UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

The correlation structure of spatial autoregressions on graphs

Martellosio, F.

Publication date
2006
Document Version
Submitted manuscript

Link to publication

Citation for published version (APA):
Martellosio, F. (2006). *The correlation structure of spatial autoregressions on graphs*. (UvA-Econometrics Working Paper; No. 2006/04). Faculteit Economie en Bedrijfskunde. http://www1.feb.uva.nl/pp/bin/458fulltext.pdf

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 Mar 2023

UvA ECONOMETRICS

Discussion Paper: 2006/04

The correlation structure of spatial autoregressions on graphs

Federico Martellosio

www.fee.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

The Correlation Structure of Spatial Autoregressions on Graphs

Federico Martellosio*

University of Amsterdam

August 31, 2006

Abstract

This paper studies the correlation structure of spatial autoregressions defined over arbitrary configurations of observational units. We derive a number of new properties of the models and provide new interpretations of some of their known properties. A little graph theory helps to clarify how the correlation between two random variables observed at two units depends on the walks connecting the two units, and allows to discuss the statistical consequences of the presence (or, more importantly in econometrics, the absence) of symmetries or regularities in the configuration of the observational units. The analysis is centered upon one-parameter models, but extensions to multi-parameter models are also considered.

Keywords: exponential families; graphs; quadratic subspace; spatial autoregressions; spatial weights matrices.

JEL Classification: C12, C21.

1 Introduction

Time series autoregressive processes can be generalized to a spatial setting in two different ways, giving rise to Simultaneous Autoregressive (SAR) and Conditional Autoregressive (CAR) models. The formulation of the models can be traced back to, respectively, Whittle (1954) and Besag (1974). Since then the models have been employed in a wide variety of applications (see, for instance, Cliff and Ord, 1981, Cressie, 1993). In particular, SAR models constitute the building block of most parametric models, and also of some semi-parametric models, currently used in spatial econometrics (e.g., Anselin, 1988, Kelejian and Prucha, 1999 and 2004, Lee, 2002, 2004 and 2006, Baltagi et al., 2003, Giacomini and Granger, 2004, Bao and Ullah, 2006, Robinson, 2006). CAR models, also known as auto-normal schemes, are more popular in other scientific fields such as disease mapping and image analysis. Despite being very different in many respects, CAR and SAR models are similar in terms of their correlation structure (e.g., Cressie, 1993, Wall, 2004).

The purpose of this paper is to analyze correlation properties of CAR and SAR models defined over arbitrary sets of observational units. In the case of regular sets of observational units (e.g.,

*Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB, Amsterdam, The Netherlands. E-mail: `f.martellosio@uva.nl`.

uniform grids), the correlation structure of spatial autoregressions is well-understood (e.g., Besag, 1972 and 1981). In this paper we focus on irregular spatial configurations, which clearly constitute the case of interest in non-experimental fields such as econometrics. We will however also consider the effect that the presence of regularities in the neighborhood structure of the units has on the properties of the models. Recent literature has pointed out that when they are defined over irregular lattices, spatial autoregressions may exhibit some undesirable or unexpected properties (e.g., Besag and Kooperberg, 1995, and Wall, 2004). In fact, it is fair to say that some of the correlation properties of spatial autoregressions on irregular lattices are not completely understood. This has serious practical consequences, because specifying the models may be difficult if their properties are not clear.

The theoretical study of the correlation structure of spatial autoregressions is not straightforward, because, in contrast to the approach that is popular in geostatistics (e.g., Cressie, 1993, Chapter 2), the specification of spatial autoregressions is not based directly on a covariance matrix, but on its inverse. In addition, one could argue that spatial autoregressions should be regarded as simple interpolating devices, and hence that what is ultimately important about the models is their degree of fitting to a given data set, rather than the correlation structure they imply. It is not our intention here to enter these debates, but, given that spatial autoregressions (i) have important advantages (alongside their disadvantages, of course) with respect to models parametrizing directly a covariance matrix,[1] and (ii) are so routinely used in spatial econometrics, it seems to us that a detailed study of the correlations they imply on irregular lattices is much needed.

Large part of our analysis is conducted with the help of some formal graph theory. The use of a graph theoretic terminology is common to discuss how CAR models are constructed (essentially because of the Hammersley-Clifford theorem, which characterizes the class of random fields that are Markovian with respect to a given graph; Besag, 1974), but not to discuss the correlation properties of CAR and SAR models. There are several advantages to adopting a graph theoretic perspective in our analysis. Firstly, a graph endows the set of observational units with a metric that allows to develop a simple interpretation of the correlation structure of the one-parameter, or first-order, models. Secondly, graph theory is helpful to clarify differences among different ways of constructing multi-parameter models. Thirdly, graph theory provides a convenient language to discuss what exactly is meant by regularities or symmetries in the spatial configuration of the observational units and what their consequences on the properties of the models are.

The rest of the paper is organized as follows. Section 2 is devoted to the study of the correlation structure of the one-parameter models, denoted by CAR(1) and SAR(1) models. The focus here is on describing properties of the models that should be of interest to practitioners. This is done not only by deriving formal results, but also by discussing less formally some general characteristics of the models, and by examining some examples. We will find it convenient to first analyze CAR(1)

[1] For instance, spatial autoregressions are exponential families with low-dimensional sufficient statistic (see below), they have a clear conditional independence structure, and they constitute a relatively simple framework for inference.

models, and then extend the results to SAR(1) models. In Section 3 we first consider more briefly possible extensions to multi-parameter models, which are less popular in econometric applications, and then we discuss some issues concerning the parameterization of the models. In Section 4 the special features of spatial autoregressive models on graphs exhibiting some regularities or symmetries are discussed. In particular, we give conditions on the structure of the underlying graphs for two pairs of variables to be equally correlated. This leads, among other things, to the statement of the restrictive conditions required for the models to be homoskedastic or stationary (in a sense to be defined). Section 5 concludes. All proofs are relegated to the Appendix.

2 One-Parameter Models

In this section we study the correlation properties of one-parameter, or first-order, spatial autoregressions on arbitrary configurations of observational units. We first introduce the models. Then, in Section 2.1 we review some graph theoretic notions that are necessary for our analysis. Such notions are used in Sections 2.2 and 2.3 to study the correlation properties of, respectively, CAR(1) and SAR(1) models.

Consider a *fixed* and *finite* set of n observational units, for instance the set of regions of a country. For convenience, we fix an arbitrary ordering of the units, i.e. an arbitrary labelling of the n units by the first n positive integers. CAR(1) and SAR(1) models are families of distributions for a vector $y = (y_1, ..., y_n)' \in \mathbb{R}^n$, where y_i is the random variable associated to the i-th observational unit. We consider only zero-mean models, because our focus is on the correlation structure of the models.

Both models are specified on the basis of an $n \times n$ (spatial) *weights matrix* $\overline{W}$ (later, we will remove the bar above W to denote a weights matrix in a normalized model). The matrix $\overline{W}$ is a fixed matrix chosen to reflect a priori information on relations among the n observations (see, e.g., Cressie, 1993). For instance, the entries $\overline{W}_{i,j}$ may be taken to be a certain function of some distance, deemed to be relevant for the phenomenon under study, between the i-th and the j-th observational units. Usually a weights matrix is sparse, which, in the example just mentioned, means that $\overline{W}_{i,j} = 0$ if the distance between i and j is larger than some threshold.

Let L be a known $n \times n$ diagonal matrix with positive diagonal entries. A CAR(1) model is specified through the n conditional distributions

$$y_i \mid \{y_j : j \neq i\} \sim N\left(\rho \sum_{j=1}^{n} \overline{W}_{i,j} y_j, \tau^2 L_{i,i}\right), \quad i = 1, 2, ..., n, \tag{1}$$

where $\tau^2 > 0$ and ρ are functionally independent unknown parameters. Provided that $(I - \rho\overline{W})^{-1}L$ is symmetric and positive definite (hereafter abbreviated to p.d.), the n conditional distributions in (1) yield the joint distribution

$$y \sim N_n\left(0, \tau^2 (I - \rho\overline{W})^{-1} L\right) \tag{2}$$

(see Besag, 1974). Here and throughout $N_v(\cdot,\cdot)$ denotes the v-variate Gaussian distribution, the subscript being dropped in the univariate case.

