(2013) 43, 131–154

The Review of Regional Studies

The Official Journal of the Southern Regional Science Association

Computational Issues in the Estimation of the Spatial Probit Model: A Comparison of Various Estimators*

Anna Gloria Billé

Department of Economics (DEc), G. D'Annunzio University, Italy

Abstract: In spatial discrete choice models the spatial dependent structure adds complexity in the estimation of parameters. Appropriate general method of moments (GMM) estimation needs inverses of *n*-by-*n* matrices and an optimization complexity of the moment conditions for moderate to large samples makes practical applications more difficult. Recently, Klier and McMillen (2008) have proposed a linearized version of the GMM estimator that avoids the infeasible problem of inverting *n*-by-*n* matrices when employing large samples. They show that standard GMM reduces to a nonlinear two-stage least squares problem. On the other hand, when we deal with full maximum likelihood (FML) estimation, a multidimensional integration problem arises and a viable computational solution needs to be found. Although it remains somewhat computationally burdensome, since the inverses of matrices dimensioned by the number of observations have to be computed, the ML estimator yields the potential advantage of efficiency. Therefore, through Monte Carlo experiments we compare GMM-based approaches with ML estimation in terms of their computation times and statistical properties. Furthermore, a comparison in terms of the *marginal effects* also is included. Finally, we recommend an algorithm based on *sparse matrices* that enables more efficient use of both ML and GMM estimators.

Keywords: spatial econometrics, binary probit model, maximum likelihood, GMM, Monte Carlo simulations

JEL Codes: C1, C31, C35, C55, R10, R12, R15

1. INTRODUCTION

When we specify a spatial autoregressive probit model we generally derive the maximum likelihood function implied by the *reduced form* of the spatial model. A major problem in maximizing the log-likelihood function implied by the reduced form of the spatial model is represented by fact that it repeatedly involves the calculation of determinants of *n* by *n* matrices (related to the $\mathbf{W_n}$ weight matrix) whose dimension depend on the sample size.[1] When *n* is very large this operation can be highly demanding even with the current computational power (Fleming, 2004; Beron and Vijverberg, 2004). For this purpose some solutions have been proposed (Ord, 1975; Griffith, 2000; Smirnov and Anselin, 2001; Pace and LeSage, 2004; Klier and McMillen, 2008). The problem is becoming unmanageable in all those regional economic studies that involve very large, dense matrices, such as applications that pertain to social interactions (Brock and Durlauf, 2007; Goetzke and Andrade, 2010) and that employ microlevel

*I thank Giuseppe Arbia for advising me on my Ph.D. research and for suggesting some ideas for the draft of this paper. Many thanks also to Roger Bivand who suggested the use of the *powerWeights* function.

Ph.D., Economic Statistics and Econometrics, Department of Economics (DEc), Gabriele D'Annunzio University, Chieti, Italy. *Corresponding Author:* A.G. Billé E-mail: ag.bille@unich.it.

[1] That is $|\mathbf{I}_n - \rho\mathbf{W}_n|$ for a Spatial Autoregressive model.

ISSN 1553-0892, 0048-749X (online)
www.srsa.org/rrs

data (Bell and Bockstael, 2000; Bell and Dalton, 2007). Nowadays, overcoming the problem is particularly important because it essentially precludes the opportunity of performing large scale studies that compare detailed spatial units. Although a multidimensional integration problem precludes an analytical solution of the Full Maximum Likelihood (FML) approach and an overly heavy computational load can restrain the use of Maximum Simulated Likelihood (MSL) estimators, in principle ML methods lead to more-efficient estimates. On the other hand, the GMM estimator is simpler to treat, but it suffers from the malady of the ML estimator—even if in a reduced measure—of needing to compute the same n-by-n matrices. In a spatial binary probit context, a comparison between a type of ML estimator—called Partial MLE (PMLE)—and the GMM estimator of Pinkse and Slade (1998) has been recently analyzed by Wang, Iglesias, and Wooldridge (2013). By contrast, Klier and McMillen (2008) have proposed a comparison between that same GMM and their Linearized GMM estimator. However, it seems that no literature made an extensive comparison among the relevant estimators in order to produce some useful information and give some guidelines that could help researchers to choose appropriate estimators in different statistical situations and for different economic purposes. So this is a purpose of the present paper. Moreover, it includes extensive results in terms of the marginal effects, different spatial weight matrices used, and different matrices of instruments specified.

To explain spatial autoregressive probit models and their main estimation techniques, the paper is structured in the following way. Section 2 starts with a brief review of the main classical estimation techniques for spatial discrete choice models. It then contains a description of the spatial mixed autoregressive-regressive probit model, introducing the multidimensional integration problem (Section 3), the maximum likelihood-based estimators (Subsection 3.1), and the generalized method of moments-based estimators (Subsection 3.2). Section 4 introduces Monte Carlo planning, and Section 5 is a collection of the main statistical and computational results based on comparisons between ML and GMM-based approaches (Subsections 5.1 and 5.2). It also displays model comparisons in terms of the marginal effects, different spatial weight matrices, and different matrices of instruments. Section 6 covers interesting results derived by using sparse matrices to efficiently estimate spatial discrete choice models with an ML or GMM estimator. Section 7 concludes. It should be clarified that, despite the increasing importance of Bayesian inference in spatial econometrics (see, e.g., LeSage and Pace, 2009), this paper concentrates on frequentist estimation methods.

2. A SHORT REVIEW OF ESTIMATION TECHNIQUES FOR SPATIAL DISCRETE CHOICES MODELS

In spatial discrete choice models the structure of spatial dependence adds complexity to the estimation of parameters (Fleming, 2004). Beron and Vijverberg (2004) performed a set of Monte Carlo simulations of a spatial linear probability model and compared them with standard and spatial probit models. They found that while the spatial linear probability model is much easier and faster to estimate compared to a spatial probit model, it fails to take into account the dichotomous nature of the dependent variable and is unable to adequately capture spatial dependence. The standard probit model is able to capture the dependent variable's binary nature, but it obviously ignores spatial structure. Thus for modeling dichotomous dependent variables, spatial probit models are superior to spatial linear models, which will most probably become obsolete as access to spatial probit software becomes more widespread.

Pinkse and Slade (1998) proposed a *GMM estimator* for a spatial error probit model with heteroskedasticity by observing that the score vector of the log-likelihood function can be viewed as a set of moment conditions. Unfortunately, since parameters are estimated simultaneously,[2] inverses of n-by-n matrices must be calculated. This means the optimization complexity of the moment conditions for moderate-to-large sized samples can become intractable. More recently, Klier and McMillen (2008) have proposed a linearized version of Pinkse and Slade's GMM that avoids the problem of inverting the n-by-n matrices. Indeed the Klier-McMillen procedure requires no matrix inversions, and needs only standard probit/logit estimation associated with a linear two-stage least squares procedure (2SLS) (see Subsection 5.2 for details). Monte Carlo experiments suggest that the linearized model accurately identifies spatial effects as long as $\rho<0.5$. On the other hand, Klier and McMillen (2008) also acknowledge that the primary advantage of maximum likelihood estimation is the potential for efficiency (see also Wang, Iglesias, and Wooldridge, 2013).[3] The prospect of efficiency may become questionable, however, when the true model is incorrectly defined. Finally, although Klier and McMillen's linearized model provides accurate estimates when the spatial coefficient is small, it yields higher standard errors compared to those from a standard GMM model.

Traditionally, spatial models with continuous dependent variables are estimated by *maximum likelihood (ML) methods* (Anselin, 1988; Arbia, 2006). However, they can be extremely computationally intensive when dealing with spatial models with discrete or limited dependent variables. In fact, under an FML approach, a multidimensional integration problem arises, so approximate computational techniques must be introduced.[4] In practical cases, the likelihood functions associated with spatial autoregressive models cannot be analytically maximized due to the high degree of nonlinearity in the parameters, so numerically approximated solutions must be found.

Beron, Murdoch, and Vijverberg (2003) and Beron and Vijverberg (2004) proposed the *recursive importance sampling (RIS) simulator* (a generalization of the GHK simulator first suggested by Geweke, 1991), which directly deals with n-dimensional integration. Despite its advantages of providing unbiased standard errors of the spatial parameter and of solving directly the high-dimensional integration as well accounting for heteroskedasticity, its main drawback when compared with other computational techniques was its computational burden (see Fleming, 2004). For this purpose, Wang, Iglesias, and Wooldridge (2013) proposed a *Partial Maximum Likelihood Estimation (PMLE) approach* through which they trade off efficiency for less computational burden. The basic idea was to divide observations into many small groups (i.e., clusters) in which adjacent observations belonged to a single group (i.e., pairwise groups), and bivariate normal distributions were specified within each group. Correctly specifying the conditional joint distribution within groups (i.e., using relatively more information provided through spatial correlations), they estimated the model by PMLE, which yields consistent and

[2] So-called "simultaneous equation models" are used to estimate all the parameters at the same time because of the endogeneity of some regressors in spatial models. In the case of ML estimators, this leads to the use of a Full ML (FML) approach which specifies joint probability functions. Due to the simultaneity of these models and the impossibility of observing the latent continuous variables, an easy analytical solution is precluded (Anselin, 2002).

[3] Notice that if the normality assumption is correct, semiparametric estimators are less efficient than parametric equivalents.

[4] The class of *Maximum Simulated Likelihood (MSL) estimator* is typically preferred.

asymptotically normal estimators that are more efficient than Pinkse and Slade's (1998) GMM estimator.

Indeed, Pinkse and Slade (1998) fail to exploit information related to spatial correlations among observations. Obviously, PMLE is not as efficient as the Full MLE. However, since information from adjacent observations typically captures relevant spatial correlations in the sample, PMLE provides a consistent and "relatively" efficient estimator, which avoids computational problems at the expense of some efficiency loss. A more detailed discussion of estimation techniques in spatial discrete choice models can be found in Smirnov (2010) and in Billé and Arbia (2013), who in particular emphasized the discussion on health economics applications.

3. THE SPATIAL AUTOREGRESSIVE-REGRESSIVE PROBIT MODEL

As we know from the spatial econometrics literature (e.g., Anselin, 1988; Anselin, Florax, and Ray, 2004; LeSage and Pace, 2009), we can consider two basic models that introduce spatial dependence in the data: the spatial autoregressive lagged dependent variable (SAL) model and the spatial autoregressive error (SAE) model. In a probit context, we can express the first as:

$$\mathbf{y}^* = \rho\mathbf{W}\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}) \tag{1}$$

$$\mathbf{y} = \mathbf{I}(\mathbf{y}^* > \mathbf{0})$$

where $\mathbf{W}\mathbf{y}^*$ is a column vector of dimension *n* of spatially lagged unobserved variables with an *n*-by-*n* spatial weight matrix $\mathbf{W}$, ρ is the spatial autoregressive coefficient, $\mathbf{X}$ is an *n*-by-*K* matrix of regressors including the constant term, $\boldsymbol{\beta}$ is the corresponding vector of *K* parameters including the intercept, $\boldsymbol{\varepsilon}$ is a column vector of dimension *n* of *iid* normal error terms, and $\mathbf{y}$ is the observed binary column vector of the same dimension. Note that the observed vector $\mathbf{y}$ is an indicator-function vector of the unobserved continuous variables $\mathbf{y}^*$, and therefore probit models are generally defined as *nonlinear models*[5] that describe the probability that the i^{th} observed dependent variable is equal to 1. The model in (1) with spatial dependence and latent form of the continuous variable is known as a *spatial autoregressive probit model (SAPM)* in a structural form (e.g., Fleming, 2004, among others). Under the assumptions that $\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I})$, all of the diagonal elements of W are zero, and $\lambda_{min}^{-1} < \rho < \lambda_{max}^{-1}$[6], so the spatial model can be written in a *reduced form* as

$$\mathbf{y}^* = (\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{2}$$

where $(\mathbf{I} - \rho\mathbf{W})^{-1}$ is the so-called *Leontief inverse*, $\mathbf{u} = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$ and $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ with variance-covariance matrix of the autocorrelated errors $\text{E}(\mathbf{u}\mathbf{u}') = \boldsymbol{\Sigma} = [(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} -$

[5] See for example Greene (2003), Verbeek (2004).

