectors of $\mathbf{W}$ can be expressed as a function of the eigenvectors of $\mathbf{W}_{sim}$. In addition, if $\mathbf{W}_0$ is symmetric, then $\mathbf{W}_{sim}$ is symmetric as well and its eigenvectors have desirable properties that will allow to simplify the problem. However, if $\mathbf{W}_0$ is not symmetric, a different approach should be taken.

Based on the two scenarios of $\mathbf{W}_0$, the approaches to obtain $\mathbf{W}^{\infty}$ are discussed below.

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

The following theorem establishes the result on the eigendecomposition of a symmetric matrix:

Theorem 1.3 (Bernstein, 2009, p. 314)**.** *Let $\mathbf{W}_{sim} \in \mathbb{C}^{n\times n}$ and assume that $\mathbf{W}_{sim}$ is symmetric. Then there exists a unitary matrix (resp. orthogonal matrix in $\mathbb{R}^{n\times n}$) $\mathbf{V}_{sim}$ such that $\mathbf{W}_{sim} = \mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{\intercal}$.*

Because $\mathbf{V}_{sim}$ is a unitary matrix (*resp.* orthogonal matrix), the euclidean norm of each eigenvector vector of $\mathbf{V}_{sim}$ is equal to one,

$$\|\mathbf{v}_{sim,j}\|_2 = \sqrt{v_{sim,(1,j)}^2 + v_{sim,(2,j)}^2 + \ldots + v_{sim,(N,j)}^2} = 1, \qquad j = 1, 2, \ldots, N \tag{30}$$

Rewrite $\mathbf{W}_{sim}$ as $\mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2}$. Consider the eigendecomposition for these matrices. This yields:

$$\mathbf{W}_{sim} = \mathbf{D}_R^{1/2}\mathbf{W}\mathbf{D}_R^{-1/2} \Leftrightarrow \mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{\intercal} = \mathbf{D}_R^{1/2}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}\mathbf{D}_R^{-1/2} \tag{31}$$

Due to similarity, it is straightforward that $\mathbf{V}_{sim} = \mathbf{D}_R^{1/2}\mathbf{V}$ and $\mathbf{V}_{sim}^{\intercal} = \mathbf{V}^{-1}\mathbf{D}_R^{-1/2}$. But, $\mathbf{V}_{sim}^{\intercal} = \mathbf{V}^{\intercal}\mathbf{D}_R^{1/2}$. Therefore, at the "long run",

$$\begin{aligned}\mathbf{W}_{sim}^{\infty} &= \frac{1}{\left\|\mathbf{D}_R^{1/2}\mathrm{col}\,(\mathbf{V})_1\right\|_2\left\|\mathrm{col}\,(\mathbf{V})_1^{\intercal}\,\mathbf{D}_R^{1/2}\right\|_2} \times \mathbf{D}_R^{1/2}\mathrm{col}\,(\mathbf{V})_1\,\mathrm{col}\,(\mathbf{V})_1^{\intercal}\,\mathbf{D}_R^{1/2} \\ &= \frac{1}{\left[\sqrt{\left(d_{R,1}^{1/2}\right)^2+\left(d_{R,2}^{1/2}\right)^2+\ldots+\left(d_{R,N}^{1/2}\right)^2}\right]^2}\mathbf{D}_R^{1/2}\boldsymbol{\iota}\boldsymbol{\iota}^{\intercal}\mathbf{D}_R^{1/2} \\ &= \left[\sum_{i=1}^{N} d_{R,i}\right]^{-1}\mathbf{D}_R^{1/2}\mathbf{J}\mathbf{D}_R^{1/2} = \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,(i,j)}\right]^{-1}\mathbf{D}_R^{1/2}\mathbf{J}\mathbf{D}_R^{1/2}\end{aligned} \tag{32}$$

where $w_{0,(i,j)}$ are the (i,j) elements of $\mathbf{W}_0$ and $\mathbf{J}$ is a matrix of ones. The matrix is rescaled by the sum of all elements of $\mathbf{W}_0$ because $\mathbf{V}_{sim}$ has to be unitary (*resp.* orthogonal). Now, to obtain $\mathbf{W}^{\infty}$ rewrite $\mathbf{W}$ as $\mathbf{D}_R^{-1/2}\mathbf{W}_{sim}\mathbf{D}_R^{1/2}$. This yields:

$$\mathbf{W}^{\infty} = \mathbf{D}_R^{-1/2}\mathbf{W}_{sim}^{\infty}\mathbf{D}_R^{1/2} = \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,(i,j)}\right]^{-1}\mathbf{J}\mathbf{D}_R \tag{33}$$

Observe that the matrix product $\mathbf{J}\mathbf{D}_R$ can be simplified to a less demanding matrix-vector product, since $\mathbf{D}_R$ is a diagonal matrix. Furthermore, the stochastic property of $\mathbf{W}^{\infty}$ is always equal to the stochastic property of $\mathbf{W}$. Note that, if $\mathbf{W}$ is doubly stochastic, the result of theorems 1.2 and 1.3 directly apply and $\mathbf{W}^{\infty}$ can be even more simplified. These results are crucial for the approximation of $(\mathbf{I}-\alpha\mathbf{W})^{-1}$ under various stochastic scenarios for the matrix $\mathbf{W}$.

Several authors (LeSage and Pace, 2009; Arbia, 2014; Elhorst, 2014, to name a few) suggest to consider the Taylor series expansion of the inverse and apply a truncation at small powers $\mathbf{W}$. Still, this operation can be computationally demanding, because several matrix products need to be calculated. If the powers $p \geq 2$ of $\mathbf{W}$ are all substituted by $\mathbf{W}^{\infty}$, the computational demand of this operation is minimal. Hence, the spatial lag operator inverse is approximated by:

$$
\begin{aligned}
(\mathbf{I}-\alpha\mathbf{W})^{-1} &= \mathbf{I}+\mathbf{W}+\alpha^2\mathbf{W}^2+\alpha^3\mathbf{W}^3+\dots \\
&\approx \mathbf{I}+\alpha\mathbf{W}+\alpha^2\mathbf{W}^{\infty}+\alpha^3\mathbf{W}^{\infty}+\dots \\
&= \mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty} \\
&= \mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,(i,j)}\right]^{-1}\mathbf{J}\mathbf{D}_R
\end{aligned}
\tag{34}
$$

and the convergence condition of the Taylor series expansion in equation (4) still holds. This involves no matrix products and the operations implied are based on quantities that have been already computed. Note that the quality of this approximation is determined by the eigenvalues of order $r \geq 2$. Specifically, it depends on how fast the remaining eigenvalues converge to zero. In fact, this approximation is related to the method proposed by Griffith (2000), applied to linear models.

3.2 Case 2: nonsymmetric $\mathbf{W}_0$

In the case where $\mathbf{W}_0$ is non-symmetric, there is at least one unit j that does not influence unit i, but the converse is not true. In addition, $\mathbf{W}_{sim}$ is no longer symmetric and the assumptions of theorem 1.3 do not hold. As a consequence, the eigendecomposition of $\mathbf{W}_{sim}$ is given by $\mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{-1}$.

Consider the eigendecomposition of $\mathbf{W}$ and $\mathbf{W}_{sim}$. Now, the eigenvectors relate as $\mathbf{V}_{sim} = \mathbf{D}_R^{1/2}\mathbf{V}$ and $\mathbf{V}_{sim}^{-1} = \mathbf{V}^{-1}\mathbf{D}_R^{-1/2}$. As a result, a matrix inversion is still required.

Just as in the first scenario, a symmetric matrix is needed to approximate $\mathbf{W}_{sim}^{\infty}$. Hence, a symmetrization procedure is applied to $\mathbf{W}_0$, such that the nonzero elements (i,j) are equal to the zero elements (j,i) and vice-versa, with $i,j = 1,2,\dots,N$ and $i \neq j$. This procedure follows as:

$$
\mathbf{W}_0^{*} = \frac{1}{2}\left\{\mathbf{W}_0+\mathbf{W}_0^{\intercal}+\left[(\mathbf{W}_0-\mathbf{W}_0^{\intercal})^{\circ 2}\right]^{\circ 1/2}\right\} \tag{35}
$$

where "$\circ 2$" is the Hadamard square and "$\circ 1/2$" is the Hadamard square root, respectively $\left(w_{0,(i,j)} - w_{0,(j,i)}\right)^2$ and $\left[\left(w_{0,(i,j)} - w_{0,(j,i)}\right)^2\right]^{1/2}$. If the norm of $\left[\left(\mathbf{W}_0 - \mathbf{W}_0^\intercal\right)^{\circ 2}\right]^{\circ 1/2}$ is small, it implies that $\mathbf{W}_0 \approx \mathbf{W}_0^\intercal$. On this scenario, $\mathbf{W}_0^* \approx \mathbf{W}_0$.

Even though $\mathbf{W}_0^*$ may be close to $\mathbf{W}_0$, the row sums are different. This implies that the normalized matrices also differ. Concretely, based on $\mathbf{W}_0^*$, the row stochastic matrix is $\mathbf{W}^* = \mathbf{D}_{R^*}^{-1}\mathbf{W}_0^*$ and the similar matrix is $\mathbf{W}_{sim}^* = \mathbf{D}_{R^*}^{-1/2}\mathbf{W}_0^*\mathbf{D}_{R^*}^{-1/2}$, where $\mathbf{D}_{R^*}$ is a diagonal matrix with diagonal elements equal to the row sums of $\mathbf{W}_0^*$.