In a SAR(1) model the distributions of the random variables in y are specified simultaneously, rather than conditionally, through the stochastic equation

$$y = \rho\overline{W}y + \varepsilon,$$

where $\varepsilon \sim N_n(0, \sigma^2 V)$, V being a known $n \times n$ diagonal matrix with positive diagonal entries, and $\tau^2 > 0$ and ρ are functionally independent unknown parameters. Provided that $I - \rho\overline{W}$ is invertible, the resulting joint distribution is

$$y \sim N_n\left(0, \sigma^2 (I - \rho\overline{W})^{-1} V (I - \rho\overline{W}')^{-1}\right). \tag{3}$$

The same notation is used for ρ and $\overline{W}$ in CAR(1) and SAR(1) models only for convenience. Note that SAR models can be readily extended to any non-Gaussian distribution with finite second moments. This is not the case for CAR(1) models, because of the compatibility conditions that n univariate conditional distributions must satisfy in order to produce a valid n-variate joint distribution (see Besag, 1974).

In this paper, a weights matrix is assumed to satisfy the following conditions:

(i) $\overline{W}_{i,i} = 0$, for $i = 1, ..., n$;

(ii) $\overline{W}_{i,j} = 0$ if and only if $\overline{W}_{j,i} = 0$, for $i, j = 1, ..., n$;

(iii) $\overline{W}$ is irreducible;

(iv) $\overline{W}$ is (entrywise) nonnegative.

Assumptions (i) and (ii) are necessary for specification (1) to be valid (observe that (ii) is implied by the stronger requirement that $(I-\rho\overline{W})^{-1}L$ is symmetric). As it is usually the case in applications, they are also maintained for SAR(1) models. Note that the symmetry of the zero entries of $\overline{W}$ entails that time series (unilateral) autoregressive models are not in the class of SAR models considered here. Also, assumption (ii) implies that, for the purpose of studying the correlation structure of models (2) and (3), $\overline{W}$ can be assumed to be *irreducible* (see, e.g., Gantmacher, 1974) without loss of generality. This is because if $\overline{W}$ were symmetric and reducible, then there would exist a permutation of its index set bringing it, and hence the covariance matrices of a CAR(1) or SAR(1) model, to block diagonal form, with the consequence that the models could be decomposed into the product of at least two models. Assumption (iv) is not required by the definition of the models, but is virtually always satisfied in empirical applications of spatial autoregressions and has the important theoretical advantage of making the *Perron-Frobenius theorem* for nonnegative irreducible matrices (e.g., Gantmacher, 1974) available to derive information about the spectral properties of weights matrices.

Recall that our objective is to study the correlation structure of the models. Since the correlation matrix of a multivariate normal vector y is invariant to transformations $y \to Ty$, where T a diagonal matrix with positive diagonal entries, it follows that there is no loss of generality in assuming

$L = V = I$ (just take $T = L^{-1/2}$ for a CAR(1) model and $T = V^{-1/2}$ for a SAR(1) model). This is indeed convenient for our analysis, and therefore, from now on and unless otherwise specified, by CAR(1) model we mean the family of distributions

$$N_n\left(0, \tau^2 \left(I - \rho W\right)^{-1}\right), \tag{4}$$

for some W (linked to the $\overline{W}$ in (2) through the relation $W = L^{-1/2}\overline{W}L^{1/2}$), and by SAR(1) model we mean the family

$$N_n\left(0, \sigma^2[\left(I - \rho W'\right)\left(I - \rho W\right)]^{-1}\right), \tag{5}$$

for some W (linked to the $\overline{W}$ in (3) through the relation $W = V^{-1/2}\overline{W}V^{1/2}$). Obviously, the matrices W must satisfy the same assumptions (i)-(iv) as the matrices $\overline{W}$. Note that expression (4) implies that, for the purpose of studying the correlation structure of a CAR(1) model, W can always be assumed to be symmetric. The fact that this is not the case for a SAR(1) model will have important bearings on our analysis.

2.1 Graph Theoretic Notions

In order to study the properties of a CAR(1) or SAR(1) model, it is natural to consider the graph with adjacency matrix W. This is the graph having as vertices the observational units, as edges the pairs (i, j) such that $W_{i,j} > 0$, and with each edge (i, j) weighted with the entry $W_{i,j}$. Often, we will find it more convenient to work with the graph with adjacency matrix ρW, which is the same graph as above, except for a rescaling by ρ of the weights of all edges. For graph theoretic details, we refer to Cvetković *et al.* (1980).

Assumptions (i)-(iv) on W have implications on the type of graphs considered in this paper. More specifically, and according to standard terminology, assumption (i) implies that such graphs do not contain *loops*; assumption (ii) that they are *undirected* (a graph is said to be directed if its edges are defined as ordered pairs of vertices, undirected if they are defined as unordered pairs of vertices); assumption (iii) that they are connected (Cvetković *et al.*, 1980, p. 18), as long as $\rho \neq 0$.

If they form an edge, two vertices of a graph are called *neighbors*. The *degree* of a vertex i is the number of neighbors of i and is denoted by δ_i. A graph is said to be *degree-regular* if all its vertices have the same degree.

A *walk* of length r, or r-walk, from i to j is a sequence of vertices $(i, l_1, ..., l_{r-1}, j)$ such that two consecutive vertices in the sequence form an edge. A walk is said to be *closed* if $i = j$. A *path* is a walk in which all the vertices are distinct. A *cycle* is a closed walk with all the vertices distinct apart from the first and the last.

Given the notion of path, it is natural to define the *distance* $d(i, j)$ between any two vertices i and j of a graph as the length of the shortest path joining i and j. Note that, since our graphs are undirected, $d(i, j) = d(j, i)$, $i, j = 1, ..., n$. The largest distance between any two vertices of a graph

is called the *diameter* of the graph, to be denoted by ζ. The *distance matrices* A_h, for $h = 0, 1, ..., \zeta$, are the $n \times n$ matrices with entry $(A_h)_{i,j}$ equal to 1 if $d(i,j) = h$, to 0 otherwise, for each pair of vertices i, j. Note that the distance matrices are disjoint symmetric $(0,1)$ matrices summing to a matrix of all ones, that $A_0 = I$ and that A_1—which we will usually denote simply by A—is the binary version of the adjacency matrix W. In the special case of an *unweighted* graph, i.e., a graph having all weights 0 or 1, $W = A$.

A graph theoretic notion that will be crucial in interpreting CAR and SAR models, is that of weight of a walk.

Definition 2.1 *The weight of a walk $(i, l_1, ..., l_{r-1}, j)$ in a graph with adjacency matrix W is the product $W_{i,l_1} W_{l_1,l_2} ... W_{l_{r-1},j}$ of the weights of its steps.*

Note that the effect of considering ρW, rather than W, as adjacency matrix is simply to multiply the weight of an r-walk by ρ^r. Clearly, one could define the weight of a walk differently (e.g., as the sum, rather than the product, of the weights of its steps), but we shall see that Definition 2.1 is the appropriate one for our purposes, because it implies a simple interpretation of the entries of powers of an adjacency matrix W. More specifically, according to Definition 2.1, the entry

$$(W^r)_{i,j} = \sum_{l_1,...,l_{r-1}=1}^{n} W_{i,l_1} W_{l_2,l_3} ... W_{l_{r-1},j}$$

represents the sum of the weights of all the r-walks from i to j. Note that if the graph is unweighted (i.e., $W = A$), then $(W^r)_{i,j}$ equals the number of r-walks from i to j.

The notion of bipartite graph (Cvetković *et al.*, 1980, p. 15) will also be critical to the understanding of some aspects of our analysis of spatial autoregressions.

Definition 2.2 *A graph is said to be bipartite if its vertex set can be partitioned into two non-empty disjoint sets V_1 and V_2 such that every edge of the graph joins one vertex in V_1 with one vertex in V_2.*

A necessary and sufficient condition that is useful to visually check for bipartiteness is the absence of cycles of odd length. This suggests that "regular" graphs such as those of a rectangular lattice (in any dimension) are bipartite, whereas "irregular" graphs such as those of geographical maps of regions are generally not bipartite. Another necessary and sufficient condition for bipartiteness is given in the next lemma, which follows immediately from Theorem 3.4 of Cvetković *et al.* (1980). We denote by $\lambda_{\max}$ the spectral radius (i.e., the largest modulus of the eigenvalues) of W. By the Perron-Frobenius theorem, $\lambda_{\max}$ is a (algebraically and geometrically) simple eigenvalue of W.