[6] These assumptions are necessary to guarantee the existence of the inverse matrix $(\mathbf{I} - \rho\mathbf{W})^{-1}$, in which case the spatial process is stationary. In particular, if $\mathbf{W}$ is *row-normalized*, this inverse matrix exist for all $\rho: |\rho| < 1$, since in this case $\lambda_{max}^{-1} = 1$ for different sample sizes. We need only to verify that $\lambda_{min}^{-1} \leq -1$ since λ_{min}^{-1} differs for different sample sizes. For details see, e.g., Kelejian and Robinson (1995), Kelejian and Prucha (1998), Bell and Bockstael (2000), LeSage and Pace (2009, chap. 4), Smirnov (2010).

$\rho\mathbf{W})]^{-1}\sigma_\varepsilon^2$. The σ_ε^2 term is usually normalized to 1 to avoid the identification problem[7] of the model. Now define $\mathbf{X}^* = (\mathbf{I_n} - \rho\mathbf{W_n})^{-1}\mathbf{X}$. Since in standard probit models we describe the probability that $y_i = 1$ as $\Phi(\mathbf{x_i'\beta})$, in spatial probit models we can write the same probability as

$$(3) \qquad \mathrm{P}\left(y_i = 1|\mathbf{x_i}, W_{ij}y_j^*\right) = \mathrm{P}\left(\mathrm{y_i^*} \geq 0|\mathbf{x_i}, W_{ij}y_j^*\right) = \mathrm{P}\left(\mathbf{x_i^{*\prime}\beta} + u_i \geq 0|\mathbf{x_i}, W_{ij}y_j^*\right) =$$

$$= \mathrm{P}\left(-u_i < \mathbf{x_i^{*\prime}\beta}|\mathbf{x_i}, W_{ij}y_j^*\right) \cong \Phi(\mathbf{x_i^{*\prime}\beta})\text{where}$$

$\mathbf{x_i^*} = (x_{i1}^*, x_{i2}^*, \ldots, x_{ik}^*, \ldots, x_{iK}^*)'$, $\mathbf{u}$ is a column vector of autocorrelated error terms defined as in (2), and $\Phi(\bullet)$ is the *cdf* for normal distributions. Actually, the previous cumulative density function does not precisely describe $\mathrm{P}\left(y_i = 1|\mathbf{x_i}, W_{ij}y_j^*\right)$ because spatial dependence introduces not only autocorrelation but also heteroskedasticity, which needs to be taken into account. Therefore, the probability that $y_i = 1$ in a spatial autoregressive-regressive probit model is

$$(4) \qquad \mathrm{P}\left(y_i = 1|\mathbf{x_i}, W_{ij}y_j^*\right) = \Phi\left(\frac{\mathbf{x_i^{*\prime}\beta}}{\sigma_i}\right)$$

where σ_i is the i^{th} standard deviation based on the variance-covariance matrix of $\mathbf{u}$ (i.e., $\mathbf{\Sigma}$). In interpreting parameters, the nonlinear nature of the normal *cdf,* as in the standard probit case*,* is complicated by the fact that changes in the level of a single observation, say x_{ik}, have an impact on the expected probability of the event analyzed in both *own*- and *other*-regions due to spatial autocorrelation. Beron and Vijverberg (2004) described the *K*-by-*1* impact vector of changes in the regressor vector $\mathbf{x_i}$ on the probability that $y_i = 1$ as

$$(5') \qquad \frac{\partial \mathrm{P}\left(y_i=1|\mathbf{x_i}, W_{ij}y_j^*\right)}{\partial \mathbf{x_i}} = \phi\left(\frac{\mathbf{x_i^{*\prime}\beta}}{\sigma_i}\right)\frac{\boldsymbol{\beta}^*}{\sigma_i}$$

where $\boldsymbol{\beta}^* = (\mathbf{I_n} - \rho\mathbf{W_n})_{\mathrm{ii}}^{-1}\boldsymbol{\beta}$ and $\phi(.)$ is the *pdf* for normal distributions. It seems clear that changes in explanatory variable values close to the mean may have a more important influence on the probability that $y_i = 1$ than changes in its extreme values, *ceteris paribus*. Therefore, model estimates are often interpreted using mean values of a regressor, say $\mathbf{x}_k$, and the *spatial marginal effects* are then evaluated as the change in the probabilities that the vector $\mathbf{y} = \mathbf{1}$ associated with a change in $\overline{\mathrm{x_k}}$ (see also LeSage and Pace, 2009):

$$(5'') \qquad \frac{\partial \mathrm{P}(\mathbf{y=1|X,Wy^*})}{\partial \mathbf{x_k}} = \phi(\mathbf{D}^{-½}\mathbf{I_n^*}\overline{\mathrm{x_k}}\beta_k)\mathbf{D}^{-½}\mathbf{I_n^*}\beta_k$$[8]

where $\mathbf{I_n^*} = (\mathbf{I_n} - \rho\mathbf{W_n})^{-1}\mathbf{I_n}$ and $\mathbf{D}^{-½}$ is the inverse diagonal matrix of standard deviations based on the variance-covariance matrix of $\mathbf{u}$ (i.e., $\mathbf{\Sigma}$). However, Beron and Vijverberg (2004) highlighted the importance of calculating the single marginal impact *for each observation* x_{ik} and then summarizing this through averaging. Spatial marginal effects are then split into an

[7] Generally speaking, identification problems are related to the lack of information respect to the number of parameters that need to be estimated. In this way, the parameters cannot be univocally estimated and all estimators become inconsistent; in particular, in that case only the ratio β/σ is identified.

[8] Note that in LeSage and Pace (2009) the matrix $\mathbf{D}^{-½}$, which accounts for heteroskedasticity, is not considered (see LeSage and Pace, 2009, page 294, equation (10.10); Beron and Vijverberg, 2004, page 174, equation (8.15)).

average direct impact and an average indirect impact. The average of the diagonal elements of the *n*-by-*n* matrix obtained in (5″) can be interpreted as the *average direct effect* (i.e., the impact from their own regions), the average of the cumulated off-diagonal elements as the *average indirect effect*—also known as *spatial spillover effect* (i.e., the impact from other regions), and finally the *average total effect* as the sum of them (LeSage and Pace, 2009; LeSage et al., 2011). Moreover, LeSage et al. (2011) stressed the need of considering changes in the dependent variable probability relative to the values of $\mathbf{x}_{\mathbf{k}}^{*} = (\mathbf{I}_{\mathbf{n}} - \hat{\rho}\mathbf{W}_{\mathbf{n}})_{\mathbf{ii}}^{-1}\mathbf{x}_{k}$ rather than to the simple $\mathbf{x}_{k}$. They also stressed the need to calculate measures of dispersion for these estimates.

3.1 Maximum Likelihood (ML) Estimation

Consider a standard probit model—that is, a probit model without spatial autocorrelation. If the conditional distribution of $y_i|\mathbf{x_i}$ is known up to a finite small number of parameters $\boldsymbol{\beta}$, it is possible to estimate $\boldsymbol{\beta}$ by choosing values in a way that the distribution is as consistent as possible with the sample. In other words, given the distributional assumption, we define the likelihood of our sample as a function of the unknown parameters $\boldsymbol{\beta}$ that characterize the assumed distribution, and then we choose the values of the parameters in a way that the probability—likelihood—of our sample is maximized. If the distribution assumption is correctly specified,[9] maximum likelihood estimators are *consistent*, *asymptotically efficient* and *asymptotically normal*.[10] Given their potential statistical properties, nonlinear models are generally estimated by maximum likelihood estimation (MLE).

If we instead consider a spatial probit model as in (1), complications arise in the estimation of parameters. First of all, spatial dependence introduces heteroskedasticity, which makes standard probit estimators inconsistent. By assuming independent errors, the likelihood function remains consistent but no longer efficient because it does not use the information in the off-diagonal terms of the variance-covariance matrix $\boldsymbol{\Sigma}$. Moreover, the *problem of multidimensional integration* must be solved in order to address the heteroskedasticity and also to use the additional off-diagonal information (see Appendix A). Therefore, if we refer to the model in (1) with spatial autocorrelation in the dependent unobserved variables $\mathbf{y}^{*}$, the *spatial probit log-likelihood function* for a sample of *n* individuals is

$$(6) \qquad l(\beta, \rho|\mathbf{X}, \mathbf{W}\mathbf{y}^{*}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} y_i \ln \Phi\left(\frac{\sum_{\mathrm{j}=1}^{\mathrm{n}} w_{ij} y_j^{*} + \mathbf{x}_{\mathbf{i}}'\boldsymbol{\beta}}{\sigma_i}\right) + \sum_{\mathrm{i}=1}^{\mathrm{n}} (1 - y_i) \ln\left[1 - \Phi\left(\frac{\sum_{\mathrm{j}=1}^{\mathrm{n}} w_{ij} y_j^{*} + \mathbf{x}_{\mathbf{i}}'\boldsymbol{\beta}}{\sigma_i}\right)\right]$$

where the term $\mathbf{W}\mathbf{y}^{*}$ is *exogenous*[11] and σ_i is the i^{th} standard deviation based on $\boldsymbol{\Sigma}$. However, the term $\mathbf{y}^{*}$ is unobserved so the expression in (6) is not operational (see also Anselin, 2002). As a consequence, we need to consider the reduced form in (2) to define a feasible likelihood function. For this purpose, define $\mathbf{X}^{*} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}$ and $\mathbf{u} = (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}$. The reduced form in (2) is

[9] This is a strong assumption. Many authors have highlighted the unfortunate consequences of maximum likelihood estimates when a model is *misspecified*—that is, when the actual distribution is not exactly what we have assumed (see e.g., Bera, Jarque, and Lee, 1984; Bera and Bilias, 2002; among others). In some cases, however, the maximization of a misspecified loglikelihood function can arrive at a consistent estimator, i.e., Quasi-ML estimator (also known as Pseudo-ML or Maximum Composite Likelihood (MCL) estimator).

[10] For an overview of the statistical properties of maximum likelihood estimators in econometrics see, e.g., Verbeek (2004).

[11] Opposed to ML estimation, the GMM estimation correctly treats the spatial lagged variables $\mathbf{W}\mathbf{y}^{*}$ endogenously (see Section 3.2).

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma}) \tag{7}$$

where $\boldsymbol{\Sigma} = [(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W})]^{-1}\sigma_\varepsilon^2$ as before. The *log-likelihood function implied by the reduced form* of the spatial autoregressive-regressive probit model will be

$$l(\beta, \rho|\mathbf{X}, \mathbf{W}\mathbf{y}^*) = \sum_{i=1}^{n} y_i \ln \Phi\left(\frac{\mathbf{x_i}^{*\prime}\boldsymbol{\beta}}{\sigma_i}\right) + \sum_{i=1}^{n}(1 - y_i)\ln\left[1 - \Phi\left(\frac{\mathbf{x_i}^{*\prime}\boldsymbol{\beta}}{\sigma_i}\right)\right] \tag{8}$$

where σ_i is the i^{th} standard deviation based on $\boldsymbol{\Sigma}$ as before. Note that the term $\Phi\left(\frac{\mathbf{x_i}^{*\prime}\boldsymbol{\beta}}{\sigma_i}\right)$ is exactly that found in (4)— i.e., the probability that $y_i = 1$ in the spatial autoregressive probit model.

3.2 Generalized method of moments estimation and its linearization

The generalized method of moments (GMM) approach (Hansen, 1982) includes the general class of estimators to which ordinary least squares (OLS) estimators, instrumental variable (IV) estimators, generalized instrumental variable (GIV) estimators, and two-stage least squares (2SLS) estimators belong. GMM can allow nonlinear estimation and it differentiates from ML in that it does not assume a functional form for the model, but only general hypotheses on moments. In this sense the GMM estimator is considered a semiparametric estimator. In particular, the parameters have to be directly estimated with some moment conditions specified by the model, and, in order to assure the identifiability of the model, the number of these conditions have to be equal or bigger than the number of the parameters. In some cases, as the following one, the ML function can be used to derive the set of moment conditions to be used in a GMM approach, by observing that the first order conditions of the likelihood maximization problem can be seen as sample means based on theoretical moment conditions. Although the GMM estimator cannot generally be calculated in an analytical way, it is possible to demonstrate that it is *consistent* and *asymptotically normal* under some regularity conditions, although it remains less efficient than the ML estimator.