Similarly to the previous case, rewrite $\mathbf{W}_{sim}^*$ as $\mathbf{D}_{R^*}^{1/2}\mathbf{W}^*\mathbf{D}_{R^*}^{-1/2}$. Consider the eigendecomposition for these matrices. This yields:

$$\mathbf{W}_{sim}^* = \mathbf{D}_{R^*}^{1/2}\mathbf{W}^*\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \mathbf{V}_{sim}^*\mathbf{\Lambda}^*\left(\mathbf{V}_{sim}^*\right)^\intercal = \mathbf{D}_{R^*}^{1/2}\mathbf{V}^*\mathbf{\Lambda}^*\left(\mathbf{V}^*\right)^{-1}\mathbf{D}_{R^*}^{-1/2} \tag{36}$$

It is straightforward that $\mathbf{V}_{sim}^* = \mathbf{D}_{R^*}^{1/2}\mathbf{V}^*$ and $\left(\mathbf{V}_{sim}^*\right)^\intercal = \left(\mathbf{V}^*\right)^{-1}\mathbf{D}_{R^*}^{-1/2}$. But, $\left(\mathbf{V}_{sim}^*\right)^\intercal = \left(\mathbf{V}^*\right)^\intercal\mathbf{D}_{R^*}^{1/2}$.

In addition, if $\mathbf{W}_{sim}^*$ is rewritten as a function of $\mathbf{W}_{sim}$, it follows that:

$$\mathbf{W}_{sim}^* \approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{W}_{sim}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \tag{37}$$

Substituting by the corresponding eigendecomposition, this yields:

$$\begin{aligned} \mathbf{W}_{sim}^* &\approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{W}_{sim}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \\ \Leftrightarrow \mathbf{V}_{sim}^*\mathbf{\Lambda}^*\left(\mathbf{V}_{sim}^*\right)^\intercal &\approx \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{V}_{sim}\mathbf{\Lambda}\mathbf{V}_{sim}^{-1}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \end{aligned} \tag{38}$$

Observe that $\mathbf{\Lambda}^*$, the eigenvalue matrix of $\mathbf{W}_{sim}^*$, is not equal to $\mathbf{\Lambda}$, the eigenvalue matrix of $\mathbf{W}_{sim}^*$. However, both have the same spectral radius, $\rho\left(\mathbf{W}_{sim}^*\right) = \rho\left(\mathbf{W}_{sim}\right) = 1$. This implies that $\lim_{p\to\infty}\left(\mathbf{\Lambda}^*\right)^p = \lim_{p\to\infty}\mathbf{\Lambda}^p = \operatorname{diag}\left(1, 0, \ldots, 0\right)$. Therefore, as $p \to \infty$, the eigenvectors are directly related. Concretely, $\mathbf{V}_{sim}^* = \mathbf{D}_{R^*}^{-1/2}\mathbf{D}_R^{1/2}\mathbf{V}_{sim}$ and $\left(\mathbf{V}_{sim}^*\right)^\intercal = \mathbf{V}_{sim}^{-1}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2}$. But $\left(\mathbf{V}_{sim}^*\right)^\intercal = \left(\mathbf{V}^*\right)^\intercal\mathbf{D}_{R^*}^{1/2}$, as previously noted. Hence, equating yields:

$$\left(\mathbf{V}^*\right)^\intercal\mathbf{D}_{R^*}^{1/2} = \mathbf{V}_{sim}^{-1}\mathbf{D}_R^{1/2}\mathbf{D}_{R^*}^{-1/2} \Leftrightarrow \mathbf{V}_{sim}^{-1} = \left(\mathbf{V}^*\right)^\intercal\mathbf{D}_{R^*}\mathbf{D}_R^{-1/2} \tag{39}$$

At the "long run", $\mathbf{W}_{sim}^{\infty}$ can be approximated by:

$$\begin{aligned}
\mathbf{W}_{sim}^{\infty} &\approx \frac{1}{\left\|\mathbf{D}_R^{1/2}\mathrm{col}\,(\mathbf{V})_1\right\|_2 \left\|\mathrm{col}\,(\mathbf{V}^*)_1^{\intercal}\,\mathbf{D}_{R^*}^{1/2}\right\|_2} \times \mathbf{D}_R^{1/2}\mathrm{col}\,(\mathbf{V})_1\,\mathrm{col}\,(\mathbf{V}^*)_1^{\intercal}\,\mathbf{D}_{R^*}\mathbf{D}_R^{-1/2} \\
&= \left[\sum_{i=1}^{N} d_{R,i}\right]^{-1/2} \left[\sum_{i=1}^{N} d_{R,i}^*\right]^{-1/2} \mathbf{D}_R^{1/2}\boldsymbol{\iota}\boldsymbol{\iota}^{\intercal}\mathbf{D}_{R^*}\mathbf{D}_R^{-1/2} \qquad (40) \\
&= \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,(i,j)}\right]^{-1/2} \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,(i,j)}^*\right]^{-1/2} \mathbf{D}_R^{1/2}\mathbf{J}\mathbf{D}_{R^*}\mathbf{D}_R^{-1/2}
\end{aligned}$$

Note that the eigenvector $\mathbf{V}^*$ is associated with the row stochastic matrix $\mathbf{W}^*$ and the result of theorem 1.2 holds. In addition, the matrix is rescaled by the geometric mean of all elements of $\mathbf{W}_0$ and $\mathbf{W}_0^*$ because $\mathbf{V}_{sim}$ has to be unitary (*resp.* orthogonal) and $\mathbf{V}_{sim}^{-1}$ is a function of $(\mathbf{V}_{sim}^*)^{\intercal}$. Now, to obtain $\mathbf{W}^{\infty}$ rewrite $\mathbf{W}$ as $\mathbf{D}_R^{-1/2}\mathbf{W}_{sim}\mathbf{D}_R^{1/2}$. This yields:

$$\mathbf{W}^{\infty} = \mathbf{D}_R^{-1/2}\mathbf{W}_{sim}^{\infty}\mathbf{D}_R^{1/2} = \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,(i,j)}\right]^{-1/2} \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,(i,j)}^*\right]^{-1/2} \mathbf{J}\mathbf{D}_{R^*} \qquad (41)$$

Again, the matrix product $\mathbf{J}\mathbf{D}_{R^*}$ can be simplified to a less demanding matrix-vector product. Furthermore, the stochastic property of $\mathbf{W}^{\infty}$ remains unchanged, relative to $\mathbf{W}$.

If the powers $p \geq 2$ of $\mathbf{W}$ are all substituted by $\mathbf{W}^{\infty}$, the spatial lag operator inverse is approximated by:

$$\begin{aligned}
(\mathbf{I} - \alpha\mathbf{W})^{-1} &= \mathbf{I} + \mathbf{W} + \alpha^2\mathbf{W}^2 + \alpha^3\mathbf{W}^3 + \ldots \\
&\approx \mathbf{I} + \alpha\mathbf{W} + \alpha^2\mathbf{W}^{\infty} + \alpha^3\mathbf{W}^{\infty} + \ldots \\
&= \mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty} \qquad (42) \\
&\approx \mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,(i,j)}\right]^{-1/2} \left[\sum_{i=1}^{N}\sum_{j=1}^{N} w_{0,(i,j)}^*\right]^{-1/2} \mathbf{J}\mathbf{D}_{R^*}
\end{aligned}$$

Note that the quality of this approximation is related to the number and magnitude of the (j,i) elements that do not coincide with the (i,j) elements of $\mathbf{W}_0$.

3.3 Estimation of a model with a spatially lagged binary dependent variable

In terms of estimation, the N2SLS iterative procedure is considered and the approximation of the spatial lag operator inverse replaces $(\mathbf{I}-\alpha\mathbf{W})^{-1}$ in the individual gradients for each parameter. This implies that the true probability, $\mathrm{P}(\mathbf{Y}=1\,|\,\mathbf{X},\mathbf{W})$, may be computed after the algorithm converges. The matrix inversions are replaced by sums of known quantities and by less computationally demanding operations, such as matrix-vector products and Hadamard products.

On this framework, the error $\mathbf{u}$ is known to be heteroskedastic. Therefore, the index of the link function, $G(\cdot)$, has to be standardized by the diagonal elements of the conditional variance of the error $\mathbf{u}$. Replacing by the approximation of the spatial lag operator inverse, it follows that:

$$\begin{aligned}\mathrm{Var}(\mathbf{u}\,|\,\mathbf{X},\mathbf{W}) &\approx \left(\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)\left(\mathbf{I}+\alpha\mathbf{W}+\frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)^{\intercal}\\ &= \mathbf{I}+\alpha\left(\mathbf{W}+\mathbf{W}^{\intercal}\right)+\frac{\alpha^2}{1-\alpha}\left[\mathbf{W}^{\infty}+(\mathbf{W}^{\infty})^{\intercal}\right]+\alpha^2\mathbf{W}\mathbf{W}^{\intercal}\\ &\quad+\frac{\alpha^3}{1-\alpha}\left[\mathbf{W}(\mathbf{W}^{\infty})^{\intercal}+\mathbf{W}^{\infty}\mathbf{W}^{\intercal}\right]+\left(\frac{\alpha^2}{1-\alpha}\right)^2\mathbf{W}^{\infty}(\mathbf{W}^{\infty})^{\intercal}=\boldsymbol{\Sigma}\end{aligned} \tag{43}$$

Still there are three quantities from the equation above that require additional calculations: (i) $\mathbf{W}\mathbf{W}^{\intercal}$, (ii) $\mathbf{W}(\mathbf{W}^{\infty})^{\intercal}$ and (iii) $\mathbf{W}^{\infty}(\mathbf{W}^{\infty})^{\intercal}$. The major concern about these quantities is due to the fact that they consist in matrix products of N dimensional matrices. For the first quantity, the only elements that can be obtained without much computational burden are the diagonal elements:

$$\mathbf{W}\mathbf{W}^{\intercal}\approx\mathrm{diag}\left(\sum_{j=1}^{N}\left(\mathbf{W}^{\circ 2}\right)_{i,j}\right)=\mathrm{diag}\left(\sum_{j=1}^{N}w_{1,j}^2,\sum_{j=1}^{N}w_{2,j}^2,\ldots,\sum_{j=1}^{N}w_{N,j}^2\right) \tag{44}$$

where $\mathbf{W}^{\circ 2}$ is the Hadamard square of $\mathbf{W}$. Here, the off diagonal elements are zero, to avoid additional calculations that rapidly increase the computational complexity of the operation. As a consequence, the covariances of $\mathbf{u}$ are underspecified. But most importantly, the variances are properly specified.