Lemma 2.3 *A graph with adjacency matrix W is bipartite if and only if $-\lambda_{\max}$ is an eigenvalue of W.*

2.2 CAR(1) Model

In this section we study correlation properties of CAR(1) models on general graphs. Section 2.2.1 derives a graph theoretic interpretation of the covariances and correlations implied by a CAR(1) model, and sets the scene for the subsequent analysis. The graph theoretic interpretation follows almost straightforwardly from associating to a CAR(1) model the graph with adjacency matrix ρW, and is useful both to provide an explanation of some known properties of the models and to derive some new properties. In Section 2.2.2 we study properties of CAR(1) models that are guaranteed to hold exactly for any W only as $\rho \to 0$. For a fixed W, such properties generally hold, at least approximately, in large neighborhoods of $\rho = 0$. Section 2.2.3 discusses some differences in the correlation structure of CAR(1) models between the case $\rho > 0$ and the case $\rho < 0$.

The covariance matrix $\tau^2(I - \rho W)^{-1}$ of a CAR(1) model is denoted by $\Sigma(\rho)$, and the corresponding correlation matrix by $\Sigma^*(\rho)$. For convenience, we fix $\tau^2 = 1$. Recall from above that, in a CAR(1) model, W is assumed to be symmetric. We denote the (real) eigenvalues of W by $\lambda_1, ..., \lambda_n$, labelled in non-decreasing order of magnitude. Note that $\lambda_1 < 0$ (because $tr(W) = \sum_{i=1}^n \lambda_i = 0$, and at least one eigenvalue does not vanish, since W is irreducible) and that $\lambda_n = \lambda_{\max} \geq |\lambda_1|$ by the Perron-Frobenius theorem, with equality if and only if the underlying graph is bipartite by Lemma 2.3. It follows that, in order for $\Sigma(\rho)$ to be p.d., it is necessary that $\lambda_1^{-1} < \rho < \lambda_{\max}^{-1}$. Of course, if desired, one can always reparametrize the model so that the right (resp. left) extreme of the parameter space is 1, by rescaling W by $\lambda_{\max}^{-1}$ (resp. λ_1^{-1}).

We will pay particular attention to the two most popular versions of CAR(1) models used in applications. Given the $(0-1)$ first distance matrix A of a graph, these are constructed by taking $W = A$ and $W = D^{-1/2}AD^{-1/2}$, where D is a diagonal matrix containing the row sums of A, i.e. $D(i,i) = \delta_i$, $i = 1, ..., n$. Observe that, before normalization to $L = I$, the latter specification corresponds to $\overline{W} = D^{-1}A$, a so-called *row-standardized weights matrix*, and $L = D^{-1}$. The models with $W = A$ and with $W = D^{-1/2}AD^{-1/2}$ are equivalent (up to a reparametrization) if and only if the underlying graph is degree-regular (i.e., D is a scalar multiple of I).

2.2.1 Graph Theoretic Interpretation

When $|\rho| < \lambda_{\max}^{-1}$, the covariance matrix of a CAR(1) model can be represented as $\Sigma(\rho) = \sum_{r=0}^{\infty} (\rho W)^r$, with $W^0 = I$ (e.g., Horn and Johnson, 1985, p. 301). Thus, for any $i, j = 1, ..., n$,

$$\Sigma_{i,j}(\rho) = \sum_{r=d(i,j)}^{\infty} \rho^r (W^r)_{i,j}, \tag{6}$$

where we have used the fact that $(W^r)_{i,j} = 0$ if $r < d(i,j)$, which follows immediately from the interpretation of powers of adjacency matrices (see Section 2.1). Based on the same interpretation, representation (6) asserts that the covariance $\Sigma_{i,j}(\rho)$ is equal to the total weight of all walks between i and j in the graph underlying a CAR(1) model, or, equivalently:

Property 2.4 *For $|\rho| < \lambda_{\max}^{-1}$, any walk from i to j in the graph with adjacency matrix ρW contributes its weight to $\Sigma_{i,j}(\rho)$.*

We stress that all walks from i to j, and not only the paths or some other sequences of vertices, contribute to $\Sigma_{i,j}(\rho)$. This makes clear that the covariance between any two variables y_i and y_j in a CAR(1) model is determined by the global connectivity properties of the graph with adjacency matrix ρW, and not only by the properties of the graph in some neighborhood of i and j. Observe that the inequality $|\rho| < \lambda_{\max}^{-1}$ defines a (proper) subset of the interval $(\lambda_1^{-1}, \lambda_{\max}^{-1})$ where $\Sigma(\rho)$ is p.d., unless the graph underlying W is bipartite, by Lemma 2.3 and the fact that $|\lambda_1| \leq \lambda_{\max}$.

The interpretation of the covariance structure of CAR(1) models suggested by Property 2.4 is particularly simple when $W = A$ or $W = D^{-1/2}AD^{-1/2}$, because of the simple forms of the weights of walks implied by such specifications. When $W = A$, any r-walk has weight ρ^r, and hence its contribution to $\Sigma_{i,j}(\rho)$ is strictly decreasing with r, for any fixed and non-zero $|\rho| < \lambda_{\max}^{-1}$.[2] For other specifications of W, the weight of an r-walk depends not only on r and ρ, but also on the vertices that the walk visits. For instance, when $W = D^{-1/2}AD^{-1/2}$ a walk $(i, l_1, ..., l_{r-1}, j)$ has weight $\rho^r/(\delta_i^{1/2}\delta_{l_1}...\delta_{l_{r-1}}\delta_j^{1/2})$, and hence, for fixed and non-zero r and ρ, its contribution to $\Sigma_{i,j}(\rho)$ is large if the vertices visited by the walk have a small number of neighbors.[3] To further clarify the distinction between a model with $W = A$ and a model with $W = D^{-1/2}AD^{-1/2}$, consider the contribution of an edge (i.e. a 1-walk) (i, j) to the covariance between y_i and y_j. Such a contribution is the same for each pair of neighbors i, j when $W = A$ (namely equal to ρ), whereas it is inversely related to δ_i and δ_j when $W = D^{-1/2}AD^{-1/2}$ (namely equal to $\rho/(\delta_i\delta_j)^{1/2}$). This suggests that the latter specification may be particularly appropriate for applications where spatial autocorrelation is determined by a limited resource, such as, for example, transportation amongst geographical units.

Remark 2.5 It is easily seen that, for any W, the powers $W^0, ..., W^{s-1}$, where s denotes the number of distinct eigenvalues of W, are linearly independent. Hence, for any W and any $i, j = 1, ..., n$, representation (6) can be rewritten as

$$\Sigma_{i,j}(\rho) = \sum_{r=d(i,j)}^{s-1} f_{W,r}(\rho)(W^r)_{i,j}, \tag{7}$$

for some functions $f_{W,0}(\rho), ..., f_{W,s-1}(\rho)$ from $(-\lambda_{\max}^{-1}, \lambda_{\max}^{-1})$ to $\mathbb{R}$ (details are omitted for brevity, but these are rational functions and can be derived straightforwardly for any W). Representation (7) asserts that $\Sigma_{i,j}(\rho)$ can be interpreted in terms of contributions coming only from the walks between i and j of length between $d(i, j)$ and $s - 1$. The maximum length $s - 1$ that needs to be considered is at least as large as the diameter ξ of the graph (this follows from a simple extension of a classic result for unweighted graphs, e.g., Theorem 3.13 of Cvetković *et al.*, 1980). While (7) is

[2] This is because $\lambda_{\max} \geq \min_{i=1,...,n} \sum_{j=1}^n A_{i,j} \geq 1$. For the first inequality see, e.g., Gantmacher (1974), p. 63; the second one is an obvious consequence of the symmetry and irreducibility of W.

[3] Note that $D^{-1/2}AD^{-1/2}$ is similar to the row-stochastic matrix $D^{-1}A$, and hence has $\lambda_{\max} = 1$. Thus the contribution $\rho^r/(\delta_i^{1/2}\delta_{i_1}...\delta_{i_{r-1}}\delta_j^{1/2})$ of an r-walk tends to decrease as r increases (for any $|\rho| < \lambda_{\max}^{-1} = 1$).

very interesting from the point of view of interpreting the covariances implied by CAR(1) models (because it shows that one does not need to consider an infinite number of walks), in the rest of the section we find it more convenient to work with the less parsimonious (6), because in that case the coefficients in front of the entries $(W^r)_{i,j}$ are independent of W.

So far, our discussion has been in terms of the covariances $\Sigma_{i,j}(\rho)$. However, since CAR(1) models are well known to be generally heteroskedastic (cf. Proposition 4.4 below), the correlations $\Sigma^*_{i,j}(\rho)$ provide a more appropriate measure of the strength of association between pairs of variables. Before moving to analyze explicitly the correlations $\Sigma^*_{i,j}(\rho)$, it is worth noting that the normalization $L = I$ (see expression (4), and recall that such a normalization does not affect the correlation structure of the models) entails that the models are approximately homoskedastic when ρ is close to 0, and hence the simple properties and interpretations of the covariances $\Sigma_{i,j}(\rho)$ deduced from (6) can be extended, at least approximately, to the correlations $\Sigma^*_{i,j}(\rho)$ when $|\rho|$ is sufficiently small.