Kelejian and Prucha (1998, 1999) introduced the GMM approach to spatial econometrics. When it comes to spatial discrete choices, however, Pinkse and Slade's (1998) GMM approach has been key. In particular, their GMM technique is based on the moment conditions implied by the likelihood function for a s*patial error probit model (SEPM)* with heteroskedasticity. Consider the following model in a latent form

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad \mathbf{u} = \lambda\mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \tag{9}$$

$$\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, \sigma_\varepsilon^2\,\mathbf{I})$$

where all the variables and parameters are the same as in (1). If the same assumptions for the model in (2) hold, the model can be written in a reduced form as

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{10}$$

where $\mathbf{u} = (\mathbf{I} - \lambda\mathbf{W})^{-1}\boldsymbol{\varepsilon}$ and $\mathbf{u} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Omega})$ with variance-covariance matrix $\text{E}(\mathbf{u}\mathbf{u}') = \boldsymbol{\Omega} = [(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})]^{-1}\sigma_\varepsilon^2$. Again, the σ_ε^2 term is usually normalized to 1 to avoid identification problems with the model. As shown in Section 3.1, for a spatial error probit model as in (9) the *heteroskedastic log-likelihood function* is equal to

$$l(\beta, \lambda|\mathbf{X}, \mathbf{W}\mathbf{u}) = \sum_{i=1}^{n} y_i \ln \Phi\left(\frac{\mathbf{x_i}'\boldsymbol{\beta}}{\sigma_i}\right) + \sum_{i=1}^{n}(1 - y_i)\ln\left[1 - \Phi\left(\frac{\mathbf{x_i}'\boldsymbol{\beta}}{\sigma_i}\right)\right] \tag{11}$$

where σ_i is the i^{th} standard deviation based on $\mathbf{\Omega}$. From the log-likelihood function, Pinkse and Slade (1998) defined the theoretical moments conditions for the heteroskedastic spatial error probit model as

$$\text{(12)} \qquad \mathrm{E}[\mathrm{f}(y_i, \mathbf{x_i}, \mathbf{h_i}, \beta, \lambda)] = \mathrm{E}\left[\mathbf{h_i'} \frac{\left(y_i - \Phi\left(\frac{\mathbf{x_i'\beta}}{\sigma_i}\right)\right)\phi\left(\frac{\mathbf{x_i'\beta}}{\sigma_i}\right)}{\Phi\left(\frac{\mathbf{x_i'\beta}}{\sigma_i}\right)\left(1-\Phi\left(\frac{\mathbf{x_i'\beta}}{\sigma_i}\right)\right)}\right] = 0$$

and so, the sample moments were equal to

$$\text{(13)} \qquad \mathrm{m}(\beta, \lambda) = \frac{1}{n}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathbf{h_i'} \left[\frac{\left(y_i - \Phi\left(\frac{\mathbf{x_i'\beta}}{\sigma_i}\right)\right)\phi\left(\frac{\mathbf{x_i'\beta}}{\sigma_i}\right)}{\Phi\left(\frac{\mathbf{x_i'\beta}}{\sigma_i}\right)\left(1-\Phi\left(\frac{\mathbf{x_i'\beta}}{\sigma_i}\right)\right)}\right]$$

where h_i is the i^{th} row of a matrix of instruments $\mathbf{H}$[12], $\phi(.)$ is the standard normal density function and $\Phi(.)$ is the cumulative normal density function. Note that in the square parenthesis we have the so-called *generalized probit residuals* for a spatial autoregressive-regressive error probit model. If the number of the moments conditions is greater than that for parameters, as it is often the case, it is not possible to univocally identify the estimate of the parameters by simply setting $\mathrm{m}(\beta, \lambda) = 0$. But it is possible to choose the set of GMM estimates so that the sample moments vector tend to be close to zero. Therefore, GMM minimizes the criteria

$$\text{(14)} \qquad \mathrm{S}(\beta, \lambda) = \mathbf{m}(\beta, \lambda)'\mathbf{M^{-1}}\mathbf{m}(\beta, \lambda)$$

where $\mathbf{M}$ is any positive-definite matrix that defines the weight that has to be assigned to the different sample moments $\mathbf{m}(\beta, \lambda)$.[13] Since (14) cannot be calculated analytically, the GMM estimator is found numerically with an appropriate algorithm. The GMM estimator has to estimate all the parameters together (see footnote 2), which requires the evaluation of $\mathbf{\Omega}$ for any candidate choice of λ as part of the nonlinear optimization of the minimization criteria. Due to the complex form of $\mathbf{\Omega}$, which includes inverses of *n*-by-*n* matrices that depend on the spatial parameter, the optimization problem can become quite intractable.

Recently, Klier and McMillen (2008) have proposed a linearized GMM (LGMM) estimation to avoid the problem of inverting large *n*-by-*n* matrices. They used Pinkse-Slade GMM approach and its linearized version on a spatial logit model.[14] The linearized GMM estimates the model in two steps. The first is a *standard logit model*, in which spatial autocorrelation and heteroskedasticity are ignored. The second involves *2SLS estimates* of the linearized model. Therefore, standard GMM reduces to nonlinear 2SLS. Now, consider the spatial autoregressive probit model in (1). Their basic idea consists of making an explicit

[12] The choice of the set of instruments $\mathbf{H}$ is particularly important and in some cases it cannot be identified. In particular, a necessary condition is the high correlation between the set of instruments and the endogenous regressors in the model. Moreover, the condition $E(\varepsilon_i h_i) = 0$ has to be verified.

[13] In particular, in order to find the GMM estimator with lowest covariance matrix, the optimal weight matrix $\mathbf{M}^{\text{opt}}$ is usually chosen as the inverse of the variance-covariance matrix of the sample moments. If instead $\mathbf{M} = (\mathbf{H'H})^{-1}$, the GMM estimator reduces to nonlinear 2SLS estimator.

[14] The linearized GMM approach can be easily extended to the spatial mixed autoregressive-regressive probit models in a reduced form as in (2).

approximation and linearization of the model around a convenient starting point, which is the standard probit estimation of the model in (1). In fact, when ρ=0, no matrices need to be inverted because $(\mathbf{I}-\rho\mathbf{W})=\mathbf{I}$ and standard probit estimators are consistent for $\boldsymbol{\beta}$. The estimation procedure involves the following two steps.

First Step: Standard probit estimation of $\boldsymbol{y}$ on $\boldsymbol{X}$. Estimate the standard probit model, without autocorrelation and heteroskedasticity, with the standard probit estimator. The estimated values are $\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}}$. Calculate the generalized residuals, e_i, as $\frac{\left(y_i-\Phi(\mathbf{x_i'}\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}})\right)\phi(\mathbf{x_i'}\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}})}{\Phi(\mathbf{x_i'}\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}})\left(1-\Phi(\mathbf{x_i'}\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}})\right)}$, and define the gradient terms as

- $G_{\beta_i}=\phi(\mathbf{x_i'}\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}})\left(1-\Phi(\mathbf{x_i'}\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}})\right)\mathbf{x_i'}$ for $\boldsymbol{\beta}$ and
- $G_{\rho_i}=\phi(\mathbf{x_i'}\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}})\left(1-\Phi(\mathbf{x_i'}\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}})\right)(\mathbf{WX})_i\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}}$ for ρ, where $(\mathbf{WX})_i$ is the i^{th} row of the matrix in parenthesis.

Second Step: 2SLS estimation. First, regress $\mathbf{G_\beta}$ and $\mathbf{G_\rho}$ on the matrix of instruments $\mathbf{H}$. The predicted values are $\widehat{\mathbf{G}}_{\boldsymbol{\beta}}$ and $\widehat{\mathbf{G}}_{\boldsymbol{\rho}}$. Second, define $\mathbf{v}=\mathbf{e}+\mathbf{G}'_{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}_{\mathbf{st.p.}}$ and regress $\mathbf{v}$ on $\widehat{\mathbf{G}}_{\boldsymbol{\beta}}$ and $\widehat{\mathbf{G}}_{\boldsymbol{\rho}}$. The coefficients are the estimated values of $\boldsymbol{\beta}$ and ρ.

Assuming that the true structure of the model is given by (1), Klier and McMillen (2008) have found that the linearization provides accurate estimates as long as ρ is small. Thus, in general, the linearized model can yield a reasonable approximation to the unknown underlying spatial model. When compared to the standard GMM, however, the linearized version is less efficient due to higher standard errors.

4. MONTE CARLO SIMULATION

We perform SAPM as in (1) with a single regressor $\mathbf{x}$, where $\mathbf{X}=(\mathbf{1},\mathbf{x})$ is an n-by-2 matrix with a column vector of ones for the constant term and a regressor $\mathbf{x}$, whereas $\boldsymbol{\beta}=(\beta_0, \beta_1)'$ is a vector of two parameters. In our experiments $\mathbf{x}$ is fixed for each replication of an experiment and drawn from a U(-1,1) distribution. The autoregressive parameter ρ varies from -0.9 to 0.9, whereas β_0=0 and β_1=1 are fixed for all experiments. The spatial weight matrices,[15] $\mathbf{W}_n$, are *n*-by-*n row-normalized* matrices based on both *queen-contiguity* and *k-nearest neighbors* criteria. In addition to *regular square lattice grids*, the $\mathbf{W}_n$ are also built on *irregular shapes* (see Figure 1). We distinguish the experimental plane in order to consider the ML estimation, the GMM, or the Linearized GMM estimation.[16] For all three, we previously define $\mathbf{X}^*=(\mathbf{I}-\boldsymbol{\rho}\mathbf{W})^{-1}\mathbf{X}$ in order to account for autocorrelation structure, and $\mathbf{X}^*_{iS}=\mathbf{X}^*_i/\boldsymbol{\sigma_i}$ to correct the heteroskedasticity introduced by the spatial autocorrelation.[17]

[15] For different weight matrices $\mathbf{W}_n$ we calculated the values of λ_{min}^{-1} (λ_{max}^{-1} is always equals to 1 in row-normalized $\mathbf{W}_n$), which vary with sample sizes and the different criteria used. It has been verified that the process is stationary for all $\rho=[-0.9;0.9]$ according to our experimental plane. See footnote 6 for references and details on general stationary conditions of the spatial process.

[16] We perform all the experiments in R language, and we use the *spdep* and the *McSpatial* packages developed by Roger Bivand and Daniel McMillen, respectively.

[17] Remember that σ_i is the i^{th} standard deviation based on the variance-covariance matrix $\boldsymbol{\Sigma}$.