For the second quantity, consider that $\mathbf{W}$ is row stochastic (*resp.* column stochastic or doubly stochastic). Recall that the stochastic property of $\mathbf{W}^{\infty}$ is always equal to the stochastic property of $\mathbf{W}$. Then, in general, $\mathbf{W}^{\infty} = (1/c^{*})\,\mathbf{J}\mathbf{D}_{R^{*}}$, has equal row vectors. Each element of the row vector of $\mathbf{W}^{\infty}$ is given by the row sum of $\mathbf{W}_0^{*}$ divided by the geometric mean of the sum of all elements of $\mathbf{W}_0$ and $\mathbf{W}_0^{*}$. Therefore, the matrix product $\mathbf{W}\,(\mathbf{W}^{\infty})^{\intercal}$ can be simplified to a matrix-vector product $\mathbf{W} \times \operatorname{row}(\mathbf{W}^{\infty})_1$. The result is the first column vector of $\mathbf{W}\,(\mathbf{W}^{\infty})^{\intercal}$. Therefore, the result of $\mathbf{W}^{\infty}\,(\mathbf{W}^{\infty})^{\intercal}$ is properly specified, because the remaining column vectors can be obtained by expanding the matrix-vector product $N-1$ times.

For the third quantity, since $\mathbf{W}^{\infty}$ has equal row vectors, the matrix product $\mathbf{W}^{\infty}\,(\mathbf{W}^{\infty})^{\intercal}$ can be simplified to a matrix-vector product $\mathbf{W}^{\infty} \times \operatorname{row}(\mathbf{W}^{\infty})_1$. The result is the first column vector of $\mathbf{W}^{\infty}\,(\mathbf{W}^{\infty})^{\intercal}$. A particular feature of this column vector is that all of its elements are equal – the rows of $\mathbf{W}^{\infty}$ are equal to the columns of $(\mathbf{W}^{\infty})^{\intercal}$ –. In addition, the diagonal elements of $\mathbf{W}^{\infty}\,(\mathbf{W}^{\infty})^{\intercal}$ are given by the Hadamard square of $\mathbf{W}^{\infty}$. Thus, only required to compute the Hadamard square of $\mathbf{W}^{\infty}$ and obtain the sum for the first row. Therefore, the result of $\mathbf{W}^{\infty}\,(\mathbf{W}^{\infty})^{\intercal}$ is properly specified if a $N \times N$ matrix is filled with $\sum_{j=1}^{N}\left(w_{1,j}^{\infty}\right)^2$.

Using the previous results, the diagonal elements of $\operatorname{Var}(\mathbf{u}\,|\,\mathbf{X},\mathbf{W})$ are now given by:

$$\sigma_i^2 \approx 1 + \frac{2\alpha^2}{1-\alpha} w_{i,i}^{\infty} + \alpha^2 \sum_{j=1}^{N} w_{i,j}^2 + \frac{2\alpha^3}{1-\alpha} \sum_{j=1}^{N} w_{i,j} w_{1,j}^{\infty} + \left(\frac{\alpha^2}{1-\alpha}\right)^2 \sum_{j=1}^{N} \left(w_{1,j}^{\infty}\right)^2 \qquad (45)$$

For a general function $G(\cdot)$, the individual gradients for each parameter are now given by:

$$(\mathbf{\Gamma}_{\boldsymbol{\beta}})_i = \frac{\partial u_{*,i}}{\partial \boldsymbol{\beta}^{\intercal}} = -u_{*,i}\left(\frac{\nabla g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right)}{g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right)} - u_{*,i}\right)\mathbf{X}_i^{\#\#}, \qquad i = 1, 2, \ldots, N \qquad (46)$$

and

$$(\boldsymbol{\Gamma}_{\alpha})_i = \frac{\partial u_{*,i}}{\partial \alpha} = -u_{*,i}\left(\frac{\nabla g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right)}{g\left(\mathbf{X}_i^{\#\#}\boldsymbol{\beta}\right)} - u_{*,i}\right)\left[\mathbf{H}_i\boldsymbol{\beta} - \frac{\mathbf{X}_i^{\#\#}\boldsymbol{\beta}}{2\sigma_i^2}\Upsilon_{ii}\right], \qquad i = 1, 2, \ldots, N \tag{47}$$

where the indexes of both functions $g(\cdot)$ and $G(\cdot)$ from the generalized residuals, $\mathbf{u}_*$, are $\mathbf{X}^{\#\#}\boldsymbol{\beta}$, with $\mathbf{X}^{\#\#} = \left(\mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)\mathbf{X}^{\#}$ and the scaled regressors, $\mathbf{X}_i^{\#}$, consider the variance in equation (45). The matrix $\mathbf{H}$ requires an approximation:

$$\mathbf{H} \approx \left(\mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\right)\mathbf{W}\mathbf{X}^{\#\#} = \left(\mathbf{W} + \alpha\mathbf{W}^2 + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}\mathbf{W}\right)\mathbf{X}^{\#\#} = \mathbf{F}\mathbf{X}^{\#\#} \tag{48}$$

In this case, the quantities that require additional calculations are: (i) $\mathbf{W}^2$ and (ii) $\mathbf{W}^{\infty}\mathbf{W}$. For the first quantity, similarly to $\mathbf{W}\mathbf{W}^{\intercal}$, the only elements that can be obtained without much computational burden are the diagonal elements:

$$\mathbf{W}^2 \approx \operatorname{diag}\left(\sum_{j=1}^{N}\left(\mathbf{W}\circ\mathbf{W}^{\intercal}\right)_{i,j}\right) = \operatorname{diag}\left(\sum_{j=1}^{N} w_{1,j}w_{j,1}, \sum_{j=1}^{N} w_{2,j}w_{j,2}, \ldots, \sum_{j=1}^{N} w_{N,j}w_{j,N}\right) \tag{49}$$

where "$\circ$" is the Hadamard product. Likewise, the off diagonal elements are underspecified, whereas the diagonal elements are properly specified.

For the second quantity, consider the eigendecomposition for these matrices, this yields:

$$\mathbf{W}^{\infty}\mathbf{W} = \mathbf{V}\boldsymbol{\Lambda}^{\infty}\mathbf{V}^{-1}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} = \mathbf{V}\boldsymbol{\Lambda}^{\infty}\mathbf{V}^{-1} = \mathbf{W}^{\infty} \tag{50}$$

Considering the previous results, $\mathbf{F} \approx \mathbf{W} + \alpha\operatorname{diag}\left(\sum_{j=1}^{N}\left(\mathbf{W}\circ\mathbf{W}^{\intercal}\right)_{i,j}\right) + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty}$. In addition, the matrix $\boldsymbol{\Upsilon}$ also requires an approximation. But only the diagonal elements need to be specified. This yields:

$$\begin{aligned}
\operatorname{diag}(\boldsymbol{\Upsilon}) &= 2 \times \operatorname{diag}\left(\left(\mathbf{I}_N - \alpha \mathbf{W}_N\right)^{-1} \mathbf{W}_N \left(\mathbf{I}_N - \alpha \mathbf{W}_N\right)^{-1} \left[\left(\mathbf{I}_N - \alpha \mathbf{W}_N\right)^{-1}\right]^{\intercal} \right) \\
&\approx 2 \times \operatorname{diag}\left(\mathbf{F}\boldsymbol{\Sigma}\right) \\
&= 2 \times \operatorname{diag}\left(\sum_{j=1}^{N} \left(\mathbf{F} \circ \boldsymbol{\Sigma}^{\intercal}\right)_{i,j} \right) \\
&= 2 \times \operatorname{diag}\left(\sum_{j=1}^{N} f_{1,j}\sigma_{j,1}, \sum_{j=1}^{N} f_{2,j}\sigma_{j,2}, \ldots, \sum_{j=1}^{N} f_{1,N}\sigma_{N,1} \right)
\end{aligned} \tag{51}$$

It is obvious that the term $\left(\mathbf{I}_N - \alpha \mathbf{W}_N\right)^{-1} \mathbf{W}_N$ is approximated by $\mathbf{F}$ and the term $\left(\mathbf{I}_N - \alpha \mathbf{W}_N\right)^{-1} \left[\left(\mathbf{I}_N - \alpha \mathbf{W}_N\right)^{-1}\right]^{\intercal}$ is approximated by $\boldsymbol{\Sigma}$. The diagonal elements can be obtained without much computational burden by the Hadamard product of the two approximation matrices.

Next, a Monte Carlo set of experiments is implemented to verify the usefulness of the methods presented in this section.

4 Simulation study

In the previous sections a method to approximate the spatial lag operator inverse was suggested and implemented to estimate a model with a spatially lagged binary dependent variable. In this section, the properties of the inverse approximation and the suggested estimator are verified to demonstrate the feasibility and consistency of both methods.

The simulation study is divided in twofold. Firstly, the quality of the inverse approximation is analyzed. Additional examples are used to discuss the robustness of the approximation in the case where some assumptions do not hold. Secondly, the small and large sample properties of the N2SLS estimator with approximated gradients, the N2SLS estimator and the Linearized GMM are assessed and compared. Consistency, efficiency and execution time are central issues.

The model considered in the simulations follows as:

$$\mathbf{Y} = \Phi\left((\mathbf{I} - \alpha\mathbf{W})^{-1}\left(\beta_0\boldsymbol{\iota}^{\#} + \beta_1\mathbf{X}^{\#}\right)\right) + (\mathbf{I} - \alpha\mathbf{W})^{-1}\boldsymbol{\varepsilon} \tag{52}$$

where $\Phi(\cdot)$ is the CDF of the Normal distribution. The unscaled explanatory variable, $\mathbf{X}$, is randomly drawn from $\mathcal{U}(-3, 3)$ and the error, ε_i, is randomly drawn from $\mathcal{N}(0, 1)$, for all i. Following McMillen (1995), the variance of $\mathbf{X}$ is chosen to be much larger than the variance of $\boldsymbol{\varepsilon}$ because the estimated model tends to produce better predictions for the dependent variable.