The analog of representation (6) for $\Sigma^*_{i,j}(\rho)$ is given in the following lemma.

Lemma 2.6 *For $|\rho| < \lambda_{\max}^{-1}$ and any $i, j = 1, ..., n$,*

$$\Sigma^*_{i,j}(\rho) = \sum_{r=d(i,j)}^{\infty} [(W^r)_{i,j} - c_r(i,j)]\rho^r, \tag{8}$$

with

$$c_r(i,j) = \sum_{s=2}^{r} (W^{r-s})_{i,j} \sum \frac{\alpha_k}{k_2!...k_s!} \prod_{t=2}^{s} \left(\sum_{u=0}^{t} (W^{t-u})_{i,i} (W^u)_{j,j} \right)^{k_t}, \tag{9}$$

where $k = \sum_{t=2}^{s} k_t$, $\alpha_k = \prod_{l=0}^{k-1} \left(-\frac{1}{2} - l\right)$, and the second summation in (9) is over all $(s-1)$-tuples of non-negative integers $(k_2, ..., k_s)$ such that $\sum_{t=2}^{s} t k_t = s$.

Although the expression for $\Sigma^*_{i,j}(\rho)$ resulting from Lemma 2.6 is rather involved, some interesting properties of the correlation structure of a CAR(1) model may be deduced very simply from it. Observe that every term on the right hand side of (9) contains one entry $(W^{r-s})_{i,j}$, for some $s = 2, ..., r$. Since $(W^q)_{i,j} = 0$ for any nonnegative integer $q < d(i,j)$, it follows that $c_r(i,j) = 0$ if $r \leq d(i,j) + 1$. Hence, from (8) we obtain that, for any $i, j = 1, ..., n$,

$$\Sigma^*_{i,j}(\rho) = (W^{d(i,j)})_{i,j}\rho^{d(i,j)} + (W^{d(i,j)+1})_{i,j}\rho^{d(i,j)+1} + O(\rho^{d(i,j)+2}), \tag{10}$$

or, equivalently, $\frac{d^r}{d\rho^r} \Sigma^*_{i,j}(\rho)\big|_{\rho=0} = r!(W^r)_{i,j}$, for $r \leq d(i,j)+1$. Expression (10) makes clear that the correlations $\Sigma^*_{i,j}(\rho)$ of a CAR(1) model are determined by the properties of the underlying graph in a way that is similar to, but more complicated than, the way the covariances $\Sigma_{i,j}(\rho)$ are determined (see representation (6)), especially for small ρ. In particular, it is worth emphasizing the following interpretation of the entries of a weights matrix.

Property 2.7 *In a CAR(1) model $W_{i,j}$ represents the first derivative of $\Sigma^*_{i,j}(\rho)$ at $\rho = 0$, for any $i, j = 1, ..., n$.*

Property 2.7 asserts that, when fixing an entry $W_{i,j}$, the user of a CAR(1) model is controlling directly $\frac{d}{d\rho}\, \Sigma^*_{i,j}(\rho)\big|_{\rho=0}$, and hence the behavior of $\Sigma^*_{i,j}(\rho)$ for $|\rho|$ sufficiently close to 0. As discussed above, for larger $|\rho|$ $\Sigma^*_{i,j}(\rho)$ is determined not only by $W_{i,j}$ but by the whole structure of W. In particular, if the modeler sets $W_{i,j} > W_{l,m}$ for two pairs of units i, j and l, m, then this implies $\left|\Sigma^*_{i,j}(\rho)\right| > \left|\Sigma^*_{l,m}(\rho)\right|$ for $|\rho|$ small enough, but not necessarily over the whole parameter space of the model (both when $W_{l,m} > 0$ and when $W_{l,m} = 0$). We will return to this point in Section 2.2.2.

The effect of the term $c_r(i, j)$, which constitutes the difference between representations (6) and (8), can also be deduced from Lemma 2.6. On expanding multinomially all terms with exponents k_t in (9), it becomes clear that, for any W and any $i, j = 1, ..., n$, $c_r(i, j)$ can be expressed as $\sum_{l_1,...,l_{r-1}=1}^{n} c(i, l_1, ..., l_{r-1}, j)$, for some coefficients $c(i, l_1, ..., l_{r-1}, j)$ that do not vanish if and only if $(i, l_1, ..., l_{r-1}, j)$ is a walk (in the graph with adjacency matrix ρW) with at least one repetition of i or j. Recalling Property 2.4, representation (8) then implies that, for $|\rho| < \lambda^{-1}_{\max}$, any walk from i to j that does not contain repetitions of i or j contributes its weight to $\Sigma^*_{i,j}(\rho)$, whereas any walk from i to j that contains at least a repetition of i or j gives a contribution that is smaller than its weight. We refrain from analyzing the implications of expression (9) further, because the interpretation of $\Sigma_{i,j}(\rho)$ and $\Sigma^*_{i,j}(\rho)$ developed so far is all we need for the analysis in the following sections.

We now illustrate how the adoption of a graph theoretic perspective contributes to the understanding of the correlation structure of CAR(1) models, by referring to two well-known properties, or "peculiarities", of the models.

Property 2.8 (e.g., Besag and Kooperberg, 1995, p. 735) *Two correlations $\Sigma^*_{i,j}(\rho)$ and $\Sigma^*_{l,m}(\rho)$, for two distinct pairs of neighbors (i, j) and (l, m), are generally different functions of ρ.*

Property 2.9 (e.g., Wall, 2004, p. 320-1) *In a CAR(1) model, the ranking of pairs of neighbors in terms of their degree of correlation may vary across the parameter space.*

These (and other) properties of spatial autoregressions have been regarded as undesirable or counterintuitive in recent literature (for instance in Besag and Kooperberg, 1995, and in Wall, 2004). Our perspective here is that they are natural consequences of the assumption that covariances are determined by weighting all walks between pairs of vertices on a graph. We believe that such an assumption might not be unrealistic in some applications. For the sake of simplicity, the following discussion of Properties 2.8 and 2.9 is based on representations (6), but it is clear that, by a slight modification of the same arguments, it could also be based on (8).

Starting from Property 2.8, observe that, according to representation (6), $\Sigma_{i,j}(\rho) = \Sigma_{l,m}(\rho)$ for any $|\rho| < \lambda^{-1}_{\max}$ if and only if the total weight of the r-walks from i to j is the same as the total weight of the r-walks from l to m, for each $r = 1, 2, ...$ Given any two pairs of neighbors i, j and

l, m, such a condition is not met on general graphs, firstly because different pairs of neighbors are generally not linked by the same number of r-walks, for any $r = 1, 2, ...$, (unless strong regularities are imposed on the graph; see Section 4), and secondly because the weights of r-walks between different pairs of neighbors are generally different (because they depend on the vertices visited by the r-walk, except for the simple weight function implied by $W = A$).

Moving to Property 2.9, this asserts that two correlations $\Sigma^*_{i,j}(\rho)$ and $\Sigma^*_{l,m}(\rho)$, regarded as functions of ρ and for two fixed pairs of neighbors i, j and l, m, may intersect. Using representation (6), write $\Sigma_{i,j}(\rho) - \Sigma_{l,m}(\rho)$ as $\sum_{r=1}^{\infty}[(W^r)_{i,j} - (W^r)_{l,m}]\rho^r$. Then, it is clear that $\Sigma_{i,j}(\rho)$ and $\Sigma_{l,m}(\rho)$ may intersect for some ρ because the terms $(W^r)_{i,j} - (W^r)_{l,m}$ need not have the same sign for each r. When $W = A$, this simply amounts to saying that there may be more walks of a certain length between i and j than between l and m, but less walks of another length. It should be noted that Property 2.9, as Property 2.8, holds not only for neighbors but also for units at graph distance larger than one.

2.2.2 Properties When ρ is Close to 0

In this section we discuss some correlation properties of CAR(1) models that are guaranteed to hold for any W, provided that ρ is sufficiently small. We also investigate in some detail how the specification of W affects such properties when ρ is not necessarily small.

The following result makes clear how the correlations between variables observed at pairs of neighbors depend on the local characteristics of the underlying graph, when ρ is close to 0, and when a $(0-1)$ weights matrix ($\overline{W} = A$, and hence $W = A$) or its row-standardized version ($\overline{W} = D^{-1}A$, and hence $W = D^{-1/2}AD^{-1/2}$) are used. Recall that δ_i denotes the number of neighbors of i, and that the correlations $\Sigma^*_{i,j}(\rho)$ are invariant to the normalization leading from $\overline{W}$ to W.