Therefore, the predicted probabilities that y_i is equal to 1 constitute a column vector $\mathbf{p} = \mathbf{\Phi}(\mathbf{X_S^*\beta})$ as in Equation (4). For the GMM case we introduce the initial values of the parameters in the GMM estimation function, and ρ=[-0.9;0.9], and we consider a set of instruments equals to $\mathbf{H} = [\mathbf{X}, \mathbf{WX}]$. A different set of instruments, $\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W^2X}, \mathbf{W^3X}]$, is also included. Finally, for the Linearized GMM estimation the second step involves 2SLS estimates of the linearized model, in which we consider the set of instruments $\mathbf{H} = [\mathbf{X}, \mathbf{WX}]$. Again, a different set of instruments $\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W^2X}, \mathbf{W^3X}]$ is also included. In order to construct the observed dependent variable y_i for each replication, we draw a vector of errors **e** from a U(0,1) distribution, and we define

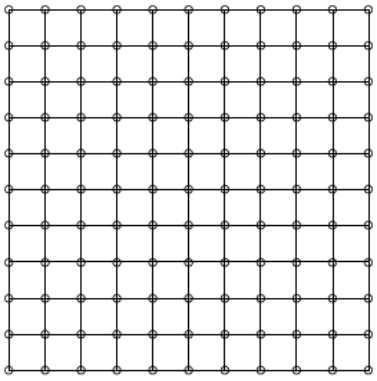
$$\begin{cases} y_i = 1 \quad if\ e_i \leq p_i \\ y_i = 0\ otherwise, \forall i = 1, \dots n \end{cases} \tag{15}$$

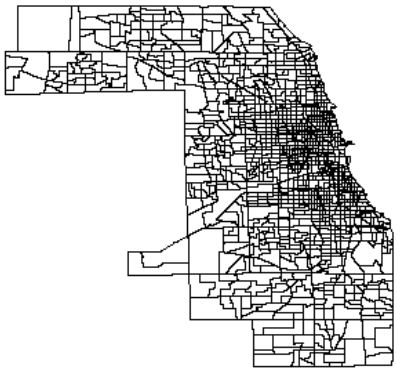
The number of observations *n* are 100 for a 10-by-10 regular square lattice grid, 400 for a 20-by-20 regular square lattice grid, 1,600 for a 40-by-40 regular square lattice grid and finally 1,343 from the actual data set. We run a different number of replications according to the type of the experiment,[18] and summarize the results in terms of the mean, the standard deviation, the mean square error (MSE), the relative absolute bias (RAB) and the relative MSE (RMSE). In particular, the **RABs** are calculated as

$$RAB_{MLvsGMM} = \left|\theta - \widehat{\theta_{ML}}\right| / \left|\theta - \widehat{\theta_{GMM}}\right|, \qquad RAB_{MLvsLGMM} = \left|\theta - \widehat{\theta_{ML}}\right| / \left|\theta - \widehat{\theta_{LGMM}}\right| \tag{16}$$

Figure 1. Shapes from Bivand's and McMillen's Packages

(a) Regular Square Lattice Grid (b) Map of Cook County's Census Tracts





[18] Because of the time used by ML estimator, we run 300 replications when *n*=1,600 and 1,000 replications in all the other cases. Whereas, when *n*=1,343 (actual data set) we run one replication for each estimator in order to compare different estimators with the same sample of observations.

where $\boldsymbol{\theta} = (\beta_0, \beta_1, \rho)'$ is the column vector of fixed parameters, $\widehat{\boldsymbol{\theta}_{\mathbf{ML}}} = \left(\widehat{\beta_0}, \widehat{\beta_1}, \hat{\rho}\right)'$ is the column vector of parameters estimated by Maximum Likelihood, $\widehat{\boldsymbol{\theta}_{\mathbf{GMM}}} = \left(\widehat{\beta_0}, \widehat{\beta_1}, \hat{\rho}\right)'$ is the column vector of parameters estimated by GMM, and $\widehat{\boldsymbol{\theta}_{\mathbf{LGMM}}} = \left(\widehat{\beta_0}, \widehat{\beta_1}, \hat{\rho}\right)'$ is the column vector of parameters estimated by Linearized GMM. Whereas the **RMSEs** are calculated as

$$(17) \qquad RMSE_{MLvsGMM} = MSE_{ML}/MSE_{GMM}, \qquad RMSE_{MLvsL.GMM} = MSE_{ML}/MSE_{LGMM}$$

where $MSE = variance + bias^2$ for each type of estimator.

5. RESULTS

In this section we present the main statistical results based on the previous Monte Carlo scenarios. We first propose in Section 5.1 a comparison between the ML estimation, Pinkse and Slade's (1998) GMM estimation, and Klier and McMillen's (2008) LGMM estimation both in terms of their *statistical properties* and *computational times*. We do so by using the same W_n matrix based both on the queen contiguity-based criterion and on a regular square lattice grid. In Section 5.2 we compare them by using *different weight matrices* based on the irregular shapes of Cook County census tracts and applying *different instruments matrices* for the GMM-based estimators. We also evaluate the relative *marginal effects*, whose particular interpretation has been recently developed by LeSage and Pace (2009) and LeSage et al. (2011) for spatial autoregressive-regressive probit models (SAPMs).

5.1 Statistical properties and computational times

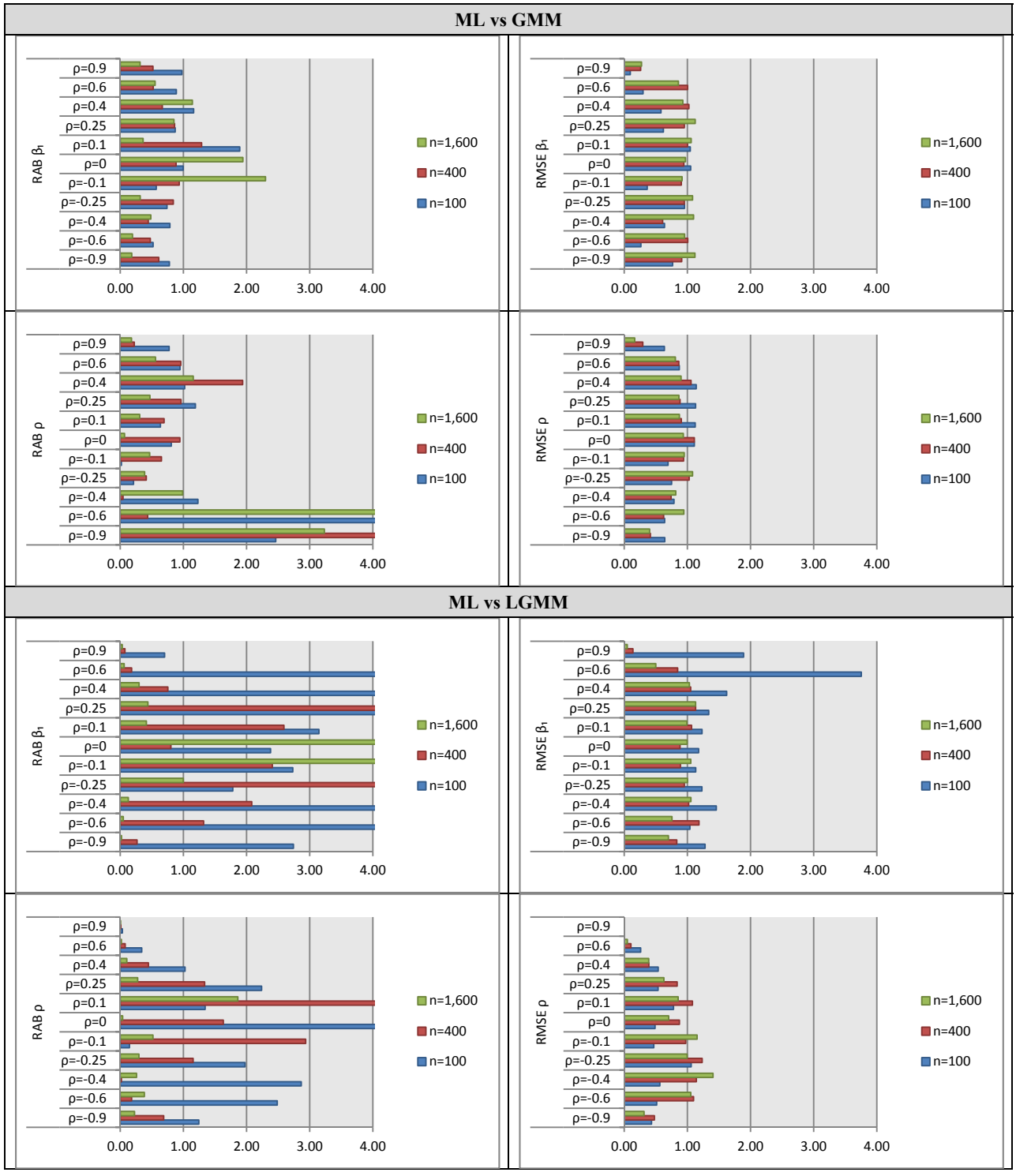
In this subsection we show the main statistical and computational results by comparing ML- to GMM-based estimators.[19] We distinguish the results across different true values of ρ and different sample sizes for both the *unbiasedness* (RAB values) and the *efficiency* (RMSE values) in Figure 2 (see also Table 1b in Appendix C for details). We also briefly compare them in terms of their computation times (Table 1).

First we compare the ML and GMM estimators. The results in terms of the RAB and RMSE values suggest a slight superiority of the ML estimator (i.e., the majority of the values are smaller than or around the value 1.0) with some exceptions. In particular, the GMM estimator seems to be less biased in the estimation of ρ in the presence of higher negative spatial autocorrelations, although with higher MSE values. Moreover, GMM also seems to generate less bias, albeit negligible in the estimation of β_1 yields a ρ value around 0.0. In general, ML seems to estimate the parameters better than does the GMM estimator, especially in the presence of positive spatial autocorrelation and for higher sample sizes (see also Table 1b in Appendix C). Indeed, as expected, in these particular cases the ML estimator appears to perform even better from a comparative perspective.

The LGMM estimator seems to work in a considerably different way. As a matter of fact, the LGMM estimator seems to be less biased than the ML estimator as long as n<1,600 and ρ<0.5, especially in the estimation of β_1. In terms of the RMSEs, the ML estimator is

[19] We should clarify that we had problems when simulating with the GMM estimator, especially for sample sizes smaller than 1,000. In particular, we found that for sample sizes smaller than 1,000 the GMM estimation function provided a lower number of replication results before stopping itself with an error in the "for" loop, whereas it seemed to work quite well as n increased. Moreover, the problem of collecting replication results was more evident as ρ increased in absolute value (i.e. $|\rho| \rightarrow 1$).

Figure 2: A Statistical Comparison between ML and GMM-based Estimators



Note: both RAB and RMSE values bigger than 1.0 indicate an advantage of the GMM-based estimators over the ML estimator in terms of the *unbiasedness* and the *efficiency* respectively. Situations in which the values are close to 1.0 indicate a similar behavior. However, since both RAB and RMSE values vary from 0 to $+\infty$ then distances from 1 to values bigger than 1.0 have a lower "weight" than the same distances to values lower than 1.0, so for significant advantages we need to consider at least RAB and RMSE values bigger than 2.0. For details see Table 1b in Appendix C.

Table 1: Computation Times for the ML, GMM, and LGMM Approaches

Computation times (seconds, tenths, and hundredths)	$n = 100$			$n = 400$			$n = 1{,}600$		
	ML	GMM	LGMM	ML	GMM	LGMM	ML	GMM	LGMM
ρ=-0.9	0.1753	0.0864	0.0105	3.4064	1.3617	0.0159	172.1174	67.7876	0.0422
ρ=-0.6	0.1572	0.0783	0.0102	3.0069	1.3706	0.0159	132.7551	66.6018	0.0412
ρ=-0.4	0.1572	0.0941	0.0102	2.7553	1.4551	0.0156	129.3850	67.3620	0.0419
ρ=-0.25	0.1456	0.0774	0.0103	2.6648	1.3479	0.0159	133.4581	66.8731	0.0414
ρ=-0.1	0.1552	0.0810	0.0102	2.6954	1.3656	0.0158	137.8382	66.8699	0.0410
ρ=0	0.1469	0.0748	0.0109	2.6686	1.2298	0.0159	142.4324	59.3239	0.0421
ρ=0.1	0.1400	0.0713	0.0104	2.7042	1.2897	0.0156	142.4667	66.0366	0.0416
ρ=0.25	0.1488	0.0830	0.0103	2.7985	1.2886	0.0159	144.1595	66.9285	0.0412
ρ=0.4	0.1401	0.0819	0.0102	2.7249	1.3574	0.0155	145.9114	68.7565	0.0413
ρ=0.6	0.1481	0.0748	0.0102	2.9458	1.2829	0.0158	153.8163	66.2017	0.0415
ρ=0.9	0.1727	0.0942	0.0104	3.3259	1.4351	0.0159	173.6253	69.4316	0.0412

Note: Times over one single replication for different values of ρ and different sample sizes.

considerably more efficient than its LGMM equivalent, especially for the estimation of ρ and for higher positive spatial autocorrelations. In addition, the LGMM estimator seems to work well in the presence of negative spatial autocorrelations in terms of the unbiasedness of β_1. In conclusion, for smaller sample sizes and as long as the spatial autocorrelation is not higher than approximately 0.5, LGMM is preferred especially for the estimation of β_1. By contrast, in all other cases the ML estimator seems to work better.