Additionally, four scenarios for the (normalized) spatial weights matrix are considered:

(1) $\mathbf{W}$ is a symmetric inverse Euclidean distance matrix, with a distance threshold:

$$\mathbf{W} = \begin{cases} w_{i,j} = 1/d_{i,j} & \text{if } d_{i,j} \leq d^* \\ w_{i,j} = 0 & \text{if } d_{i,j} > d^* \end{cases} \tag{53}$$

(2) $\mathbf{W}$ is a non-symmetric inverse Euclidean distance matrix, with a distance threshold

(3) $\mathbf{W}$ is a symmetric binary distance based matrix, with a distance threshold:

$$\mathbf{W} = \begin{cases} w_{i,j} = 1 & \text{if } d_{i,j} \leq d^* \\ w_{i,j} = 0 & \text{if } d_{i,j} > d^* \end{cases} \tag{54}$$

(4) $\mathbf{W}$ is a non-symmetric binary distance based matrix, with a distance threshold.

Using these definitions, the scenarios for non-binary and binary spatial weights matrices are comparable. The non-symmetric matrices are obtained by restricting the element $(1, 2)$ of the symmetric matrices.

The definition of the sample size follows the work of Lee (2004). The spatial weights matrix is defined such that there are R regions with, at most, m units. The sample size is given by $N = R \times m$. This structure allows to define $\widetilde{\mathbf{W}} = \mathbf{I}_R \otimes \mathbf{W}_m$, a block diagonal matrix with R blocks and each block has, at most, m neighbors. Hence, all operations with $\widetilde{\mathbf{W}}$ are simplified to Kronecker products involving $\mathbf{W}_m$ instead. Taking this approach, one is able to verify the asymptotics of the estimators based on spatial statistics frameworks.

Here the asymptotics can be approached in two ways: increasing-domain asymptotics and infill asymptotics (Cressie, 1993). The former corresponds to a scenario where the samples are collected from different regions, hence $R \to \infty$ and $N \to \infty$ (Lee, 2004). The latter corresponds to a scenario where new observations are added between the existing ones and a bounded area tends to get denser, as $N \to \infty$ (Anselin, 2001).

The following experiments consider values of $R \in \{10, 50, 100\}$ and $m \in \{10, 50, 100\}$, for the four scenarios of the spatial weights matrix. The distance threshold is fixed at $d^* = 30$, to avoid the absence of neighbors, for a given unit i. The samples are generated, for a fixed $\widetilde{\mathbf{W}}$, considering $\alpha = 0.5$, $\beta_0 = 0$ and $\beta_1 = 1$. The matrix of instruments for all estimations is $\mathbf{Z} = \left[\mathbf{X}\ \widetilde{\mathbf{W}}\mathbf{X}\ \widetilde{\mathbf{W}}^{\infty}\mathbf{X}\right]$. For each scenario and sample designs, 500 replications were used. The simulations were performed in a computer working in a 64-bit operating system, quad-core Intel based processor with 3.2 GHz and 8 gigabites of RAM.

In the sections that follow, the norm of the approximation to the true inverse is reported, considering the four spatial weights matrices scenarios. Next, the estimates of α, β_0 and β_1 are reported, for each of the three estimators (the N2SLS estimator with approximated gradients, the N2SLS estimator and the Linearized GMM). The estimates of se (α), se (β_0) and se (β_1) are reported, for each of the three estimators, as well. The results are summarized by the mean and mean squared error. To compute the MSE for the empirical standard errors, it is assumed that the asymptotic standard errors are given by the N2SLS estimator for $m = 50$ and $R = 100$ (simulations for $m = 100$ and $R = 100$ were not concluded in time). Execution times and the number of iterations, are also reported and summarized by the mean.

4.1 Inverse approximation

To evaluate the quality of the inverse approximation, is it useful to establish the following identity:

$$(\mathbf{I} - \alpha\mathbf{W})^{-1}(\mathbf{I} - \alpha\mathbf{W}) = \mathbf{I} \tag{55}$$

For $|\alpha| < 1$, this matrix product is available, because $(\mathbf{I} - \alpha\mathbf{W})$ is non-singular. If $\mathbf{I}$ is subtracted from both sides of the equation, $(\mathbf{I} - \alpha\mathbf{W})^{-1}(\mathbf{I} - \alpha\mathbf{W}) - \mathbf{I} = \mathbf{0}$, where $\mathbf{0}$ is a matrix of zeros. Applying a norm, the previous difference is zero. Substituting the spatial lag operator inverse by the corresponding approximation yields:

$$\left\| \left(\mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty} \right) (\mathbf{I} - \alpha\mathbf{W}) - \mathbf{I} \right\| \tag{56}$$

If this norm is close to zero, the approximation properly identifies the elements of the inverse.

This simulation study focuses on $\mathbf{W}$ instead of $\widetilde{\mathbf{W}}$, because the inverse of a block diagonal matrix is the inverse of each block. Here, for simplicity, each of the blocks are assumed to be equal.

In table 1 the results of the norm, in equation (56), are presented, considering the four scenarios of the normalized spatial weights matrix, $\mathbf{W}$, and $\alpha = 0.5$. The size of these matrices is determined by m, the maximum number of units in each block, R. Exceptionally m is ranging from 10 to 1000, because the dimension R is dropped.

In a nutshell, on average, the quality of the approximation slightly deteriorates as $m \to \infty$. The approximation appears to be better for binary matrices and the effect of non-symmetry is minimal, if only few (j, i) elements are not equal to the corresponding (i, j) elements.

Taking in consideration the results of table 2, regarding the computational time required to calculate the inverse matrix, the previous results are remarkable. When the matrix size increases from 500 to 1000, the computational time increases, on average, by 10 times. However, the quality of the approximation deteriorates by a far less amount, depending on the nature of the matrix. There is a striking evidence towards the importance of the inverse approximation for large sized matrices.

Still these results are based on the assumption that the largest absolute eigenvalue has algebraic multiplicity equal to one. In practice, this assumption may be quite restrictive. Especially for binary spatial weights matrices, there is the possibility that the algebraic multiplicity is, indeed, larger than one and the matrix is not block diagonal, or the smallest

eigenvalue is -1. As a consequence, the first eigenvector of $\mathbf{W}$ is no longer a vector of ones[2].

Consider the following adapted examples of normalized binary spatial weights matrices, from (Arbia, 2014, p. 29) and (Anselin, 1988, p. 19):

Arbia (2014)

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Anselin (1988)

$$\mathbf{W} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

For both matrices, the largest eigenvalue is 1 and the smallest eigenvalue is -1. Then, for each, there two largest absolute eigenvalues equal to one and the first eigenvector is no longer a vector of ones. Under this eigenstructure, the quality of the inverse approximation for such matrices is addressed. For $\alpha = 0.5$, the true inverses of the spatial lag operators are:

Arbia (2014)

$$(\mathbf{I} - \alpha\mathbf{W})^{-1} = \begin{bmatrix} 1 & 0.67 & 0 & 0.33 \\ 0 & 1.33 & 0 & 0.67 \\ 0 & 0.33 & 1 & 0.67 \\ 0 & 0.67 & 0 & 1.33 \end{bmatrix}$$

Anselin (1988)

$$(\mathbf{I} - \alpha\mathbf{W})^{-1} = \begin{bmatrix} 1.24 & 0.36 & 0.09 & 0.31 \\ 0.36 & 1.24 & 0.31 & 0.09 \\ 0.18 & 0.62 & 1.16 & 0.04 \\ 0.62 & 0.18 & 0.04 & 1.16 \end{bmatrix}$$

And the approximations are given by:

Arbia (2014)

$$\mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty} = \begin{bmatrix} 1.10 & 0.70 & 0.10 & 0.20 \\ 0.10 & 1.20 & 0.10 & 0.70 \\ 0.10 & 0.20 & 1.10 & 0.70 \\ 0.10 & 0.70 & 0.10 & 1.20 \end{bmatrix}$$

Anselin (1988)

$$\mathbf{I} + \alpha\mathbf{W} + \frac{\alpha^2}{1-\alpha}\mathbf{W}^{\infty} = \begin{bmatrix} 1.17 & 0.42 & 0.08 & 0.33 \\ 0.42 & 1.17 & 0.33 & 0.08 \\ 0.17 & 0.67 & 1.08 & 0.08 \\ 0.67 & 0.17 & 0.08 & 1.08 \end{bmatrix}$$

The norms are, respectively, $\|\cdot\| = 0.356$ and $\|\cdot\| = 0.258$. Considering that, before normalization, the matrices were binary, the quality of the approximation deteriorates at least 2 times, compared to the average norm of approximation for non-symmetric binary spatial weights matrix (see table 1). If the coefficient on $\mathbf{W}^{\infty}$ is replaced by $\alpha^2/(1-\alpha^2)$, the quality of the approximation may improve, because only the even powers of $\mathbf{W}$ are substituted by $\mathbf{W}^{\infty}$, eliminating the negative eigenvalue issue.

[2]The first eigenvector may have elements equal to 1, -1 or 0.

4.2 Estimation

Table 3 summarizes the results of the Monte Carlo simulation study, for the spatial lag Probit, with a symmetric inverse distance matrix.

The estimates of α, $\hat{\alpha}$, are slightly biased, in all three estimators (Nonlinear Two Stages Least Squares – N2SLS –, N2SLS with approximated gradients – N2SLSa – and the Linearized GMM – LinGMM –). For a fixed m, the biases of $\hat{\alpha}$ decrease, as R becomes larger. The same do not apply for a fixed R and increasing m. Therefore, under increasing-domain asymptotics consistency can be achieved, whereas under infill asymptotics the estimators may not be consistent. This in line with the conclusions of Lahiri (1996) and Lee (2004). The same argument is valid for both $\hat{\beta}_0$ and $\hat{\beta}_1$. Though the Linearized GMM estimates for β_1 are severely biased.

The empirical standard errors steadily decrease as both R and m increase. The same applied to the MSE of the estimates and to the MSE of the empirical standard errors, except for $\text{se}(\hat{\alpha})$. In fact, the MSE of $\text{se}(\hat{\alpha})$ increase as both R and m increase.