Proposition 2.10 *Let i, j and l, m be any two pairs of neighbors. For any $W = A$ in a CAR(1) model, there exists a real interval (a, b) containing 0 such that, for any $\rho \in (a, b)$, if the number of common neighbors of i and j is not less than the number of common neighbors of l and m, then $\left|\Sigma^*_{i,j}(\rho)\right| \geq \left|\Sigma^*_{l,m}(\rho)\right|$. For any $W = D^{-1/2}AD^{-1/2}$ in a CAR(1) model, there exists a real interval (a, b) containing 0 such that, for any $\rho \in (a, b)$, if $\delta_i\delta_j \leq \delta_l\delta_m$, then $\left|\Sigma^*_{i,j}(\rho)\right| \geq \left|\Sigma^*_{l,m}(\rho)\right|$.*

Remark 2.11 The interval (a, b) in Proposition 2.10 depends on W and on i, j, l, m. In particular, it strongly depends on the degree of regularity of W (see Section 4). For instance, $(a, b) = (\lambda_1^{-1}, \lambda_n^{-1})$, for any i, j, l, m, if the graph underlying a CAR(1) model is distance-regular (because in such a case $\Sigma^*_{i,j}(\rho)$ depends on i, j only through $d(i, j)$, see Theorem 4.7 below). Conversely, (a, b) can be very small when a CAR(1) model is defined on a very irregular graph.

In general, on typical planar graphs,[4] both the number of common neighbors of i and j and the

[4] A graph is said to be planar if it can be drawn on a plane without edges crossing. For instance, the graphs of geographical maps with two units being neighbors if they share a boundary are planar.

product $\delta_i\delta_j$ are large for pairs of neighbors (i,j) in the central region of the graph, and small for pairs of neighbors close to the borders of the graph. Taking the previous sentence as our definition of "typical planar graphs", Proposition 2.10 suggests the following fundamental difference between using a $(0-1)$ weights matrix or its row-standardized version.

Property 2.12 *For CAR(1) models defined on typical planar graphs, and for $|\rho|$ sufficiently small, if $W = A$ (resp. $W = D^{-1/2}AD^{-1/2}$), then pairs of neighbors in the central region of the graph tend to be the most (resp. least) correlated pairs of neighbors, whereas pairs of neighbors close to the borders of the graph tend to be the least (resp. most) correlated.*

In fact, Property 2.12 generally holds over most of the interval $(0, \lambda_{\max}^{-1})$ (the case $\rho < 0$ is more delicate, as we will see below). We illustrate by referring to the map of the US, also used in Wall (2004) (see in particular her Figure 5).

Example 2.13 Consider the graph having as vertices the 48 continental US states, and let two states be neighbors if and only if they share a common boundary (or a common corner). Figure 1(b) shows that when $W = A$ and unless ρ is very close to $\lambda_{\max}^{-1}$ ($\lambda_{\max}^{-1} \cong .185$), Maine and New Hampshire are the most correlated contiguous states, whereas Missouri and Tennessee are the least correlated contiguous states. (The non-smooth behavior of the correlation between variables observed at Maine and New Hampshire will be explained in Remark 2.18.) On the contrary, Figure 1(c) shows that when $W = D^{-1/2}AD^{-1/2}$ is used, Missouri and Tennessee is one of the most correlated pairs of neighbors, whereas Maine and New Hampshire is the least correlated pair of neighbors.

Figure 1 about here

We now turn to analyze the behavior of $\Sigma^*_{i,j}(\rho)$, for fixed ρ, as the distance $d(i,j)$ changes. The next result establishes that, when ρ is sufficiently close to 0, $\Sigma^*_{i,j}(\rho)$ is decreasing (in absolute value) with $d(i,j)$.

Proposition 2.14 *In a CAR(1) model, for any W and for any $i,j,l,m = 1,...,n$, there exists a real interval (a,b) containing 0 such that, for any $\rho \in (a,b)$, if $d(i,j) < d(l,m)$, then $\left|\Sigma^*_{i,j}(\rho)\right| > \left|\Sigma^*_{l,m}(\rho)\right|$.*

As Proposition 2.10, Proposition 2.14 need not hold over the whole parameter space of a CAR(1) model. That is, for $|\rho|$ sufficiently large, pairs of variables observed at two units l,m may be more correlated than variables observed at two units i,j that are closer together according to the graph distance $d(\cdot,\cdot)$. The reason for such a (at first sight counterintuitive) behavior lies again in the fact that $\Sigma^*_{i,j}(\rho)$ is not determined by $d(i,j)$, but by the total weight of the walks from i to j. More specifically, for some $r \geq d(l,m)$, the total weight $\rho^r(W^r)_{i,j}$ of the r-walks between i and j need not be larger than the total weight $\rho^r(W^r)_{l,m}$ of the r-walks between l and m. This is particularly

clear when $W = A$ (in which case the entries of W^r represent the number of r-walks between two vertices), because on a given graph and for units i, j, l, m such that $d(i, j) < d(l, m)$, there may be less r-walks between i and j than between l and m.

The above argument also suggests that for a fixed matrix A, a CAR(1) model with $W = D^{-1/2}AD^{-1/2}$ is closer to satisfying, at least approximately, the property in Proposition 2.14 in large intervals (a, b) than a CAR(1) model with $W = A$. This is because the weight of an r-walk implied by the specification $W = D^{-1/2}AD^{-1/2}$ is inversely related to the product of the degrees of the vertices visited (see Section 2.2.1), and hence tends to decay quicker with r than the weight ρ^r of an r-walk implied by $W = A$. To illustrate, let us go back to the example of the US map. In Figure 2(b) we display the correlations implied by a CAR(1) model with $W = A$ and when $\rho \geq 0$, for a pair of neighbors (dark line; Maine and New Hampshire) and for a pair of non-neighbors (light line; Oklahoma and Nebraska). Observe that for large ρ the correlation between the two non-neighbors can be much larger than the correlation between the two neighbors. This behavior is attenuated, but not eliminated, when $W = D^{-1/2}AD^{-1/2}$. For such a choice of W, Figure 2(d) shows one case in which, for some positive values of ρ, the correlation between a pair of non-neighbors (light line; Vermont and Connecticut) can be larger than the correlation between a pair of neighbors (dark line; Missouri and Tennessee). Note that, for any $\rho > 0$, the correlation between non-neighbors is never much larger than the correlation between neighbors; when $W = D^{-1/2}AD^{-1/2}$ on the US map, this is true over all pairs of contiguous and non-contiguous states.

Figure 2 about here

The fact that non-neighbors may be more correlated than neighbors should not necessarily be regarded as a drawback of CAR(1) models. This is particularly true when the models are defined on irregular lattices, as we next argue.

In the case of *regular lattices*, it is usually convenient to interpret a CAR(1) model for a finite-dimensional vector y as a restriction of a process defined on an infinite lattice. In such a case it would certainly be desirable that the finite model satisfies any kind of regularity or stationarity assumed on the infinite process. This requires edge corrections, which, in general, are difficult to perform (see Besag and Kooperberg, 1995, Section 2.3) and involve the loss of the original Markov property (e.g., Champagnat, Idier and Goussard, 1998). "Edge effects" definitely represent drawbacks of spatial autoregressive models.

Conversely, when the models are defined on *irregular lattices*, it is hard to see why one would want to impose that the correlations between neighbors are non-increasing (in absolute value) in the graph distance. In the case of the US map, for example, if two non-contiguous states share more low-order neighbors than two contiguous states, then it may be appropriate to allow for the correlation between the two non-contiguous states to be greater than the correlation between the two contiguous states. As discussed above, this is precisely what CAR models do, because the

correlation between two variables in a CAR model is determined by weighting all the walks between the two corresponding observational units (for instance, the number of neighbors shared by two units is equal to the number of 2-walks between the two units).

To summarize the discussion prompted by Proposition 2.14, the user of CAR(1) models should be aware that setting $W_{i,j} > 0$ (i.e., imposing that i and j are neighbors) does not necessarily imply that y_i and y_j are more correlated than a pair y_l and y_m such that $W_{l,m} > 0$. If the modeler wants to impose that pair of neighbors that are close together (in some metric) are more correlated, then models that parametrize directly a covariance matrix are preferable to CAR(1) models.

2.2.3 Positive versus Negative Autocorrelation

This section discusses differences in the correlation structure of CAR(1) models between the case $\rho > 0$ and the case $\rho < 0$. We start from the following simple result.