Consider the following results in terms of the *computation times*. In Table 1 we show together the overall results of the computational times in seconds, tenths, and hundredths required by ML and GMM-based estimators. When n=100 and n=400, time consumption between positive and negative values of ρ is approximately the same, especially for the Linearized GMM estimator. Whereas, when n=1,600 we need to distinguish the case when ρ assumes positive values from the case when it assumes the negative ones, since the time used especially by ML estimator in some particular cases is now significantly different. For all different sample sizes the Linearized GMM estimator has a significant advantage over both ML and GMM estimators. Moreover, as the value of ρ rises, the distance between those estimators increases. The Linearized GMM estimator seems to approximate a *constant time trend* for different positive and negative values of ρ and different sample sizes. On the contrary, the ML and GMM estimators seem to have an *exponential time trend* as ρ increases in absolute value, especially for higher sample sizes. The computation time for the ML estimator is about twice that of the GMM estimator for different values of ρ and different sample sizes. Finally, the augmentation of the time used by ML with respect to different values of ρ appears asymmetrical around the value zero, minimizing where $\rho = -0.4$.

5.2 Different weight matrices, different instruments, and marginal effects

In spatial econometrics with discrete choices, when space is tessellated into units, spatial autocorrelation is usually expressed in terms of relationships on a graph rather over distance, and a *spatial weights matrix* is needed to describe the assumed underlying spatial processes (Bivand,

2008). Moreover, the arbitrary nature of the representation of spatial processes, and hence the choice of the $\mathbf{W}_n$ matrix, seems to be unavoidable.[20] In a parametric spatial econometric context, one possible solution consists in using different definitions of the spatial weight matrix (which are reasonable for our economic purposes), in order to check if they lead to significant differences in the estimates of the parameters, and in particular in the relative *marginal effects*. Moreover, according to Section 3.2, appropriate instruments have to be found when we use GMM-based estimators. Therefore, it could be interesting to compare those estimators based on differently defined sets of instruments.

In this subsection we compare ML and GMM-based estimators by using different weight matrices (W_n) and different H instruments matrices according to our Monte Carlo scenarios (Section 4). For this comparison we use the Cook County Census Tracts data set (n=1,343) and differentiate the results across different values of ρ ($\rho = \pm 0.1$; $\rho = \pm 0.4$). Table 2(a,b) first displays the *estimates* of the ML, the GMM based on $\mathbf{H} = [\mathbf{X}, \mathbf{WX}]$ instruments, the GMM based on $\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W^2X}, \mathbf{W^3X}]$, the LGMM based on $\mathbf{H} = [\mathbf{X}, \mathbf{WX}]$ instruments, and finally the LGMM based on $\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W^2X}, \mathbf{W^3X}]$ by using the *same sample of observations*. Then, for each, we determine the marginal effects based on the Equation (5″) in two ways: (*i*) without considering the matrix $\mathbf{D}^{-½}$ (column (1) in Table 2) based on LeSage and Pace (2009, p. 294, Equation (10.10)); (*ii*) considering the matrix $\mathbf{D}^{-½}$ (column (3) in Table 2) based on Beron and Vijverberg (2004, p. 174, Equation (8.15)) (see also note 8). The resulting matrix is then split into the *average direct effect* and the *average indirect effect* (which is due to *spatial spillover effects*) as suggested in Section 3. We also include the same marginal effects based on the estimated values of ρ (columns (2) and (4)).

First, it is clear that for different levels of lower-medium spatial autocorrelations, especially for β_0 and β_1, the estimated values of all the estimators considered are approximately the same. This suggests that GMM-based estimators do not lead to changes in the estimates when we specify different sets of instruments. However, significant differences, especially in the ρ parameter, can be found if we consider different spatial weight matrices, in our case one based on the queen's criterion, $\mathbf{W}_q$, and another based on the *k*-nearest neighbors with $k = 10$, $\mathbf{W}_{knn}$. As for the interpretation of marginal effects, the *direct impact* represents the impact of changes in x_{ik} on $P(y_i = 1)$, whereas the *indirect impact* represents the impact of changes in x_{ik} on $P(y_j = 1; j \neq i)$ due to the proximity in space.

As we can observe from Table 2(a,b), the *total* marginal effects are the same if we used the formulae from either Beron-Vijverberg or LeSage et al. Thus, the correction for heteroskedasticity introduced in the marginal effects have no significant impact on the total average. However, it seems that in the presence of low or medium positive spatial autocorrelations (i.e., ρ=0.1; ρ=0.4) the average *indirect* impact is somewhat higher (consequently the *direct* is lower) when we correct for heteroskedasticity, in which case the

[20] The problem of choosing an appropriate weight matrix **W** is one of the most controversial discussions in the literature (Bell and Bockstael, 2000; Anselin, 2002; Bell and Dalton, 2007; Robinson, 2008; Bivand, 2008; Anselin, 2010). In some cases it has the disadvantage of imposing restrictive structure that can bias results when inappropriate (McMillen and McDonald, 2004; Klier and McMillen, 2008). For that reason, *locally weighted regressions (LWRs)* or *geographically weighted regressions (GWRs)* (e.g., Fotheringham, Charlton, and Brunsdon, 1998; LeSage, 2004), as well as other nonparametric estimation techniques, are becoming prevalent also in discrete choice setting.

Table 2a: A Comparison between ML and GMM Estimators with the Same Sample of Observations

MARGINAL EFFECTS ON x		ESTIMATES			(1) LeSage-*et al.*			(2)LeSage-*et al.* with $\hat{\rho}$			(3) Beron-Vijverberg*			(4)Beron-Vijverberg* with $\hat{\rho}$		
		β_0	β_1	ρ	direct	indirect	total	direct	indirect	total	direct	indirect	total	direct	indirect	total
$\rho = -0.4 \mid W_q$	ML	-0.0210	1.0280	-0.2731	0.4189	-0.1261	0.2928	0.4142	-0.0922	0.3220	0.4058	-0.1130	0.2929	0.4079	-0.0859	0.3220
	GMM**	-0.0210	1.0280	-0.2749	0.4189	-0.1261	0.2929	0.4143	-0.0927	0.3216	0.4058	-0.1130	0.2929	0.4079	-0.0863	0.3216
	GMM2***	-0.0209	1.0270	-0.2629	0.4185	-0.1259	0.2926	0.4135	-0.0892	0.3243	0.4054	-0.1129	0.2926	0.4076	-0.0833	0.3243
	L.GMM**	-0.0212	1.0216	-0.2538	0.4163	-0.1253	0.2910	0.4111	-0.0861	0.3249	0.4033	-0.1123	0.2910	0.4056	-0.0807	0.3249
	L.GMM2***	-0.0211	1.0208	-0.2528	0.4160	-0.1252	0.2908	0.4107	-0.0858	0.3250	0.4030	-0.1122	0.2908	0.4053	-0.0804	0.3250
$\rho = -0.4 \mid W_{knn}$	ML	-0.0197	1.0420	0.0860	0.4204	-0.1236	0.2968	0.4157	0.0388	0.4545	0.4131	-0.1163	0.2968	0.4152	0.0393	0.4545
	GMM**	-0.0197	1.0421	0.0894	0.4204	-0.1236	0.2969	0.4157	0.0405	0.4562	0.4131	-0.1163	0.2969	0.4153	0.0410	0.4562
	GMM2***	-0.0195	1.0418	0.0862	0.4203	-0.1235	0.2968	0.4156	0.0389	0.4545	0.4130	-0.1162	0.2968	0.4152	0.0394	0.4545
	L.GMM**	-0.0198	1.0415	0.0933	0.4202	-0.1235	0.2967	0.4155	0.0424	0.4580	0.4129	-0.1162	0.2967	0.4150	0.0429	0.4580
	L.GMM2***	-0.0197	1.0413	0.0936	0.4201	-0.1235	0.2966	0.4154	0.0425	0.4580	0.4128	-0.1162	0.2966	0.4149	0.0431	0.4580
$\rho = -0.1 \mid W_q$	ML	-0.0065	1.0372	-0.3333	0.4143	-0.0382	0.3761	0.4200	-0.1097	0.3103	0.4134	-0.0373	0.3761	0.4107	-0.1004	0.3103
	GMM**	-0.0061	1.0378	-0.3578	0.4145	-0.0382	0.3763	0.4212	-0.1163	0.3049	0.4136	-0.0373	0.3763	0.4105	-0.1056	0.3049
	GMM2***	-0.0065	1.0371	-0.3333	0.4142	-0.0382	0.3760	0.4200	-0.1097	0.3102	0.4133	-0.0373	0.3760	0.4106	-0.1004	0.3102
	L.GMM**	-0.0070	1.0280	-0.3233	0.4106	-0.0379	0.3727	0.4159	-0.1061	0.3098	0.4097	-0.0370	0.3727	0.4072	-0.0974	0.3098
	L.GMM2***	-0.0070	1.0279	-0.3232	0.4106	-0.0379	0.3727	0.4159	-0.1060	0.3099	0.4097	-0.0370	0.3727	0.4072	-0.0973	0.3099
$\rho = -0.1 \mid W_{knn}$	ML	-0.0074	1.0474	-0.5089	0.4181	-0.0383	0.3798	0.4254	-0.1485	0.2769	0.4175	-0.0378	0.3798	0.4140	-0.1371	0.2769
	GMM**	-0.0082	1.0469	-0.4705	0.4179	-0.0383	0.3796	0.4242	-0.1402	0.2840	0.4173	-0.0378	0.3796	0.4143	-0.1303	0.2840
	GMM2***	-0.0076	1.0473	-0.5045	0.4180	-0.0383	0.3797	0.4252	-0.1476	0.2776	0.4175	-0.0378	0.3797	0.4140	-0.1364	0.2777
	L.GMM**	-0.0093	1.0366	-0.4002	0.4138	-0.0379	0.3758	0.4183	-0.1230	0.2953	0.4132	-0.0374	0.3758	0.4110	-0.1157	0.2953
	L.GMM2***	-0.0093	1.0359	-0.4000	0.4135	-0.0379	0.3756	0.4180	-0.1229	0.2951	0.4130	-0.0374	0.3756	0.4107	-0.1156	0.2951

Note: The comparisons are made in terms of both parameter estimate values and marginal effects according to different weight matrices ($\mathbf{W}_q$ "queen" - $\mathbf{W}_{knn}$ "k-nearest neighbours"), different **H** instruments matrices and different negative lower/medium values of ρ. Marginal effects are split into direct impacts and indirect impacts.

* Consider the heteroskedasticity in the evaluation of the marginal effects (see Equation 5′).