In terms of the computational time, it is obvious that the Linearized GMM outperforms the two remaining estimators. For the N2SLS estimator, it exponentially increases as R and m increase. The same is valid to the N2SLSa estimator, but at a slower rate. This is related to the fact that the N2SLSa struggles to converge, by two iterations, on average, when compared to the N2SLS estimator. These results appear to be a disappointment. Nevertheless, the progressive integration of C++ codes and high performance linear algebra routines in statistical and mathematical software, already allows to manage large amounts of information. Still, it is restricted to simple matrix operations, such as the sum or the product. So, the computational time of the N2SLSa estimator can be drastically improved, whereas the effect on the N2SLS estimator is still unknown. To guarantee comparability, this approach is not yet considered.

5 Empirical application: Competitiveness in U.S. Metropolitan Areas

In this section, a short empirical application about competitiveness in the United States Continental Metropolitan Statistical Areas (MSAs) is presented. The previous estimators are used and their adaptability to empirical data is assessed.

Competitiveness is currently one of the main concerns for policy makers. It is a measure of economic performance, usually related to the GDP growth or the unit labor costs (the cost of labour per units of output). Porter et al. (2016), in an extensive survey about U.S. competitiveness, highlight the progressive deterioration of the whole economy, especially in the 2000s. This is related to the consecutive economic crisis and the accumulation of trade deficits.

Additionally, the environmental issues in the U.S. are a controversial subject since the mid-2000s. In the Economics theory, these environmental issues are addressed by the Environmental Kuznets Curve (EKC) hypothesis (Grossman and Krueger, 1991; Shafik and Bandyopadhyay, 1992; Panayotou, 1993). The EKC hypothesis states that there is an inverted "U" shaped relationship between environmental degradation and economic growth. Similarly, this led to controversy especially in terms of the shape of the curve (de Bruyn, van den Bergh, and Opschoor, 1998).

Besides the theoretical discussions, recognizing that there is a progressive decline in the environmental quality, contributed to the definition of strategic plans that focus on both economic and ecological growth. As a result, the EKC hypothesis can be reversed: competitiveness may be affected by environmental quality. However, empirically, this relationship is not yet tested.

The literature focus on the analysis of competitiveness and environmental quality as separate subjects. In addition, only few works consider a regional level analysis, namely Dudensing and Barkley (2010) and Rice, Venables, and Patacchini (2006), addressing regional competitiveness in a spatial framework (U.S. Southern Metropolitan Areas and British regions, respectively), and Rupasingha, Goetz, Debertin, and Pagoulatos (2004) and Millimet, List, and Stengos (2003), for a spatial analysis on the EKC hypothesis at

regional level (U.S. counties and U.S. states, respectively). Still nothing is done for spatial discrete variables.

This study consists in the analysis of a binary regional competitiveness indicator in terms of environmental quality variables, for 2009. The regional units are the U.S. Continental MSAs with more than 350 thousand inhabitants in the 2000 Population Census ($N = 129$). The year 2009 is of particular interest, as it is the first year that the U.S. economy exhibited strong recovery signs, after the economic crisis of the mid-2000s. The dataset is a combination of socioeconomic data and environmental data for the selected MSAs. The sources are respectively, the U.S. Bureau of Economic Analysis (BEA) and the U.S. Environmental Protection Agency (EPA). Considering the available data, the binary competitiveness indicator is given by:

$$Y_i = \begin{cases} 1, & \text{if } \left(\dfrac{\text{LC}_i}{\text{GDP}_i} - \dfrac{\text{LC}_{\text{USA}}}{\text{GDP}_{\text{USA}}}\right) \geq 0 \\ 0, & \text{if } \left(\dfrac{\text{LC}_i}{\text{GDP}_i} - \dfrac{\text{LC}_{\text{USA}}}{\text{GDP}_{\text{USA}}}\right) < 0 \end{cases}, \qquad i = 1, \ldots, 129 \tag{57}$$

where LC is the Total Labor Compensation and GDP is the Total Gross Domestic Product. The binary variable, Y_i, indicates whether a given MSA i is more competitive than the whole country. Figure 1 shows the spatial distribution of the competitiveness indicator for each MSA. The darker areas correspond to competitive areas. There are at least two clusters of competitive MSAs (with respect to the whole country) in the same state – California and Florida – and two multi-state clusters of competitive MSAs – Virginia, Maryland, Pennsylvania and New Jersey; Connecticut, Massachusetts and New Hampshire –. As these are MSAs with typically high GDP, the productivity of an additional dollar spent in labor is much higher than in the whole country.

Table 7 presents the descriptive statistics for the variables used on the analysis. The reported AQI is the annualized Air Quality Index (divided by 100). It is calculated, on a daily basis, for each of five major pollutants: (1) ground-level ozone (O_3), (2) particule pollution ($PM_{2.5}$ and PM_{10}), (3) carbon monoxide (CO), (4) sulfur dioxide (SO_2) and (5) nitrogen dioxide (NO_2). The daily AQI corresponds to the maximum daily index for each pollutant, depending on the observed concentration levels and admissible values

for that particular day. The AQI_{min} and AQI_{max} are, respectively, the minimum and maximum annualized AQI for a given MSA i, in 2009.

The variables % days O_3, % days $PM_{2.5}$ and % days NO_2 are the percentage of days that each of the pollutants were responsible for the overall value of the daily AQI. Their averages sum up to approximately 0.93. Therefore, they are the pollutants that contribute the most for the overall value of the daily AQI on a given MSA i, in 2009.

To study the effects of environmental indicators over the binary competitiveness in the selected MSAs, a spatial lag Probit is estimated. Recall that, on this framework $G(\cdot) = \Phi(\cdot)$, the CDF of the Normal distribution. The spatial weights matrix $\mathbf{W}$ is assumed to be a binary radial matrix, with a distance threshold, $d^* = \bar{d} - 0.5\,\text{sd}(d) \approx 6$, where $\bar{d}$ is the average Euclidean distance between MSAs and $\text{sd}(d)$ is the corresponding standard deviation. For all the estimators, the matrix of instruments is $\mathbf{Z} = \left[\mathbf{X}\ \widetilde{\mathbf{W}}\mathbf{X}\ \widetilde{\mathbf{W}}^{\infty}\mathbf{X}\right]$.

The estimation results are shown in table 8. All of the three estimators produce coherent estimates, in terms of sign and significance. Exception made to the variable AQI_{min}, not significant in the Linearized GMM estimation. However, in terms of magnitude, some estimates differ drastically. Not surprisingly, this is the case of the spatial lag parameter, ranging from $\hat{\alpha} = 0.524$ to $\hat{\alpha} = 0.799$. This is a consequence of the approximations made in both N2SLS with approximated gradients and Linearized GMM, but also due to the small sample size.

In addition, the Hansen test is presented. The null of correct over-identifying restrictions is only rejected for the N2SLS estimator. Again, these results are debatable, due to small sample distortions.

Still, striking conclusions can be taken from these results. The partial effects cannot be directly addressed, as the estimated model is nonlinear. Nevertheless, as the spatial lag operator inverse is a sum of non-negative matrices, the estimated signs correspond to the estimated signs of the partial effects in the nonlinear model.

Even though the estimates of the spatial parameter are far from stable, one may argue that there is a medium-high degree of spatial dependence between the selected MSAs, in 2009. This means that, if the neighbors are competitive, then, on average, the probability

that the MSA i is competitive (with respect to whole country) increases. Also, an environmental degradation implies that, on average, there is a reduction on the probability that the MSA i is competitive.

Furthermore, the quality of the predictions can be used as a measure for the overall adjustment of the previous estimators. Wooldridge (2010) defines the predictions under a binary dependent variable framework as it follows:

$$\hat{Y}_i = \begin{cases} 1, \text{ if } \Phi\left(\hat{\mathbf{X}}_i^{\#\#}\hat{\boldsymbol{\beta}}\right) \geq \bar{Y} \\ 0, \text{ if } \Phi\left(\hat{\mathbf{X}}_i^{\#\#}\hat{\boldsymbol{\beta}}\right) < \bar{Y} \end{cases} \tag{58}$$

where $\hat{\mathbf{X}}_i^{\#\#}$ is a function of the estimated spatial lag parameter, $\hat{\alpha}$. The indicator that measures the quality of the predictions is given by:

$$P = \frac{N_{00} + N_{11}}{N} \tag{59}$$

where N_{00} and N_{11} are, respectively, the number of zeroes and ones well predicted. Then for each estimator we have, $P_{N2SLSa} = 0.6828$, $P_{N2SLS} = 0.6828$ and $P_{LinGMM} = 0.5659$. These are outstanding results. Even though the estimates of the two N2SLS estimators are different, they produce exactly the same predictions. Therefore, the approximation of the spatial lag operator inverse do not seem to penalize the ability of the estimator, with regard to prediction. The same is likely to happen to the estimates of the partial effects.

6 Conclusion

The idea of this paper is to simplify the estimation procedures of nonlinear models in spatial frameworks. Taking into account the eigenstructure of the spatial weights matrix, it is possible to reduce the problem of inverting the spatial lag operator to a simpler operation involving known matrix sums. At the same time, two problems are solved. First, the elements of the spatial lag operator inverse now have an approximated close formula. Second, the computational time required to invert the spatial lag operator can be drastically reduced.

Even though the quality of the approximation may not the best, when applied to the estimation of a nonlinear model, it produces reliable estimates. The proposed estimator, with approximated gradients (N2SLSa), is a feasible and consistent alternative to the N2SLS estimator (a particular case of the full GMM estimator), specially for large samples. Moreover, the N2SLSa estimator and the N2SLS estimator produce exactly the same predictions. The difference is, once again, computational time.