Proposition 2.15 *In a CAR(1) model, for any W and any $i, j = 1, ..., n$, (a) if $\rho > 0$ then $\Sigma^*_{i,j}(\rho) > 0$; (b) there exists a left neighborhood of $\rho = 0$ where $\Sigma^*_{i,j}(\rho) < 0$ if $d(i,j)$ is odd and $\Sigma^*_{i,j}(\rho) > 0$ if $d(i,j)$ is even.*

When $\rho < 0$, Proposition 2.15 determines the sign of $\Sigma^*_{i,j}(\rho)$ only for small $|\rho|$. In fact, for a fixed W and for fixed i and j, $\Sigma^*_{i,j}(\rho)$ need not have the same sign over the interval $(\lambda_1^{-1}, 0)$.[5] This is because when $d(i,j)$ is odd (resp. even), the total positive contribution $\sum_{r=2,r\ even}^{\infty} \rho^r (W^r)_{i,j}$ of the walks of even lengths to $\Sigma_{i,j}(\rho)$ may, for some negative ρ, become larger (resp. smaller) than the total negative contribution $\sum_{r=2,r\ odd}^{\infty} \rho^r (W^r)_{i,j}$ of the walks of odd lengths.

The different behavior of the correlations when $\rho > 0$ and when $\rho < 0$ can be regarded as an intrinsic characteristic of spatial autoregressions on general graphs. This is in contrast with time series autoregressions, whose correlations functions are either odd or even functions of the autocorrelation parameter. Obviously, if the graph underlying a CAR(1) model is not bipartite, then a correlation $\Sigma^*_{i,j}(\rho)$ cannot be either an odd or an even function of ρ, because, by Lemma 2.3, it is defined on a non-symmetric interval. Conversely, when the graph is bipartite, we have the following neat result.

Proposition 2.16 *For any CAR(1) model on a bipartite graph, $\Sigma^*_{i,j}(\rho)$, $i, j = 1, ..., n$, is: (a) an odd function if $d(i,j)$ is odd; (b) an even function if $d(i,j)$ is even.*

The next proposition is concerned with the behavior of the correlations at the two extremes of the parameter space of a CAR(1) model. It is probably not surprising, at this stage of the analysis, that for a CAR(1) model on a non-bipartite graph, the limits of the autocorrelations as $\rho \to \lambda_{\max}^{-1}$ (from the left) and as $\rho \to \lambda_1^{-1}$ (from the right) can be very different. We denote by m_1 the multiplicity of λ_1, and by q_1 an eigenvector of W associated to λ_1. Unless the graph underlying a CAR(1) model satisfies particular symmetries, generally $m_1 = 1$ (see, for instance, Biggs, 1993).

[5] This property is also pointed out by Wall (2004), p. 321, where it is regarded as "counterintuitive".

Proposition 2.17 *In a CAR(1) model, for any W and any $i, j = 1, ..., n$, (a) as $\rho \to \lambda_{\max}^{-1}$, $\Sigma^*_{i,j}(\rho) \to 1$; (b) as $\rho \to \lambda_1^{-1}$,*

$$\Sigma^*_{i,j}(\rho) \to \begin{cases} 1 & \text{if } m_1 = 1 \text{ and } (q_1)_i (q_1)_j > 0, \\ -1 & \text{if } m_1 = 1 \text{ and } (q_1)_i (q_1)_j < 0, \\ 0 & \text{if } m_1 = 1 \text{ and either } (q_1)_i = 0 \text{ or } (q_1)_j = 0, \end{cases} \tag{11}$$

whereas $\Sigma^*_{i,j}(\rho)$ can go to any number in $[-1, 1]$ otherwise (i.e., if $m_1 = 1$ and $(q_1)_i = (q_1)_j = 0$, or if $m_1 > 1$).

Remark 2.18 Generally, the limits in Proposition 2.17 are informative about the behavior of the correlations in a rather large neighborhood of $\lambda_{\max}^{-1}$ and λ_1^{-1}, because the correlations go to such limits slowly. There are cases, however, in which this is not so. Consider, for instance, two neighbors i, j when W is a $(0-1)$ matrix A. By Property 2.7, the derivative of $\Sigma^*_{i,j}(\rho)$ at $\rho = 0$ is 1. If the number of short walks between i and j is very small (compared to number of longer walks), then the coefficients of ρ^r in (8) for small r are very small (compared to the coefficients for larger r). Since $\lambda_{\max}^{-1}$ is typically much smaller than 1 (in particular $\lambda_{\max}^{-1} \leq 1/\min_{i=1,..,n} \sum_{j=1}^n A_{i,j}$; Gantmacher, 1974, p. 63), it follows that $\Sigma^*_{i,j}(\rho)$ has to increase quickly close to $\lambda_{\max}^{-1}$ in order to approach the limit 1 as $\rho \to \lambda_{\max}^{-1}$. This is precisely what happens in Figure 1(b) for the correlations between Maine and New Hampshire. The same argument also suggests that a similar non-smooth behavior (i) can only occur when CAR(1) models are defined on very irregular graphs, (ii) is very unlikely to occur when $W = D^{-1/2} A D^{-1/2}$.

Remark 2.19 Proposition 2.17 (a) relies on the assumption of nonnegativity and irreducibility of W. Without such an assumption, $\Sigma^*_{i,j}(\rho)$ could go, as $\rho \to \lambda_{\max}^{-1}$, to any value in $[-1, 1]$.

Remark 2.20 Let E_i be the eigenspace of W associated to λ_i, for $i = 1, ..., n$. An alternative proof of Proposition 2.17 is worth recording, because it shows how the limiting correlations, as $\rho \to \lambda_{\max}^{-1}$ and as $\rho \to \lambda_1^{-1}$, are determined by the position of, respectively, E_n and E_1 in the sample space $\mathbb{R}^n$. Since, as ρ goes to $\lambda_{\max}^{-1}$ (resp. λ_1^{-1}), the precision matrix $I - \rho W$ of a CAR(1) model tends to a singular matrix, it is easily shown that CAR(1) models tend to improper distributions supported on E_n (resp. E_1). By the Perron-Frobenius theorem, E_n is the span of a vector with positive entries, which implies that, as $\rho \to \lambda_{\max}^{-1}$, all pairs of variables y_i and y_j are perfectly correlated. On the other hand as $\rho \to \lambda_1^{-1}$ a pair of variables y_i and y_j need not be perfectly autocorrelated, since their distribution is supported on a subspace, E_1, that may have dimension larger than 1, or may be 1-dimensional but spanned by a vector q_1 with some zero entries (and then the particular cases in (11) admit a straightforward geometric interpretation; e.g., if $(q_1)_i = 0$ then y_j must be uncorrelated with any variable y_j such that $(q_1)_j \neq 0$).

Remark 2.21 The combination of Propositions 2.15 and 2.17 implies that, when $m_1 = 1$, $\Sigma^*_{i,j}(\rho)$ changes sign over the interval $(0, \lambda_1^{-1})$ if $d(i,j)$ is odd (resp. even) and $(q_1)_i(q_1)_j > 0$ (resp. $(q_1)_i(q_1)_j < 0$).

Remark 2.22 When $W = D^{-1/2}AD^{-1/2}$ (which has $\lambda_{\max} = 1$, since it is similar to the row-stochastic matrix $D^{-1}A$), Proposition 2.17 (a) is a simple consequence of the celebrated Matrix-Tree-Theorem for the *Laplacian matrix* $D - A$ of an unweighted connected graph; Cvetković *et al.* (1980), p. 38. The limiting distributions with singular precision matrix $D - A$ are known in the literature as *intrinsic autoregressions*, and are often used as (improper) priors in Bayesian analysis; e.g., Besag and Kooperberg (1995).

We end this section with a comment on the interpretation of ρ as as a measure of autocorrelation in the joint distribution of y.[6] Obviously, when it is negative, in general ρ cannot be interpreted as an autocorrelation coefficient, because, as observed above, the correlations implied by a CAR(1) model need not be monotonic over the interval $(\lambda_1^{-1}, 0)$. Conversely, when $\rho > 0$ it appears that all correlations are always monotonic in ρ, and hence ρ can be interpreted as a measure of autocorrelation. This is indicated in the following conjecture, to which we have not been able to find either a counterexample or a proof.

Conjecture 2.23 *In a CAR(1) model, for any W and for any $i, j = 1, ..., n$, $\Sigma^*_{i,j}(\rho)$ is strictly increasing in ρ over the interval $(0, \lambda_{\max}^{-1})$.*

Note that, by representation (6), the conjecture holds trivially for the covariances $\Sigma_{i,j}(\rho)$, but due to the fact that the models are generally heteroskedastic, this is not very informative.