** GMM-based estimator with instruments $\mathbf{H} = [\mathbf{X}, \mathbf{WX}]$.

*** GMM-based estimator with instruments $\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W^2X}, \mathbf{W^3X}]$.

Table 2b: A Comparison between ML and GMM Estimators with the Same Sample of Observations (continued)

MARGINAL EFFECTS ON x		ESTIMATES			(1) LeSage-*et al.*			(2)LeSage-*et al.* with $\hat{\rho}$			(3) Beron-Vijverberg*			(4)Beron-Vijverberg* with $\hat{\rho}$		
		β_0	β_1	ρ	direct	indirect	total	direct	indirect	total	direct	indirect	total	direct	indirect	total
$\rho = 0.1 \mid W_q$	ML	0.0166	0.9678	0.1596	0.3866	0.0423	0.4289	0.3875	0.0717	0.4593	0.3857	0.0432	0.4289	0.3851	0.0742	0.4593
	GMM**	0.0166	0.9679	0.1660	0.3866	0.0423	0.4289	0.3877	0.0751	0.4628	0.3857	0.0432	0.4289	0.3850	0.0778	0.4628
	GMM2***	0.0167	0.9679	0.1622	0.3866	0.0423	0.4289	0.3876	0.0731	0.4607	0.3857	0.0432	0.4289	0.3851	0.0757	0.4607
	L.GMM**	0.0166	0.9650	0.1768	0.3855	0.0422	0.4276	0.3868	0.0807	0.4675	0.3845	0.0431	0.4276	0.3837	0.0837	0.4675
	L.GMM2***	0.0168	0.9651	0.1763	0.3855	0.0422	0.4277	0.3868	0.0805	0.4672	0.3846	0.0431	0.4277	0.3838	0.0835	0.4673
$\rho = 0.1 \mid W_{knn}$	ML	0.0064	0.9697	-0.0162	0.3871	0.0426	0.4297	0.3868	-0.0062	0.3806	0.3865	0.0432	0.4297	0.3868	-0.0061	0.3806
	GMM**	0.0064	0.9697	-0.0141	0.3871	0.0426	0.4297	0.3868	-0.0054	0.3814	0.3865	0.0432	0.4297	0.3867	-0.0054	0.3814
	GMM2***	0.0066	0.9689	-0.0183	0.3868	0.0426	0.4294	0.3865	-0.0070	0.3795	0.3862	0.0431	0.4294	0.3864	-0.0069	0.3795
	L.GMM**	0.0064	0.9697	-0.0141	0.3871	0.0426	0.4297	0.3868	-0.0054	0.3814	0.3865	0.0432	0.4297	0.3867	-0.0054	0.3814
	L.GMM2***	0.0067	0.9689	-0.0161	0.3868	0.0426	0.4293	0.3864	-0.0061	0.3803	0.3862	0.0431	0.4293	0.3864	-0.0061	0.3803
$\rho = 0.4 \mid W_q$	ML	0.0143	1.0352	0.4615	0.4251	0.2630	0.6882	0.4300	0.3367	0.7666	0.4053	0.2829	0.6882	0.4019	0.3648	0.7667
	GMM**	0.0137	1.0374	0.4739	0.4260	0.2636	0.6896	0.4320	0.3543	0.7864	0.4061	0.2835	0.6896	0.4020	0.3844	0.7864
	GMM2***	0.0133	1.0326	0.4673	0.4240	0.2624	0.6864	0.4294	0.3437	0.7731	0.4042	0.2822	0.6864	0.4005	0.3726	0.7731
	L.GMM**	0.0043	1.0006	0.5933	0.4109	0.2542	0.6651	0.4303	0.5507	0.9810	0.3917	0.2734	0.6651	0.3777	0.6035	0.9811
	L.GMM2***	0.0035	0.9978	0.5935	0.4097	0.2535	0.6633	0.4291	0.5496	0.9787	0.3906	0.2727	0.6633	0.3766	0.6022	0.9788
$\rho = 0.4 \mid W_{knn}$	ML	0.0172	1.0218	0.4325	0.4149	0.2643	0.6792	0.4164	0.3017	0.7181	0.4022	0.2770	0.6792	0.4010	0.3172	0.7181
	GMM**	0.0165	1.0231	0.4482	0.4154	0.2647	0.6801	0.4177	0.3218	0.7395	0.4027	0.2774	0.6801	0.4008	0.3387	0.7395
	GMM2***	0.0167	1.0223	0.4425	0.4151	0.2645	0.6796	0.4171	0.3142	0.7313	0.4024	0.2772	0.6796	0.4007	0.3306	0.7313
	L.GMM**	0.0090	0.9994	0.5867	0.4058	0.2585	0.6643	0.4176	0.5466	0.9642	0.3934	0.2710	0.6644	0.3826	0.5817	0.9642
	L.GMM2***	0.0090	0.9990	0.5864	0.4057	0.2584	0.6641	0.4174	0.5457	0.9631	0.3933	0.2709	0.6641	0.3824	0.5808	0.9632

Note: The comparisons are made in terms of both parameter estimate values and marginal effects according to different weight matrices ($\mathbf{W}_q$ "queen" - $\mathbf{W}_{knn}$ "k-nearest neighbours"), different **H** instruments matrices and different positive lower/medium values of ρ. Marginal effects are split into direct impacts and indirect impacts.
* Consider the heteroskedasticity in the evaluation of the marginal effects (see Equation 5′).
** GMM-based estimator with instruments $\mathbf{H} = [\mathbf{X}, \mathbf{WX}]$.
*** GMM-based estimator with instruments $\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W^2X}, \mathbf{W^3X}]$.

spatial spillover effects play a more important role. By contrast, the opposite can be found if we instead considered negative spatial autocorrelations (i.e., ρ = -0.1; ρ = -0.4). That is, the values are still higher but the magnitude of the impact is reduced due to a negative sign. Finally, note that evaluating marginal effects as functions of $\widehat{\boldsymbol{\rho}}$ produce substantial differences.

6. A COMPUTATIONAL SOLUTION IN R FOR ML AND GMM ESTIMATES

As we showed in Section 5, both ML and GMM estimators suffer from the calculation of n-by-n matrices. In a recent comparison between GMM and LGMM, Klier and McMillen (2008) propose two approximations of the $(\text{I}-\rho\text{W})^{-1}$ matrix to compute GMM estimates in a Monte Carlo experiment with 10,000 observations: (a) a third order expansion $(\mathbf{I}-\rho\mathbf{W})^{-1}=\mathbf{I}+\rho\mathbf{W}+\rho^2\mathbf{W}^2+\rho^3\mathbf{W}^3$, and ($b$) a definition of the weight matrix as $W_{i,j}=0.5$ if $|i-j|=1$ (with $W_{1,2}=W_{n,n-1}=1$). However, the definition of the statistical properties for large samples of our estimators without modifying the $\mathbf{W}_n$ structure seems more appropriate. Without any modification, we found that R was able to manipulate no more than 8,100 observations arranged on a 90-by-90 regular grid.[21] Multicore processing does not help, but more efficient optimization algorithms and programming does. We therefore tried to optimize our R script progressively and removed workspace objects to obtain more RAM. After some matrix manipulations including *matrix partitioning to calculate inverses* (see Appendix B), we were able to obtain the estimates using R from a sample of 10,000 observations. Unfortunately, these manipulations did not also lead to significant computational advantages.[22] In order to estimate computationally efficient spatial autoregressive probit models, a current solution is to use *sparse matrices*.[23] In fact, creating inverses of the $(\mathbf{I}-\rho\mathbf{W})$ matrices by using R's *solve* function is the focus of most computational problems.

In this section we provide some important results obtained by modifying McMillen's *spprobitml* function for ML estimators. In Figure 3 we compare the computation times used by both the *spprobitml* function and the modified McMillen's spprobitml function, which we label *sparsespprobitml*. In this modified function we used the *powerWeights* function available in the *spdep* package, which directly and efficiently calculates the matrix $(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}=\mathbf{X}^*$ for large matrices by using sparse spatial weight matrices. As before, we can calculate $(\text{I}-\rho\text{W})^{-1}$ but this time by post-multiplying it with an identity matrix, that is, $(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{I}=\mathbf{I}^*=(\mathbf{I}-\rho\mathbf{W})^{-1}$. This procedure can be easily extended to the GMM estimator. As we can observe in Figure 3, the computation times are all significantly reduced, especially for negative values of ρ, except for the strongest positive spatial autocorrelation and when n=2,500.

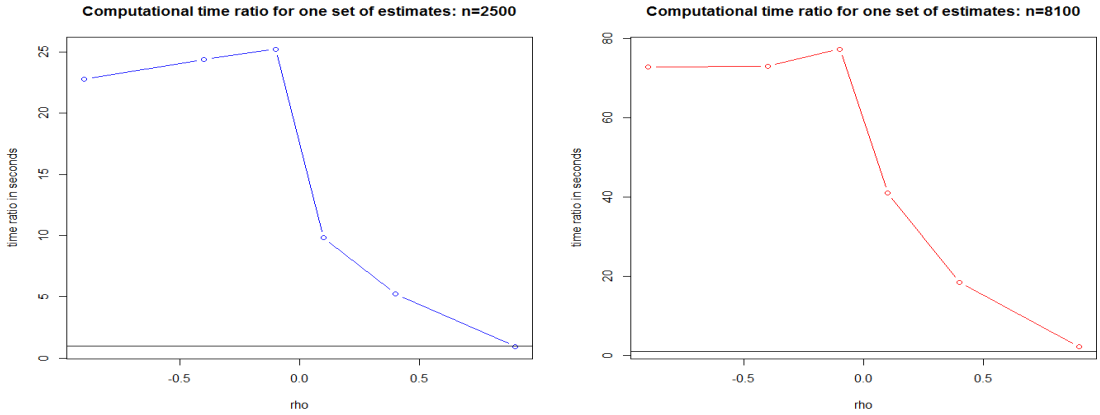
As a consequence, the *powerWeights* function, correctly placed, is able to efficiently substitute for the *solve* function to estimate inverses of large matrices, without a significant effect on estimates (see Table 3). Perhaps using both *partitioned matrices*, which help to manage larges datasets, and *sparse matrices*, which reduce time requirements, can lead to better algorithmic solutions. Regardless we demonstrated that in R (i) the *solve* function is not strictly necessary,

[21] The PC's main characteristics: (i) Intel® Core™ i5-2467M CPU @ 1.60 GHz, (ii) 4 GB RAM.

[22] Instead of directly taking the inverse on a 10,000-by-10,000 matrix, we progressively used R's *solve* function on smaller square matrices (see Appendix B). R provided the estimates with a prohibitive computational time.

[23] See LeSage and Pace (2009) for the use of *sparse matrices* in MatLab for Bayesian estimation methods.

Figure 3: A Comparison of Computation Times between ML Estimators



Note: Computational time ratios in seconds between the *spprobitml* function and the *sparsespprobitml* function in terms of different values of ρ when n=2,500 and n=8,100.

Table 3: A Comparison of Estimate Values between ML Estimators

sample size	n=2,500 (50 x 50 regular grid)						n=8,100 (90 x 90 regular grid)					
parameters	β_0=0		β_1=1		ρ		β_0=0		β_1=1		ρ	
ρ\estimators	ML*	spML**	ML	spML	ML	spML	ML	spML	ML	spML	ML	spML
ρ=-0.9	0.1381	0.3289	1.0345	1.0397	-0.8327	-0.8757	0.0101	0.0191	1.0124	1.0141	-0.8217	-0.8153
ρ=-0.4	-0.0127	-0.0130	0.9805	0.9800	-0.5445	-0.5166	-0.0225	-0.0226	0.9582	0.9582	-0.2137	-0.2151
ρ=-0.1	0.0368	0.0368	1.0111	1.0111	-0.1703	-0.1694	0.0218	0.0218	1.0310	1.0310	-0.2139	-0.2123
ρ=0.1	0.0158	0.0158	0.8826	0.8826	0.1469	0.1469	-0.0154	-0.0154	0.9863	0.9863	0.0721	0.0721
ρ=0.4	-0.0126	-0.0126	0.9843	0.9843	0.3915	0.3915	-0.0134	-0.0134	0.9696	0.9696	0.3717	0.3717
ρ=0.9	-0.0036	-0.0036	0.9878	0.9878	0.9146	0.9146	-0.0014	-0.0014	1.0080	1.0080	0.8975	0.8975

Note: *ML estimates with McMillen's *spprobitml* function; **spML estimates with *sparsespprobitml* function.

especially for large matrices, and certainly costs time; (*ii*) the optimization algorithm and efficient programming techniques are critical when performing ML and GMM estimation; and (*iii*) the time used to calculate the inverse matrices directly depends on the relative location of weights within the spatial weight matrix.

7. CONCLUSIONS

In recent years many researchers have recognized the importance of introducing spatial autocorrelation effects in econometric models. However, when the dependent variable is unobserved some complications arise in the estimation of the parameters. A plausible reason for the relatively scarce diffusion of spatial autoregressive latent models is certainly their complexity. Closed-form solutions are precluded due to the presence of both unobserved dependent variables and spatial autocorrelation. Recently both ML and GMM-based estimators have been proposed. Although a problem of extreme computational intensity exists for ML estimators, especially in the calculation of inverses of n-by-n matrices, they have the potential advantage of providing efficient estimates (Wang, Iglesias, and Wooldridge, 2013). In this vein,

Klier and McMillen (2008) proposed a Linearized GMM estimator, which alleviates some of the computational intractability.