The empirical application allowed to extend the literature in twofold. First, the positive spillovers of the probability of being competitive. This implies that the MSAs do not appear to compete, instead they cooperate, allowing to create regional clusters and promote integration between industries. As this integration strengthens, the U.S. economy becomes less exposed to exogenous industry-specific shocks. Second, the negative correlation between environmental degradation and the probability of being competitive. The idea that pollution implies more output is misleading, if economic performance is at cause. As environmental degradation is related to health problems, the value of a worker tends to reduce. However, the cost of the same worker may probably increase due to health expenditures. Hence, it is less probable that a given MSA i is competitive, with respect to the whole country.

Furthermore, improving the computational time of the N2SLSa estimator without oversimplifying the model is a true possibility, as C++ codes and high performance linear algebra routines can be integrated to several statistical software. The ultimate goal would the ability of using big data in spatial frameworks.

Finally, the proposed approximation can be applied to linear models, as well. In particular, it can be used to approximate the log-determinant of the likelihood function. In fact, due to its simplicity, this approximation can be quickly introduced in the estimation routine of any researcher.

7 Appendix

7.1 Appendix 1

TABLE 1: Quality of the approximation to the true spatial lag operator inverse

		(1)	(2)	(3)	(4)
m	10	0.211	0.378	0.101	0.103
	50	0.318	0.317	0.104	0.104
	100	0.334	0.334	0.106	0.106
	500	0.342	0.342	0.107	0.107
	1000	0.350	0.350	0.106	0.106

TABLE 2: Computational time (in seconds) of the spatial lag operator inverse

		(1)	(2)	(3)	(4)
m	10	0.006	0.005	0.006	0.005
	50	0.005	0.007	0.006	0.007
	100	0.011	0.015	0.014	0.015
	500	0.439	0.475	0.447	0.468
	1000	4.337	4.416	4.339	4.437

TABLE 3: Simulation results for symmetric inverse distance **W**

m		10						50						100					
R		N2SLSa		N2SLS		LinGMM		N2SLSa		N2SLS		LinGMM		N2SLSa		N2SLS		LinGMM	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
10	$\hat{\alpha}$	0.470	0.020	0.485	0.020	0.451	0.036	0.514	0.011	0.493	0.010	0.492	0.021	0.498	0.009	0.481	0.010	0.457	0.021
	se $(\hat{\alpha})$	0.129	0.139	0.116	0.149	0.148	0.124	0.098	0.161	0.087	0.170	0.117	0.146	0.100	0.160	0.091	0.167	0.106	0.154
	$\hat{\beta}_0$	−0.002	0.024	−0.003	0.024	0.001	0.028	0.003	0.004	0.003	0.004	−0.001	0.010	0.002	0.002	0.002	0.002	0.000	0.005
	se $\left(\hat{\beta}_0\right)$	0.118	0.012	0.120	0.012	0.208	0.040	0.043	0.001	0.046	0.001	0.093	0.006	0.031	0.000	0.033	0.000	0.067	0.003
	$\hat{\beta}_1$	1.102	0.074	1.126	0.089	0.693	0.117	1.021	0.007	1.029	0.008	0.606	0.159	1.004	0.003	1.008	0.003	0.578	0.179
	se $\left(\hat{\beta}_1\right)$	0.203	0.038	0.221	0.064	0.224	0.046	0.079	0.003	0.080	0.003	0.097	0.005	0.054	0.001	0.055	0.001	0.068	0.002
	Time:																		
	Loop	0.083		0.125				0.265		0.416				0.850		7.494			
	#iter	9		6				9		6				8		6			
	Total	0.821		0.818		0.058		2.334		2.630		0.063		7.022		45.182		0.092	
50	$\hat{\alpha}$	0.484	0.003	0.494	0.003	0.449	0.008	0.515	0.002	0.497	0.002	0.487	0.004	0.513	0.002	0.494	0.002	0.461	0.005
	se $(\hat{\alpha})$	0.056	0.196	0.050	0.202	0.060	0.192	0.043	0.208	0.037	0.213	0.051	0.201	0.042	0.209	0.038	0.212	0.047	0.205
	$\hat{\beta}_0$	−0.002	0.004	−0.002	0.004	0.000	0.005	−0.001	0.001	−0.001	0.001	−0.001	0.002	−0.001	0.000	0.002	0.000	−0.001	0.001
	se $\left(\hat{\beta}_0\right)$	0.047	0.001	0.047	0.001	0.080	0.004	0.018	0.000	0.019	0.000	0.039	0.001	0.013	0.000	0.013	0.000	0.028	0.000
	$\hat{\beta}_1$	1.020	0.008	1.028	0.009	0.647	0.128	1.000	0.001	1.005	0.001	0.593	0.166	1.000	0.001	1.001	0.001	0.570	0.185
	se $\left(\hat{\beta}_1\right)$	0.079	0.003	0.083	0.004	0.086	0.004	0.034	0.000	0.035	0.000	0.042	0.000	0.024	0.000	0.024	0.000	0.030	0.000
	Time:																		
	Loop	0.187		0.363				3.137		4.887				15.460		54.041			
	#iter	9		6				8		6				8		6			
	Total	1.697		2.246		0.070		26.488		30.050		0.111		128.927		328.801		3.318	
100	$\hat{\alpha}$	0.489	0.001	0.500	0.002	0.451	0.005	0.514	0.001	0.499	0.001	0.488	0.002						
	se $(\hat{\alpha})$	0.039	0.211	0.035	0.215	0.042	0.209	0.031	0.000	0.026	0.224	0.036	0.000						
	$\hat{\beta}_0$	0.000	0.002	0.000	0.002	−0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.001						
	se $\left(\hat{\beta}_0\right)$	0.032	0.000	0.033	0.000	0.056	0.002	0.013	0.000	0.013	0.000	0.027	0.000						
	$\hat{\beta}_1$	1.005	0.004	1.012	0.004	0.645	0.128	0.993	0.001	0.999	0.001	0.593	0.166						
	se $\left(\hat{\beta}_1\right)$	0.055	0.001	0.057	0.001	0.059	0.001	0.024	0.000	0.024	0.000	0.029	0.000						
	Time:																		
	Loop	1.023		1.065				13.567		23.351									
	#iter	9		6				8		6									
	Total	8.907		6.477		0.058		113.067		143.274		1.990							

TABLE 4: Simulation results for non-symmetric inverse distance **W**

m		10						50						100					
R		N2SLSa		N2SLS		LinGMM		N2SLSa		N2SLS		LinGMM		N2SLSa		N2SLS		LinGMM	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
10	$\hat{\alpha}$	0.470	0.017	0.488	0.018	0.447	0.031	0.512	0.011	0.492	0.010	0.491	0.021	0.497	0.009	0.481	0.010	0.457	0.021
	se $(\hat{\alpha})$	0.123	0.143	0.115	0.149	0.140	0.130	0.099	0.161	0.088	0.170	0.117	0.146	0.100	0.160	0.091	0.167	0.106	0.154
	$\hat{\beta}_0$	−0.005	0.024	−0.003	0.025	0.002	0.024	0.003	0.004	0.003	0.004	−0.002	0.010	0.002	0.002	0.002	0.002	0.000	0.005
	se $\left(\hat{\beta}_0\right)$	0.118	0.012	0.121	0.013	0.205	0.038	0.044	0.001	0.046	0.001	0.093	0.006	0.031	0.000	0.033	0.000	0.067	0.003
	$\hat{\beta}_1$	1.114	0.080	1.140	0.100	0.694	0.117	1.021	0.007	1.028	0.008	0.606	0.159	1.004	0.003	1.008	0.003	0.578	0.179
	se $\left(\hat{\beta}_1\right)$	0.208	0.041	0.247	0.107	0.220	0.044	0.079	0.003	0.080	0.003	0.097	0.005	0.054	0.001	0.055	0.001	0.068	0.002
	Time:																		
	Loop	0.083		0.126				0.264		0.427				0.853		7.318			
	#iter	9		6				9		6				8		6			
	Total	0.795		0.834		0.054		2.327		2.700		0.063		7.069		44.154		0.084	
50	$\hat{\alpha}$	0.482	0.003	0.494	0.003	0.442	0.008	0.515	0.002	0.497	0.002	0.485	0.004	0.513	0.002	0.493	0.002	0.461	0.005
	se $(\hat{\alpha})$	0.053	0.199	0.050	0.202	0.058	0.195	0.043	0.208	0.037	0.213	0.051	0.201	0.042	0.209	0.039	0.212	0.047	0.205
	$\hat{\beta}_0$	−0.003	0.004	−0.003	0.004	0.001	0.004	−0.001	0.001	−0.001	0.001	−0.001	0.002	−0.001	0.000	0.002	0.000	−0.001	0.001
	se $\left(\hat{\beta}_0\right)$	0.047	0.001	0.047	0.001	0.079	0.004	0.018	0.000	0.019	0.000	0.039	0.001	0.013	0.000	0.013	0.000	0.028	0.000
	$\hat{\beta}_1$	1.015	0.008	1.025	0.009	0.645	0.129	1.000	0.001	1.005	0.001	0.593	0.166	1.000	0.001	1.001	0.001	0.570	0.185
	se $\left(\hat{\beta}_1\right)$	0.079	0.003	0.085	0.004	0.084	0.004	0.034	0.000	0.035	0.000	0.042	0.000	0.024	0.000	0.024	0.000	0.030	0.000
	Time:																		
	Loop	0.191		0.374				3.130		4.932				15.193		52.099			
	#iter	8		6				8		6				8		6			
	Total	1.670		2.300		0.071		26.448		30.347		0.108		126.202		317.337		3.343	
100	$\hat{\alpha}$	0.486	0.001	0.499	0.001	0.445	0.006	0.514	0.001	0.499	0.001	0.487	0.002						
	se $(\hat{\alpha})$	0.037	0.213	0.035	0.215	0.040	0.210	0.031	0.000	0.026	0.223	0.036	0.000						
	$\hat{\beta}_0$	−0.001	0.002	−0.001	0.002	−0.002	0.002	0.000	0.000	0.000	0.000	0.000	0.001						
	se $\left(\hat{\beta}_0\right)$	0.032	0.000	0.032	0.000	0.055	0.002	0.013	0.000	0.013	0.000	0.027	0.000						
	$\hat{\beta}_1$	1.003	0.004	1.011	0.004	0.646	0.127	0.993	0.001	0.999	0.001	0.592	0.166						
	se $\left(\hat{\beta}_1\right)$	0.055	0.001	0.059	0.001	0.058	0.001	0.024	0.000	0.024	0.000	0.029	0.000						
	Time:																		
	Loop	1.008		1.074				13.469		23.468									
	#iter	8		6				8		6									
	Total	8.540		6.544		0.057		112.594		143.803		2.031							