2.3 SAR(1) Model

The aim of this section is to show that essentially all of the properties of CAR(1) models discussed above extend to SAR(1) models. We will repeat the main steps of Section 2.2, presenting the necessary modifications.

For the purpose of studying their correlation structure, one important difference between CAR(1) models and SAR(1) models is that in the latter case we have to allow the possibility of nonsymmetric W (see Section 2). Setting for convenience $\sigma^2 = 1$, we denote the covariance matrix $[(I - \rho W')(I - \rho W)]^{-1}$ of a SAR(1) model by $\Gamma(\rho)$, and the corresponding correlation matrix by $\Gamma^*(\rho)$. For $\Gamma(\rho)$ to be p.d., ρ must be different from the reciprocal of the non-zero real eigenvalues of W. Usually, for various reasons, it is desirable to restrict the parameter space much further. For instance, one

[6] We stress that we are here referring to the *joint* distribution of y. The relation between ρ and an autocorrelation coefficient in the *conditional* distribution of two variables observed at two neighbors, given all remaining variables, is well-known and, because of the way CAR models are constructed, much neater. In fact, denoting by ρ^{ij} be the partial correlation coefficient between y_i and y_j, for two neighbors i and j, it is easily seen that $\rho = \rho^{ij}[W_{i,j}W_{j,i}]^{-1/2}$ (e.g., Cressie and Chan, 1989, eq. 3.14).

may want to work with a connected parameter space and hence impose $\rho < \lambda_{\max}^{-1}$, and, if W admits at least a (real) negative eigenvalue, $\rho > \lambda_{\min}^{-1}$, with $\lambda_{\min}$ denoting the minimum negative eigenvalue of W.[7]

Let $Z_r = \sum_{k=0}^{r} W^k (W')^{r-k}$. When $|\rho| < \lambda_{\max}^{-1}$, $\Gamma(\rho)$ admits the representation

$$\Gamma(\rho) = \sum_{r=0}^{\infty} \rho^r W^r \sum_{r=0}^{\infty} \rho^r (W')^r = \sum_{r=0}^{\infty} \rho^r Z_r. \tag{12}$$

Observe that, for any $i, j = 1, ..., n$, $(Z_r)_{i,j} = 0$ if $r < d(i,j)$. (This is because any two units $i, j = 1, ..., n$ cannot be joined by a walk of length less than $d(i,j)$, and hence $(W^k)_{i,j} = 0$, for any $k = 0, ..., r$, if $r < d(i,j)$.) It follows that (12) can be rewritten as

$$\Gamma_{i,j}(\rho) = \sum_{r=d(i,j)}^{\infty} \rho^r (Z_r)_{i,j}, \tag{13}$$

for any $i, j = 1, ..., n$. Comparing (13) and (6), it is clear that the matrices Z_r play, for the correlation structure of a SAR(1) model, a role similar to that played by the powers W^r for the correlation structure of a CAR(1) model. In particular, $\rho^r (Z_r)_{i,j}$ may be interpreted as the contribution to $\Gamma_{i,j}(\rho)$ of an r-walk from i to j. Note that CAR(1) and SAR(1) are directly comparable only if W is symmetric (because W must be symmetric in a CAR(1) model). In this case, Z_r equals simply $(r+1)W^r$, which clearly indicates that, for a fixed (symmetric) W, the correlation structures of CAR(1) and SAR(1) depend on W in a very similar way (more specifically, the contribution of a certain r-walk to $\Gamma_{i,j}(\rho)$ is just $r+1$ times its contribution to $\Sigma_{i,j}(\rho)$).

Given a graph with first distance matrix A, the two most popular specifications of SAR(1) models in applications are obtained by setting $V = I$, and W equal to A or $D^{-1}A$.[8] In these two cases, the relationship between the covariances implied by the models and the structure of the underlying graph is particularly simple. When $W = A$, $\Gamma_{i,j}(\rho)$ is determined by weighting the number $(A^r)_{i,j}$ of r-walks from i to j, precisely as in a CAR(1) model constructed with the same W (see the previous section), but with different functions of ρ, namely ρ^r in the CAR(1) model and $(r+1)\rho^r$ in the SAR(1) model.[9] When $W = D^{-1}A$, it is easily seen that the contribution $\rho^r (Z_r)_{i,j}$ of an r-walk $(i_0, i_1, ..., i_r)$ to $\Gamma_{i_0,i_r}(\rho)$ depends not only on ρ and r, but also on the vertices visited by the walk (more specifically, such a contribution equals $\rho^r \sum_{l=0}^{r} \delta_{i_l} / \prod_{l=0}^{r} \delta_{i_l}$).

[7] Virtually all weights matrices used in applications have at least a negative eigenvalue. In particular, since $tr[W] = 0$ by assumption and $\lambda_{\max} > 0$ by the Perron-Frobenius theorem, this is certainly the case if W is (similar to) a symmetric matrix. An important example is when W or $\overline{W}$ is equal to a row-standardized matrix $D^{-1}A$ (with symmetric A), because then W can be written as $V^{-1/2}D^{-1/2}D^{-1/2}AD^{-1/2}D^{1/2}V^{1/2}$, which is similar to the symmetric matrix $D^{-1/2}AD^{-1/2}$.

[8] Contrary to what happens to L and $\overline{W}$ in a CAR model (because of the constraint $L^{-1}\overline{W} = \overline{W}'L^{-1}$), the specification of V in a SAR(1) model does not restrict the specification of $\overline{W}$. In other words, the correlation structure of a SAR(1) model is determined not only by the choice of $\overline{W}$ but also, independently, by that of V.

[9] For a fixed $\rho < 2/3$, the function $(r+1)\rho^r$, $r = 1, 2, ...$, is decreasing in r. Except for very special cases, the inequality $\rho < 2/3$ is always satisfied when A is a $(0-1)$ irreducible matrix, due to the restriction $\rho < \lambda_{\max}^{-1}$.

We now turn to consider explicitly the correlations $\Gamma^*_{i,j}(\rho)$. Given representation (13), it is straightforward to check that Lemma 2.6 extends to a SAR(1) model, provided that $\Sigma^*_{i,j}(\rho)$ is replaced by $\Gamma^*_{i,j}(\rho)$ and W^r by Z_r. Thus $\Gamma^*_{i,j}(\rho)$, for any $i, j = 1, ..., n$, admits the simple expansion

$$\Gamma^*_{i,j}(\rho) = Z_{d(i,j)}(i,j)\rho^{d(i,j)} + Z_{d(i,j)+1}(i,j)\rho^{d(i,j)+1} + O(\rho^{d(i,j)+2}). \tag{14}$$

Correspondingly, we have:

Property 2.24 *In a SAR(1) model $W_{i,j} + W_{j,i}$ represents the first derivative of $\Gamma^*_{i,j}(\rho)$ at $\rho = 0$, for any $i, j = 1, ..., n$.*

Thus, the specification of the entries $W_{i,j}$ and $W_{j,i}$ has a direct effect on $\Gamma^*_{i,j}(\rho)$ when $|\rho|$ is small. As ρ gets larger, $\Gamma^*_{i,j}(\rho)$ is determined by the whole structure of W, as illustrated above. We stress that setting $W_{i,j} + W_{j,i} > W_{l,m} + W_{m,l}$, for any $i, j, l, m = 1, ..., n$, guarantees $\left|\Gamma^*_{i,j}(\rho)\right| > \left|\Gamma^*_{l,m}(\rho)\right|$ for sufficiently small ρ, but not necessarily for any ρ.

It should now be clear that not only Lemma but also all other formal results obtained in the previous section for CAR(1) models can be extended to SAR(1) models with little effort. In particular, Propositions 2.10,[10] 2.14 and 2.15 can be shown to hold for any SAR(1) model simply by replacing W^r with Z_r in the proofs. The counterparts of Propositions 2.16 and 2.17 require some modifications in the proofs, and are therefore stated here (and proved in the appendix).

Proposition 2.25 *For any SAR(1) model on a bipartite graph, $\Gamma^*_{i,j}(\rho)$, $i, j = 1, ..., n$, is: (a) an odd function if $d(i, j)$ is odd; (b) an even function if $d(i, j)$ is even.*

Note that if the graph underlying a SAR(1) model is not bipartite, then the (real) eigenvalue $-\lambda_{\max}$ does not belong to the spectrum of W by Lemma 2.3, which implies that no correlation $\Gamma^*_{i,j}(\rho)$ can be either an odd or an even function over the parameter space of a SAR(1) model.