In this paper we simulate a spatial mixed autoregressive-regressive probit model in order to compare statistical properties and computational times of a type of ML estimator with those of GMM-based estimators. Different spatial weight matrices and different sets of instruments are also considered. Comparisons between ML and LGMM suggest that the LGMM works quite well for smaller sample sizes and in the presence of lower levels of spatial autocorrelation. On the other hand, comparisons between ML and GMM reveal strong similarities in both estimates and statistical properties, although the ML in general yielded less biased and more efficient estimates than did the GMM, especially for positive values of ρ. Unfortunately, while somewhat reduced, the computation time required by both the ML and GMM estimators remained rather heavy and rose exponentially with rises in ρ and the sample size. By contrast, the LGMM estimator was substantially more time-efficient, and its time consumption remained fairly constant under different values of ρ and only rose linearly with sample size.

When estimation algorithms are needed, considerations of both the choice of optimization algorithm and the use of time-efficient code are critical. In order to reduce computation times in both ML and GMM frameworks we suggest the use of *sparse matrices.* In particular, we used the *powerWeights* function, which is available in the spdep package, which efficiently calculates the $(\mathbf{I_n} - \rho\mathbf{W_n})^{-1}$ matrix. We find that the times are significantly reduced for different values of ρ and different large sample sizes. Moreover, the estimates did not seem to undergo changes. A method based on *partitioned matrices* also has been introduced, but it seemed to yield no time benefits.

REFERENCES

Anselin, Luc. (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers: Dordrecht, The Netherlands.

Anselin, Luc. (2010) "Thirty Years of Spatial Econometrics," *Papers in Regional Science*, 89, 1, 3–26.

_____. (2002) "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models," *Agricultural Economics*, 27, 247–267.

Anselin, Luc, Raymond J.G.M. Florax, and Sergio J. Rey, eds. (2004) *Advances in Spatial Econometrics*: *Methodology, Tools and Applications*. Springer-Verlag: Berlin.

Arbia, Giuseppe. (2006) *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Springer-Verlag: Berlin.

Bell, Kathleen P. and Nancy E. Bockstael. (2000) "Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data," *Review of Economics and Statistics*, 82, 72–82.

Bell, Kathleen P. and Timothy J. Dalton. (2007) "Spatial Economic Analysis in Data-rich Environments," *Journal of Agricultural Economics*, 58, 487–501.

Bera, Anil K. and Yannis Bilias. (2002) "The MM, ME, ML, EL, EF, and GMM Approaches to Estimation: A Synthesis," *Journal of Econometrics*, 107, 51–86.

Bera, Anil K., Carlos M. Jarque, and Lung-Fei Lee. (1984) "Testing the Normality Assumption in Limited Dependent Variable Models," *International Economic Review*, 25, 563–578.

Beron, Kurt J., James C. Murdoch, and Wim P.M. Vijverberg. (2003) "Why Cooperate? Public Goods, Economic Power, and the Montreal Protocol," *The Review of Economics and Statistics*, 85, 2, 286–297.

Beron, Kurt J. and Wim P.M. Vijverberg. (2004) "Probit in a Spatial Context: A Monte Carlo Approach," in Luc Anselin, Raymond J.G.M. Florax, and Sergio J. Rey, eds., *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer-Verlag: Berlin, pp. 169–196.

Billé, Anna Gloria and Giuseppe Arbia. (2013) "Spatial Discrete Choice and Spatial Limited Dependent Variable Models: A Review with an Emphasis on the Use in Regional Health Economics" working paper available in February 2013 at http://arXiv.org/abs/1302.2267.

Bivand, Roger. (2008) "Implementing Representations of Space in Economic Geography," *Journal of Regional Science*, 48, 1–27.

Brock, William A. and Steven N. Durlauf. (2007) "Identification of binary choice models with social interactions," *Journal of Econometrics*, 140, 52–75.

Fleming, Mark M. (2004) "Techniques for Estimating Spatially Dependent Discrete Choice Models," in Luc Anselin, Raymond J.G.M. Florax, and Sergio J. Rey, eds., *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer-Verlag: Berlin, pp. 145–168.

Fotheringham, A. Stewart, Martin E. Charlton, and Chris Brunsdon. (1998) "Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis," *Environment and Planning A*, 30, 1905–1927.

Geweke, John. (1991) "Efficient Simulation from the Multivariate Normal and Student-*t* Distributions Subject to Linear Constraints" in Elaine M. Keramidas, ed., *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface.* Interface Foundation of North America, Inc.: Fairfax. VA, pp. 571-578. Also available in February 2014 at http://www.censoc.uts.edu.au/pdfs/geweke_papers/gp_47.pdf.

Goetzke, Frank and Patrick M. Andrade. (2010) "Walkability as a Summary Measure in a Spatially Autoregressive Mode Choice Model: An Instrumental Variable Approach," in Antonio Páez, Ron N. Buliung, Julie Le Gallo, and Sandy Dall'erba, eds., *Progress in Spatial Sciences: Methods and Applications.* Springer-Verlag, Berlin, pp. 217–229.

Greene, William H. (2003) *Econometric Analysis*. Prentice-Hall: Upper Saddle River, NJ.

Griffith, Daniel A. (2000) "Eigenfunction Properties and Approximations of Selected Incidence Matrices Employed in Spatial Analysis," *Linear Algebra and Its Applications*, 321, 95–112.

Hansen, Lars P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Kelejian, Harry H. and Ingmar R. Prucha. (1998) "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating A Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17, 1, 99–121.

_____. (1999) "A Generalized Moment Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40, 509–533.

Kelejian, Harry H. and Dennis P. Robinson. (1995) "Spatial Correlation: A Suggested Alternative to the Autoregressive Model," in Luc Anselin, and Raymond J.G.M. Florax, eds., *New Directions in Spatial Econometrics*. Springer Verlag: New York, pp. 75–95.

Klier, Thomas and Daniel P. McMillen. (2008) "Clustering of Auto Supplier Plants in the United States: Generalized Method of Moments Spatial Logit for Large Samples," *Journal of Business & Economic Statistics*, 26, 460–471.

LeSage, James P. (2004) "A Family of Geographically Weighted Regression Models," in Anselin, Luc, Raymond J.G.M. Florax, and Sergio J. Rey, eds., *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer-Verlag: Berlin, pp. 241–264.

LeSage, James P. and R. Kelley Pace. (2009) *Introduction to Spatial Econometrics*. Chapman & Hall/CRC: Boca Raton, FL.

LeSage, James P., R. Kelley Pace, Nina Lam, Richard Campanella, and Xingjian Liu. (2011) "New Orleans Business Recovery in the Aftermath of Hurricane Katrina," *Journal of Royal Statistics Society A*, 174, 1007–1027.

McMillen, Daniel P. and John F. McDonald. (2004) "Locally Weighted Maximum Likelihood Estimation: Monte Carlo Evidence and an Application," in Luc Anselin, Raymond J.G.M. Florax, and Sergio J. Rey, eds., *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer-Verlag: Berlin, pp. 225–239.

Ord, J. Keith. (1975) "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70, 120–126.

Pace, R. Kelley and James P. LeSage. (2004) "Chebyshev Approximation of Log-Determinants of Spatial Weights Matrices," *Computational Statistics and Data Analysis*, 45, 179–196.

Pinkse, Joris and Margaret E. Slade. (1998) "Contracting in Space: An Application of Spatial Statistics to Discrete Choice Models," *Journal of Econometrics*, 85, 125–154.

Robinson, Peter M. (2008) "Developments in the Analysis of Spatial Data," *Journal of Japan Statistical Society*, 38, 87–96.

Smirnov, Oleg A. (2010) "Modeling Spatial Discrete Choice," *Regional Science and Urban Economics*, 40, 292–298.

Smirnov, Oleg A. and Luc Anselin. (2001) "Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: A Characteristic Polynomial Approach," *Computational Statistics & Data Analysis*, 35, 301–319.

Verbeek, Marno. (2004) *A Guide to Modern Econometrics*, 2nd ed., John Wiley & Sons: Hoboken, NJ.

Wang, Honglin, Emma M. Iglesias, and Jeffrey M. Wooldridge. (2013) "Partial Maximum Likelihood Estimation of Spatial Probit Models," *Journal of Econometrics*, 172, 77–89.

APPENDIX A. THE MULTIDIMENSIONAL INTEGRATION PROBLEM

In order to illustrate a multidimensional integration problem for a spatial probit model, consider the reduced form in (2). Then, the joint probability will be

$$\mathrm{P}(\mathrm{Y}_1 = y_1, \mathrm{Y}_2 = y_2, \dots, \mathrm{Y}_\mathrm{n} = y_n | \mathbf{X}, \mathbf{W}\mathbf{y}^*; \beta, \rho) = \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_i} \dots \int_{-\infty}^{a_n} \phi(\mathbf{u}) \mathrm{d}\mathbf{u}$$

where $a_i = [2y_i - 1]\frac{\mathbf{x_i^{*\prime}\beta}}{\sigma_{u_i}}$ and $\phi(\mathbf{u}) = (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1}\exp\left\{-\frac{1}{2}(\mathbf{u'\Sigma^{-1}u})\right\}$ is the multivariate normal density function. In that case, due to the lack of independence, the spatially correlated covariance structure $\mathbf{\Sigma}$ does not allow the typical simplification of the multivariate distribution into the product of univariate distributions as in the standard probit case (see Fleming, 2004 for details).

APPENDIX B. THE INVERSE OF A PARTITIONED MATRIX

Consider a partitioned square matrix **A** of dimension *n* by *n* as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{bmatrix}$$

where the elements $\mathbf{A_{11}}, \mathbf{A_{12}}, \mathbf{A_{21}}, \mathbf{A_{22}}$ are themselves matrices of dimension (*n/2*)-by-(*n/2*) each, and in particular the diagonal blocks $\mathbf{A_{11}}, \mathbf{A_{22}}$ are such that $\exists \mathbf{A_{11}^{-1}}, \exists \mathbf{A_{22}^{-1}}$. Suppose we are interested in the inverse of the matrix **A**. Then, matrix $\mathbf{A^{-1}}$ can be written as

$$\mathbf{A^{-1}} = \begin{bmatrix} \mathbf{A_{11}^{-1}}(\mathbf{I} + \mathbf{A_{12}B_{22}A_{21}A_{11}^{-1}}) & -\mathbf{A_{11}^{-1}A_{12}B_{22}} \\ -\mathbf{B_{22}A_{21}A_{11}^{-1}} & \mathbf{B_{22}} \end{bmatrix}$$

where $\mathbf{B_{22}} = \left(\mathbf{A_{22}} - \mathbf{A_{21}A_{11}^{-1}A_{12}}\right)^{-1}$.