TABLE 5: Simulation results for symmetric binary **W**

m		10						50						100					
R		N2SLSa		N2SLS		LinGMM		N2SLSa		N2SLS		LinGMM		N2SLSa		N2SLS		LinGMM	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
10	$\hat{\alpha}$	0.470	0.020	0.483	0.021	0.458	0.040	0.493	0.045	0.472	0.021	0.575	0.060	0.483	0.055	0.468	0.026	0.571	0.068
	se $(\hat{\alpha})$	0.135	0.134	0.120	0.145	0.157	0.118	0.122	0.144	0.105	0.157	0.171	0.107	0.131	0.139	0.110	0.153	0.177	0.104
	$\hat{\beta}_0$	0.001	0.025	−0.001	0.027	0.005	0.031	0.002	0.006	0.004	0.005	−0.001	0.020	−0.001	0.002	0.001	0.003	0.004	0.012
	se $\left(\hat{\beta}_0\right)$	0.118	0.012	0.121	0.013	0.209	0.040	0.049	0.002	0.049	0.001	0.103	0.008	0.035	0.001	0.036	0.001	0.075	0.004
	$\hat{\beta}_1$	1.089	0.068	1.120	0.088	0.692	0.117	1.020	0.008	1.031	0.012	0.594	0.168	1.008	0.003	1.014	0.007	0.569	0.187
	se $\left(\hat{\beta}_1\right)$	0.197	0.035	0.218	0.056	0.224	0.046	0.079	0.003	0.082	0.005	0.098	0.006	0.054	0.001	0.056	0.002	0.069	0.002
	Time:																		
	Loop	0.083		0.125				0.264		0.429				0.847		7.276			
	#iter	9		6				9		6				9		6			
	Total	0.804		0.823		0.055		2.372		2.773		0.062		7.503		44.911		0.089	
50	$\hat{\alpha}$	0.485	0.003	0.494	0.003	0.451	0.008	0.508	0.002	0.499	0.003	0.576	0.016	0.502	0.002	0.478	0.003	0.591	0.019
	se $(\hat{\alpha})$	0.057	0.194	0.051	0.200	0.063	0.189	0.048	0.203	0.039	0.211	0.072	0.182	0.049	0.202	0.043	0.206	0.075	0.179
	$\hat{\beta}_0$	−0.001	0.004	−0.001	0.004	0.005	0.005	−0.001	0.001	−0.001	0.001	−0.002	0.003	−0.001	0.0003	0.004	0.000	−0.002	0.002
	se $\left(\hat{\beta}_0\right)$	0.047	0.001	0.047	0.001	0.081	0.005	0.019	0.000	0.019	0.000	0.040	0.001	0.013	0.000	0.014	0.000	0.029	0.000
	$\hat{\beta}_1$	1.023	0.009	1.029	0.010	0.648	0.127	1.003	0.001	1.006	0.001	0.582	0.175	1.003	0.001	1.002	0.000	0.561	0.193
	se $\left(\hat{\beta}_1\right)$	0.079	0.003	0.083	0.004	0.087	0.004	0.034	0.000	0.034	0.000	0.042	0.000	0.024	0.000	0.024	0.000	0.030	0.000
	Time:																		
	Loop	0.189		0.354				3.119		4.878				15.020		59.528			
	#iter	9		6				9		6				9		6			
	Total	1.683		2.210		0.069		29.300		29.992		0.110		142.034		361.537		3.307	
100	$\hat{\alpha}$	0.491	0.002	0.500	0.002	0.453	0.005	0.506	0.001	0.498	0.001	0.576	0.011						
	se $(\hat{\alpha})$	0.040	0.210	0.035	0.214	0.043	0.206	0.034	0.000	0.027	0.221	0.050	0.001						
	$\hat{\beta}_0$	0.001	0.002	0.001	0.002	−0.001	0.002	0.000	0.000	0.000	0.000	−0.001	0.001						
	se $\left(\hat{\beta}_0\right)$	0.032	0.000	0.032	0.000	0.056	0.002	0.013	0.000	0.013	0.000	0.028	0.000						
	$\hat{\beta}_1$	1.009	0.004	1.013	0.004	0.644	0.129	0.997	0.001	1.000	0.001	0.581	0.176						
	se $\left(\hat{\beta}_1\right)$	0.055	0.001	0.057	0.001	0.060	0.001	0.024	0.000	0.024	0.000	0.030	0.000						
	Time:																		
	Loop	1.008		1.069				13.432		23.313									
	#iter	8		6				10		6									
	Total	8.578		6.537		0.055		130.911		142.885		2.058							

TABLE 6: Simulation results for non-symmetric binary **W**

m		10						50						100					
R		N2SLSa		N2SLS		LinGMM		N2SLSa		N2SLS		LinGMM		N2SLSa		N2SLS		LinGMM	
		Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
10	$\hat{\alpha}$	0.470	0.021	0.482	0.020	0.466	0.040	0.490	0.043	0.472	0.021	0.574	0.060	0.490	0.078	0.468	0.026	0.571	0.068
	se $(\hat{\alpha})$	0.134	0.135	0.120	0.145	0.160	0.116	0.122	0.145	0.105	0.157	0.171	0.108	0.131	0.139	0.110	0.153	0.177	0.104
	$\hat{\beta}_0$	−0.001	0.025	−0.004	0.027	0.006	0.032	0.005	0.004	0.004	0.005	−0.001	0.020	0.000	0.003	0.001	0.003	0.004	0.012
	se $\left(\hat{\beta}_0\right)$	0.118	0.012	0.121	0.012	0.211	0.041	0.048	0.001	0.049	0.001	0.103	0.009	0.035	0.001	0.036	0.001	0.075	0.004
	$\hat{\beta}_1$	1.098	0.078	1.118	0.087	0.694	0.117	1.021	0.007	1.031	0.012	0.594	0.168	1.007	0.003	1.015	0.007	0.569	0.187
	se $\left(\hat{\beta}_1\right)$	0.200	0.037	0.215	0.049	0.226	0.047	0.079	0.003	0.082	0.005	0.098	0.006	0.055	0.001	0.056	0.002	0.069	0.002
	Time:																		
	Loop	0.084		0.127				0.264		0.437				0.854		7.435			
	#iter	9		6				9		6				9		6			
	Total	0.795		0.838		0.057		2.389		2.810		0.061		7.498		45.820		0.086	
50	$\hat{\alpha}$	0.486	0.003	0.493	0.003	0.455	0.008	0.508	0.002	0.499	0.003	0.576	0.016	0.502	0.002	0.478	0.003	0.591	0.019
	se $(\hat{\alpha})$	0.057	0.195	0.050	0.200	0.064	0.189	0.048	0.203	0.039	0.211	0.072	0.182	0.049	0.202	0.043	0.207	0.075	0.179
	$\hat{\beta}_0$	−0.001	0.004	−0.001	0.004	0.000	0.005	−0.001	0.001	−0.001	0.001	−0.002	0.003	−0.001	0.000	0.004	0.000	−0.002	0.002
	se $\left(\hat{\beta}_0\right)$	0.047	0.001	0.047	0.001	0.081	0.005	0.019	0.000	0.019	0.000	0.040	0.001	0.013	0.000	0.014	0.000	0.029	0.000
	$\hat{\beta}_1$	1.022	0.009	1.028	0.010	0.648	0.128	1.003	0.001	1.006	0.001	0.582	0.175	1.003	0.001	1.002	0.000	0.561	0.193
	se $\left(\hat{\beta}_1\right)$	0.079	0.003	0.082	0.004	0.088	0.004	0.034	0.000	0.034	0.000	0.042	0.000	0.024	0.000	0.024	0.000	0.030	0.000
	Time:																		
	Loop	0.188		0.366				3.123		4.935				15.257		52.889			
	#iter	8		6				9		6				9		6			
	Total	1.639		2.269		0.071		29.359		30.408		0.111		144.270		321.995		3.378	
100	$\hat{\alpha}$	0.491	0.002	0.500	0.002	0.457	0.005	0.506	0.001	0.498	0.001	0.576	0.011						
	se $(\hat{\alpha})$	0.040	0.210	0.035	0.214	0.044	0.206	0.034	0.000	0.027	0.221	0.050	0.001						
	$\hat{\beta}_0$	0.000	0.002	0.000	0.002	−0.001	0.002	0.000	0.000	0.000	0.000	−0.001	0.001						
	se $\left(\hat{\beta}_0\right)$	0.032	0.000	0.032	0.000	0.056	0.002	0.013	0.000	0.013	0.000	0.028	0.000						
	$\hat{\beta}_1$	1.008	0.004	1.012	0.004	0.643	0.129	0.997	0.001	1.000	0.001	0.581	0.176						
	se $\left(\hat{\beta}_1\right)$	0.055	0.001	0.057	0.001	0.060	0.001	0.024	0.000	0.024	0.000	0.030	0.000						
	Time:																		
	Loop	1.007		1.080				13.464		23.434									
	#iter	8		6				10		6									
	Total	8.367		6.605		0.058		131.219		143.572		2.031							