APPENDIX C. Table 1b: Means, Standard Deviations, MSEs of θ=(β_0, β_1,ρ) for the ML, the GMM, and the LGMM Estimators

		ML									GMM									LGMM								
	ρ	μ(β₀)	sd(β₀)	mse(β₀)	μ(β₁)	sd(β₁)	mse(β₁)	μ(ρ)	sd(ρ)	mse(ρ)	μ(β₀)	sd(β₀)	mse(β₀)	μ(β₁)	sd(β₁)	mse(β₁)	μ(ρ)	sd(ρ)	mse(ρ)	μ(β₀)	sd(β₀)	mse(β₀)	μ(β₁)	sd(β₁)	mse(β₁)	μ(ρ)	sd(ρ)	mse(ρ)
n=100	-0.9	0.0028	0.2919	0.0852	1.0716	0.2593	0.0724	-0.6842	0.4174	0.2208	0.0075	0.6169	0.3807	1.0921	0.2935	0.0946	-0.8124	0.5803	0.3445	0.0059	0.2453	0.0602	0.9739	0.2364	0.0566	-0.7266	0.6947	0.5127
	-0.6	0.0152	0.2401	0.0579	1.0656	0.2454	0.0645	-0.4999	0.4919	0.2520	0.0070	0.4277	0.1830	1.1262	0.4785	0.2449	-0.6084	0.6273	0.3936	-0.0099	0.2187	0.0479	1.0021	0.2492	0.0621	-0.5597	0.6999	0.4914
	-0.4	-0.0036	0.2119	0.0449	1.0991	0.2910	0.0945	-0.3004	0.5580	0.3212	-0.0030	0.4216	0.1778	1.1262	0.3635	0.1480	-0.4810	0.6335	0.4078	-0.0107	0.2158	0.0467	1.0159	0.2542	0.0649	-0.4347	0.7539	0.5696
	-0.25	-0.0020	0.2083	0.0434	1.0661	0.2669	0.0756	-0.2320	0.5275	0.2786	0.0062	0.2893	0.0837	1.0894	0.2670	0.0793	-0.3358	0.6036	0.3717	-0.0009	0.1796	0.0323	1.0371	0.2452	0.0615	-0.2409	0.5125	0.2627
	-0.1	-0.0018	0.1863	0.0347	1.0677	0.2759	0.0807	-0.1027	0.5369	0.2882	-0.0721	0.6351	0.4086	1.1190	0.4563	0.2223	-0.2555	0.6254	0.4154	0.0061	0.1854	0.0344	1.0248	0.2658	0.0713	-0.1182	0.7844	0.6157
	0	-0.0005	0.1749	0.0306	1.0669	0.2714	0.0781	-0.0753	0.5371	0.2942	0.0166	0.2063	0.0428	1.0672	0.2641	0.0742	-0.0931	0.5066	0.2653	0.0010	0.1739	0.0302	1.0281	0.2559	0.0663	0.0051	0.7768	0.6034
	0.1	0.0078	0.1651	0.0273	1.1107	0.2813	0.0914	0.0268	0.5539	0.3121	-0.0113	0.3565	0.1272	1.0585	0.2897	0.0873	-0.0150	0.5142	0.2776	0.0048	0.1560	0.0244	1.0352	0.2702	0.0742	0.1544	0.6306	0.4006
	0.25	0.0178	0.1691	0.0289	1.0670	0.2749	0.0800	0.1184	0.4995	0.2668	0.0086	0.1673	0.0281	1.0771	0.3512	0.1293	0.1394	0.4734	0.2364	-0.0030	0.1346	0.0181	1.0133	0.2442	0.0598	0.3088	0.7041	0.4992
	0.4	0.0059	0.1510	0.0228	1.1230	0.3279	0.1226	0.2573	0.5168	0.2875	-0.0013	0.2004	0.0402	1.1058	0.4470	0.2110	0.2601	0.4824	0.2523	-0.0022	0.1213	0.0147	1.0095	0.2748	0.0756	0.5391	0.7186	0.5357
	0.6	-0.0067	0.1314	0.0173	1.1796	0.4678	0.2511	0.4786	0.4128	0.1851	-0.0095	0.3407	0.1162	1.2025	0.8948	0.8417	0.4716	0.4426	0.2124	-0.0669	0.1581	0.0295	0.9563	0.2549	0.0669	0.9581	0.7664	0.7157
	0.9	-0.0001	0.0446	0.0020	1.2193	0.5038	0.3020	0.8487	0.2104	0.0469	0.0212	0.6316	0.3993	1.2248	1.7675	3.1747	0.8337	0.2633	0.0737	0.3168	4.3445	18.9755	0.6873	0.2487	0.1597	2.4037	15.5473	243.9828
n=400	-0.9	-0.0013	0.1414	0.0200	1.0119	0.1227	0.0152	-0.8035	0.2451	0.0694	-0.0003	0.1673	0.0280	1.0195	0.1278	0.0167	-0.8866	0.4092	0.1677	0.0018	0.1201	0.0144	0.9551	0.1275	0.0183	-0.7597	0.3548	0.1456
	-0.6	-0.0015	0.1178	0.0139	1.0118	0.1307	0.0172	-0.5826	0.3236	0.1050	0.0021	0.1430	0.0205	1.0248	0.1284	0.0171	-0.6401	0.4088	0.1687	-0.0041	0.1045	0.0109	0.9911	0.1203	0.0145	-0.5036	0.2939	0.0957
	-0.4	0.0066	0.1002	0.0101	1.0125	0.1249	0.0158	-0.4009	0.3414	0.1166	-0.0055	0.1992	0.0397	1.0281	0.1585	0.0259	-0.4206	0.3957	0.1570	-0.0024	0.0917	0.0084	0.9940	0.1242	0.0155	-0.3495	0.3155	0.1021
	-0.25	-0.0005	0.0920	0.0085	1.0213	0.1188	0.0146	-0.2657	0.3472	0.1208	0.0024	0.0943	0.0089	1.0254	0.1210	0.0153	-0.2881	0.3407	0.1175	0.0018	0.0841	0.0071	0.9997	0.1238	0.0153	-0.2364	0.3127	0.0979
	-0.1	0.0010	0.0823	0.0068	1.0167	0.1152	0.0136	-0.1366	0.3267	0.1081	-0.0011	0.0845	0.0071	1.0179	0.1212	0.0150	-0.1562	0.3346	0.1151	-0.0015	0.0749	0.0056	1.0069	0.1232	0.0152	-0.0875	0.3331	0.1111
	0	0.0003	0.0776	0.0060	1.0102	0.1204	0.0146	-0.0363	0.3166	0.1016	-0.0006	0.0728	0.0053	1.0115	0.1238	0.0155	-0.0384	0.3002	0.0916	-0.0017	0.0711	0.0051	1.0127	0.1279	0.0165	-0.0222	0.3402	0.1163
	0.1	0.0017	0.0684	0.0047	1.0191	0.1232	0.0155	0.0653	0.3027	0.0928	0.0021	0.0707	0.0050	1.0148	0.1237	0.0155	0.0499	0.3166	0.1027	-0.0015	0.0657	0.0043	1.0074	0.1205	0.0146	0.1042	0.2933	0.0860
	0.25	-0.0013	0.0587	0.0035	1.0149	0.1208	0.0148	0.2072	0.2555	0.0671	0.0060	0.0602	0.0037	1.0173	0.1236	0.0156	0.2055	0.2721	0.0760	0.0015	0.0537	0.0029	1.0016	0.1145	0.0131	0.2821	0.2812	0.0801
	0.4	-0.0012	0.0484	0.0023	1.0177	0.1283	0.0168	0.3573	0.2255	0.0527	0.0015	0.0482	0.0023	1.0266	0.1252	0.0164	0.3780	0.2220	0.0498	-0.0045	0.0439	0.0020	0.9765	0.1241	0.0160	0.4958	0.3550	0.1352
	0.6	-0.0012	0.0334	0.0011	1.0086	0.1256	0.0159	0.5817	0.1385	0.0195	-0.0016	0.0349	0.0012	1.0165	0.1246	0.0158	0.5809	0.1491	0.0226	0.0158	0.0316	0.0012	0.9523	0.1283	0.0187	0.8337	0.3611	0.1851
	0.9	-0.0006	0.0196	0.0004	1.0255	0.1335	0.0185	0.8947	0.0424	0.0018	0.0008	0.0193	0.0004	1.0492	0.2619	0.0710	0.8760	0.0753	0.0062	-0.1647	0.1107	0.0394	0.6551	0.1225	0.1341	1.9452	0.5977	1.4507

Note: $n = 100$ and$n = 400$. Simulations are based on both a queen contiguity-based $\mathbf{W}_q$ matrix and a regular square lattice grid. The number of replications is equal to 1,000.

Table 1b: Means, Standard Deviations, MSEs of θ=(β_0, β_1,ρ) for the ML, the GMM, and the LGMM Estimators (continued)

		ML									GMM									LGMM								
	ρ	$\mu(\beta_0)$	$sd(\beta_0)$	$mse(\beta_0)$	$\mu(\beta_1)$	$sd(\beta_1)$	$mse(\beta_1)$	$\mu(\rho)$	$sd(\rho)$	$mse(\rho)$	$\mu(\beta_0)$	$sd(\beta_0)$	$mse(\beta_0)$	$\mu(\beta_1)$	$sd(\beta_1)$	$mse(\beta_1)$	$\mu(\rho)$	$sd(\rho)$	$mse(\rho)$	$\mu(\beta_0)$	$sd(\beta_0)$	$mse(\beta_0)$	$\mu(\beta_1)$	$sd(\beta_1)$	$mse(\beta_1)$	$\mu(\rho)$	$sd(\rho)$	$mse(\rho)$
n=1,600	-0.9	0.0011	0.0646	0.0042	0.9990	0.0666	0.0044	-0.8617	0.1319	0.0189	-0.0056	0.0756	0.0057	1.0055	0.0627	0.0040	-0.8882	0.2163	0.0469	-0.0048	0.0572	0.0033	0.9488	0.0610	0.0064	-0.7300	0.1765	0.0601
	-0.6	-0.0113	0.0556	0.0032	1.0009	0.0592	0.0035	-0.5695	0.1968	0.0397	0.0055	0.0580	0.0034	1.0047	0.0604	0.0037	-0.6025	0.2050	0.0420	-0.0039	0.0497	0.0025	0.9808	0.0654	0.0046	-0.5200	0.1766	0.0376
	-0.4	-0.0023	0.0476	0.0023	1.0011	0.0607	0.0037	-0.4092	0.1850	0.0343	0.0038	0.0493	0.0024	1.0023	0.0579	0.0034	-0.4093	0.2050	0.0421	0.0036	0.0476	0.0023	0.9914	0.0585	0.0035	-0.3643	0.1520	0.0244
	-0.25	0.0015	0.0418	0.0017	1.0030	0.0582	0.0034	-0.2476	0.1906	0.0363	-0.0002	0.0439	0.0019	1.0095	0.0552	0.0031	-0.2438	0.1831	0.0336	-0.0044	0.0404	0.0017	0.9970	0.0582	0.0034	-0.2581	0.1909	0.0365
	-0.1	0.0030	0.0388	0.0015	1.0059	0.0617	0.0038	-0.1087	0.1688	0.0286	0.0021	0.0397	0.0016	1.0026	0.0648	0.0042	-0.1187	0.1725	0.0301	-0.0001	0.0345	0.0012	0.9994	0.0604	0.0037	-0.0832	0.1564	0.0248
	0	-0.0011	0.0332	0.0011	1.0106	0.0608	0.0038	0.0006	0.1467	0.0215	0.0030	0.0351	0.0012	1.0055	0.0624	0.0039	-0.0089	0.1515	0.0230	-0.0029	0.0348	0.0012	1.0012	0.0618	0.0038	0.0167	0.1742	0.0306
	0.1	0.0001	0.0315	0.0010	1.0012	0.0599	0.0036	0.1062	0.1448	0.0210	-0.0011	0.0276	0.0008	1.0033	0.0581	0.0034	0.0798	0.1537	0.0240	0.0008	0.0298	0.0009	1.0029	0.0601	0.0036	0.0967	0.1569	0.0246
	0.25	0.0015	0.0257	0.0007	1.0044	0.0607	0.0037	0.2416	0.1268	0.0161	-0.0008	0.0261	0.0007	1.0052	0.0572	0.0033	0.2321	0.1354	0.0187	0.0011	0.0254	0.0006	0.9899	0.0564	0.0033	0.2804	0.1574	0.0257
	0.4	-0.0005	0.0209	0.0004	1.0061	0.0641	0.0041	0.3913	0.1096	0.0121	-0.0014	0.0205	0.0004	1.0053	0.0666	0.0045	0.3925	0.1157	0.0134	-0.0019	0.0174	0.0003	0.9795	0.0599	0.0040	0.4835	0.1555	0.0312
	0.6	-0.0003	0.0163	0.0003	1.0037	0.0578	0.0034	0.5954	0.0688	0.0048	-0.0004	0.0157	0.0002	1.0067	0.0622	0.0039	0.5917	0.0762	0.0059	0.0012	0.0079	0.0001	0.9402	0.0560	0.0067	0.8543	0.1821	0.0980
	0.9	-0.0003	0.0054	0.0000	1.0093	0.0645	0.0042	0.8990	0.0164	0.0003	-0.0009	0.0102	0.0001	1.0298	0.1210	0.0155	0.8944	0.0398	0.0016	-0.1581	0.0442	0.0270	0.7056	0.0567	0.0902	1.9623	0.2486	1.1940

Note: $n = 1{,}600$. Simulations are based on both a queen-contiguity $\mathbf{W}_q$ matrix and a regular square lattice grid. The number of replications is equal to 300.