7.2 Appendix 2

FIGURE 1: Binary competitiveness indicator for the selected MSAs

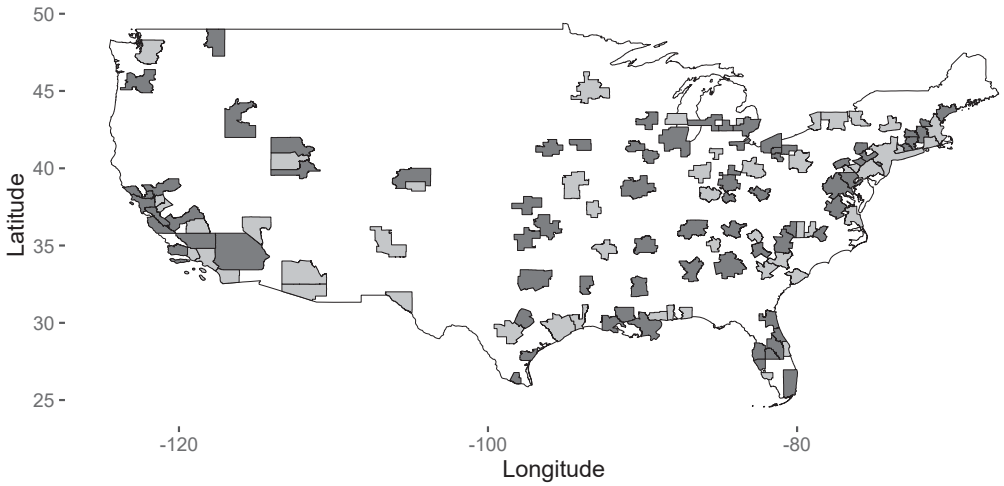

TABLE 7: Descriptive statistics

	Y	AQI_{min}	AQI_{max}	% days O_3	% days $PM_{2.5}$	% days NO_2
Mean	0.628	0.160	1.420	0.361	0.521	0.045
Std. Dev.	0.485	0.076	1.120	0.199	0.212	0.055
Min	0.000	0.030	0.710	0.033	0.027	0.000
Q1	0.000	0.100	1.100	0.212	0.401	0.000
Median	1.000	0.150	1.240	0.295	0.555	0.021
Q3	1.000	0.210	1.530	0.473	0.680	0.077
Max	1.000	0.360	13.410	0.970	0.866	0.219
N	129	129	129	129	129	129

TABLE 8: Spatial lag Probit estimation results

	Dependent variable:		
		Y	
	(N2SLSa)	(N2SLS)	(Lin. GMM)
Constant	6.098^{***}	5.852^{***}	5.422
	(1.860)	(2.220)	(3.338)
AQI_{min}	-4.228^{***}	-3.582^{*}	-2.438
	(1.598)	(1.844)	(4.759)
AQI_{max}	-0.223	-0.229	-0.257
	(0.160)	(0.158)	(1.781)
% days O_3	-5.569^{***}	-5.171^{**}	-4.763^{**}
	(1.858)	(2.112)	(2.316)
% days $PM_{2.5}$	-5.694^{***}	-5.399^{***}	-5.380^{**}
	(1.942)	(2.049)	(2.326)
% days NO_2	-1.865	-1.707	-3.516
	(2.693)	(2.539)	(3.018)
Spatial Lag	0.799^{*}	0.524^{**}	0.601^{*}
	(0.455)	(0.267)	(0.346)
Observations	129	129	129
# Iterations	30	16	
Hansen test	5.402	1.360	4.758
(significance level)	(0.202)	(0.002)	(0.145)
Note:			*p<0.1; **p<0.05; ***p<0.01

Bibliography

Amemiya, T. (1985): *Advanced econometrics*, Harvard University Press.

Anselin, L. (1988): *Spatial Econometrics: methods and models*, Dordrecht: Kluwer Academic.

——— (1990): "Some robust approaches to testing and estimation in spatial econometrics," *Regional Science and Urban Economics*, 20, 141–163.

——— (2001): "Spatial Econometrics," in *A Companion to Theoretical Econometrics*, ed. by B. Baltagi, Oxford: Blackwell, 310–330.

Anselin, L. and A. K. Bera (1998): "Spatial dependence in linear regression models with an introduction to Spatial Econometrics," in *Handbook of Applied Economic Statistics*, ed. by A. Ullah and D. E. A. Giles, New York: Marcel Dekker AG, 237–290.

Arbia, G. (2014): *A primer for spatial econometrics: with applications in R*, Springer.

Bernstein, D. S. (2009): *Matrix Mathematics: Theory, Facts, and Formulas*, Princeton University Press.

Beron, K. J. and W. P. Vijverberg (2004): "Probit in a spatial context: a Monte Carlo analysis," in *Advances in spatial econometrics*, Springer, 169–195.

Bhat, C. R. (2011): "The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models," *Transportation Research Part B: Methodological*, 45, 923–939.

Case, A. (1992): "Neighborhood influence and technological change," *Regional Science and Urban Economics*, 22, 491–508.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *The Review of Economic Studies*, 47, 225–238.

Cliff, A. D. and J. K. Ord (1973): *Spatial autocorrelation*, vol. 5, Pion London.

Cressie, N. A. C. (1993): *Statistics for Spatial Data*, New York: Wiley.

de Bruyn, S. M., J. C. J. M. van den Bergh, and J. B. Opschoor (1998): "Economic growth and emissions: reconsidering the empirical basis of environmental Kuznets curves," *Ecological Economics*, 25, 161–175.

Dudensing, R. M. and D. L. Barkley (2010): "Competitiveness of Southern Metropolitan Areas: The Role of New Economy Policies," *The Review of Regional Studies*, 40, 197–226.

Elhorst, J. P. (2014): *Spatial econometrics: from cross-sectional data to spatial panels*, Springer.

Fleming, M. M. (2004): "Techniques for estimating spatially dependent discrete choice models," in *Advances in spatial econometrics*, Springer, 145–168.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987): "Generalised residuals," *Journal of econometrics*, 34, 5–32.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo maximum likelihood methods: Theory," *Econometrica*, 52, 681–700.

Greene, W. H. (2003): *Econometric analysis*, Pearson Education India.

Griffith, D. A. (2000): "Eigenfunction properties and approximations of selected incidence matrices employed in spatial analyses," *Linear Algebra and its Applications*, 321, 95–112.

Grossman, G. M. and A. B. Krueger (1991): *Environmental impacts of a North American free trade agreement*, w3914, National Bureau of Economic Research.

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

Kelejian, H. H. and D. P. Robinson (1993): "A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model," *Papers in regional science*, 72, 297–312.

——— (1995): "Spatial correlation: a suggested alternative to the autoregressive model," in *New directions in Spatial Econometrics*, ed. by L. Anselin and R. J. G. M. Florax, Berlin: Springer-Verlag, 75–95.

Klier, T. and D. P. McMillen (2012): "Clustering of auto supplier plants in the United States," *Journal of Business & Economic Statistics*.

Lahiri, S. N. (1996): "On inconsistency of estimators based on spatial data under infill asymptotics," *Sankhyā: The Indian Journal of Statistics, Series A*, 403–417.

Lee, L.-F. (2004): "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models," *Econometrica*, 72, 1899–1925.

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

LeSage, J. P. (1997): "Bayesian estimation of spatial autoregressive models," *International Regional Science Review*, 20, 113–129.

——— (2000): "Bayesian estimation of limited dependent variable spatial autoregressive models," *Geographical Analysis*, 32, 19–35.

Maddala, G. S. (1986): *Limited-dependent and qualitative variables in econometrics*, 3, Cambridge University Press.

Martinetti, D. and G. Geniaux (2017): "Approximate likelihood estimation of spatial probit models," *Regional Science and Urban Economics*, 64, 30–45.

McMillen, D. P. (1992): "Probit with spatial autocorrelation," *Journal of Regional Science*, 32, 335–348.

——— (1995): "Spatial effects in probit models: a Monte Carlo investigation," in *New Directions in Spatial Econometrics*, Springer, 189–228.

Millimet, D. L., J. A. List, and T. Stengos (2003): "The environmental Kuznets curve: Real progress or misspecified models?" *Review of Economics and Statistics*, 85, 1038–1047.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85.

Ord, J. (1981): "Towards a theory of spatial statistics: a comment," *Geographical Analysis*, 13, 86–91.

Ord, K. (1975): "Estimation methods for models of spatial interaction," *Journal of the American Statistical Association*, 70, 120–126.

Pace, R. K. and J. P. LeSage (2016): "Fast simulated maximum likelihood estimation of the spatial probit model capable of handling large samples," in *Spatial Econometrics: Qualitative and Limited Dependent Variables*, Emerald Group Publishing Limited, 3–34.

Paelinck, J. H. P. (1967): "L'efficacité des mesures de politique économique régionale," in *Actes du 6ième Colloque Annuel de l'Association de Science Régionale de Langue Française*, ed. by J. H. P. Paelinck, Namur: Centre de Recherches, Facultes Universitaires Notre-Dame de la Paix, 27–75.

Panayotou, T. (1993): *Empirical tests and policy analysis of environmental degradation at different stages of economic development*, 292778, Geneva: International Labour Organization.

Pinkse, J. and M. E. Slade (1998): "Contracting in space: An application of spatial statistics to discrete-choice models," *Journal of Econometrics*, 85, 125–154.

Porter, M. E., J. W. Rivkin, M. A. Desai, and M. Raman (2016): "Problems Unsolved and a Nation Divided," *Harvard Business School*.

Rice, P., A. J. Venables, and E. Patacchini (2006): "Spatial determinants of productivity: analysis for the regions of Great Britain," *Regional Science and Urban Economics*, 36, 727–752.

Rupasingha, A., S. J. Goetz, D. L. Debertin, and A. Pagoulatos (2004): "The environmental Kuznets curve for US counties: A spatial econometric analysis with extensions," *Papers in Regional Science*, 83, 407–424.

Shafik, N. and S. Bandyopadhyay (1992): *Economic growth and environmental quality: time-series and cross-country evidence*, Washington: World Bank.

Smith, T. E. and J. P. LeSage (2004): "A Bayesian probit model with spatial dependencies," in *Spatial and spatiotemporal econometrics*, Emerald Group Publishing Limited, 127–160.

Wang, H., E. M. Iglesias, and J. M. Wooldridge (2013): "Partial maximum likelihood estimation of spatial probit models," *Journal of Econometrics*, 172, 77–89.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*, Cambridge, Massachusetts: MIT press, 2nd ed.