Let us now assume that W has at least one negative eigenvalue (see footnote 7, and note that if W did not have a negative eigenvalue one could set $\lambda_{\min} = -\infty$, and then Proposition 2.26 below would require only a minor modification, omitted for the sake of brevity). Denoting by $m_{\min}$ the multiplicity of $\lambda_{\min}$ (with generally $m_{\min} = 1$, unless the graph underlying the SAR(1) model satisfies particular symmetries; e.g., Biggs, 1993), and by $q_{\min}$ an eigenvector associated to $\lambda_{\min}$, we have:

Proposition 2.26 *In a SAR(1) model, for any W and any $i, j = 1, ..., n$, (a) as $\rho \to \lambda^{-1}_{\max}$,*

[10] The SAR(1) models for which Proposition 2.10 holds are the model with $W = A$, corresponding to Model 1, and the model with $W = D^{-1/2}AD^{-1/2}$, corresponding to Model 2. The latter model is obtained from specifying $\overline{W} = D^{-1}A$ and $V = D^{-1}$; this is used for instance in Wall (2004), p. 316. Extensions to other SAR(1) models, e.g., the model with $W = D^{-1}A$, are straightforward.

$\Gamma^*_{i,j}(\rho) \to 1$; *(b) as* $\rho \to \lambda_{\min}^{-1}$,

$$\Gamma^*_{i,j}(\rho) \to \begin{cases} 1 & \text{if } m_{\min} = 1 \text{ and } (q_{\min})_i (q_{\min})_j > 0, \\ -1 & \text{if } m_{\min} = 1 \text{ and } (q_{\min})_i (q_{\min})_j < 0, \\ 0 & \text{if } m_{\min} = 1 \text{ and either } (q_{\min})_i = 0 \text{ or } (q_{\min})_j = 0, \end{cases} \tag{15}$$

whereas $\Gamma^*_{i,j}(\rho)$ *can go to any number in* $[-1, 1]$ *otherwise (i.e., if* $m_1 = 1$ *and* $(q_{\min})_i = (q_{\min})_j = 0$*, or if* $m_1 > 1$*).*

Some numerical investigation of properties connected to Proposition 2.26 can be found in Kelejian and Robinson (1995). As Proposition 2.17, Proposition 2.26 relies on both nonnegativity and irreducibility of W (cf. Remark 2.19). For instance, it would not hold if W were equal to the matrix with entries $W_{i,j} = 1$ if $i - j = 1$, $W_{i,j} = 0$ otherwise, which is reducible (and does not satisfy condition (ii) of Section 2). Such a matrix is the one that, if it were used in a SAR(1) model, would produce an AR(1) model (with random start-up $y_1 \sim N(0, \sigma^2)$, not covariance stationary).

For comparison, the counterparts of Figures 1 and 2 for SAR(1) models with $W = A$ and $W = D^{-1}A$ are given in Figures 3 and 4.[11]

Figures 3 and 4 about here

As in the case of a CAR(1) model (see Conjecture 2.23), the correlations $\Gamma^*_{i,j}(\rho)$ seem to be increasing in ρ over the interval $(0, \lambda_{\max}^{-1})$ for any W and any $i, j = 1, .., n$, implying that, when it belongs to such an interval, the parameter ρ can be thought of as a measure of spatial correlation. The correlations do not need to be monotonic in ρ when $\rho < 0$, or when $\rho > \lambda_{\max}^{-1}$ (which is one of the reasons why it is desirable to restrict the parameter space of SAR(1) models to $\rho < \lambda_{\max}^{-1}$).

3 Multi-Parameter Models

This section is concerned with the extension of one-parameter spatial autoregressions to multi–parameter models. In Section 3.1, we discuss various possibilities of performing such an extension. In Section 3.2, we analyze multi-parameter models from the perspective of the theory of exponential families.

Extending the parameter space of the one-parameter models can be regarded as a natural way to try to overcome some of the limitations of the one-parameter models, yet maintaining some of their advantages. After all, it would be surprising if a family of distributions indexed by only two parameters (ρ plus the parameter τ^2 or σ^2 scaling the covariance matrix) were sufficiently rich to successfully represent the interaction structure of variables observed over possibly very irregular lattices! In addition, the introduction of further parameters may help to compensate for the uncertainty associated to the specification of a weights matrix.

[11] We do not report the plots obtained for $W = D^{-1/2}AD^{-1/2}$ (see footnote 10), but we remark that they are very similar to those for $W = D^{-1}A$.

It is convenient to introduce the models by referring to the following general formulation of CAR and SAR models (normalized, as in Section 2, to include the hypothesis of i.i.d. data). Given $n \times n$ matrices C and S depending on known constants and unknown parameters (and such that $I - C$ is symmetric and p.d., and $I - S$ is nonsingular), a CAR model is specified through

$$y_i \mid \{y_j : j \neq i\} \sim N\left(\sum_{j=1}^{n} C_{i,j} y_j, \tau^2\right), \quad i = 1, 2, ..., n, \tag{16}$$

yielding

$$N_n\left(0, \tau^2 (I - C)^{-1}\right), \tag{17}$$

and a SAR model is

$$N_n\left(0, \sigma^2 [(I - S')(I - S)]^{-1}\right). \tag{18}$$

Then, when C and S have the linear structure $\sum_{h=1}^{p} \rho_h W_h$, $p > 1$, we call the models *multi-parameter* CAR or SAR models, denoted by CAR(p) and SAR(p). Here, $\rho_1, ..., \rho_p$ are functionally independent unknown parameters and $W_1, ..., W_p$ are (known) non-zero linearly independent weights matrices assumed to satisfy assumptions (i), (ii) and (iv) of Section 2. Assumption (iii) is now imposed on $\sum_{h=1}^{p} W_h$. Observe that, due to the employed normalization, $W_1, ..., W_p$ must be symmetric in a CAR(p) model.

3.1 Different Specifications

Given a graph with adjacency matrix W, there are two fundamentally different approaches to constructing multi-parameter autoregressions, the distinction being based on whether or not such models maintain the same conditional independence structure as the one-parameter model with weights matrix W. We recall that when $y \sim N_n(\mu, \Sigma)$ and for any $i \neq j$, y_i and y_j are *conditionally independent* given all the remaining random variables in y if and only if $(\Sigma^{-1})_{i,j} = 0$ (i.e., $(W_h)_{i,j} = 0$, for each $h = 1, ..., p$, in the case of a CAR(p) model). For details on the conditional independence structure of a Gaussian model, see Speed and Kiiveri (1986).

The *first* approach, which maintains the original conditional independence structure, consists of splitting the edge set of the graph in $p > 1$ disjoint subsets, say $E_1, ..., E_p$, and associating a parameter to each subset, so that C (or S) is parametrized as $\sum_{h=1}^{p} \rho_h W_h$ with $(W_h)_{i,j} \neq 0$ if and only if $(i, j) \in E_h$. For example, on a rectangular two-dimensional grid different parameters may be associated to horizontal and vertical edges, to account for potential anisotropy along the two main axes. On irregular lattices, reasonable criteria for splitting the edge set will vary from application to application. We are not aware of work addressing this issue, but a criterion that might prove to be useful to classify edges is their closeness to the borders of the graph, as could be measured, for instance, by the number $(A^2)_{i,j}$ of common neighbors of two vertices i and j forming an edge.

In the *second* approach to extending spatial autoregressions based on a single weights matrix W, additional parameters are associated to different degrees of neighborhood. We refer to such multi-parameter autoregressions as *higher-order* CAR or SAR models. Contrary to the time series case, in a spatial setting there is no obvious way of constructing higher-order autoregressions. The use of a graph theoretic perspective helps to clarify the differences among the following three alternatives that have been proposed in the literature (recall that A^h denotes the h-th power of A, whereas A_h denoted a h-th distance matrix, with $A_1 = A$):

(a) $W_h = A^h$, $h = 1, ..., p$ (e.g., Huang, 1984);

(b) $W_h = A_h$, $h = 1, ..., p$ (e.g., Anselin and Smirnov, 1996);

(c) $W_h = P_h$, $h = 1, ..., p$, P_h being the matrices with (i, j)-th entry equal to the number of paths of length h between i and j (Blommestein and Koper, 1997).

As for specification (a), the matrices A^h generally have non-zero elements on their diagonals (for instance, $A^h(i, i) > 0$, for any even h and any $i = 1, ..., n$), which requires modifications in our definition of CAR(p) models (see Martin, 1990). The fact that the entries of the powers A^h count the numbers of h-walks between two vertices, including the ones that are not paths, has been seen as a problem by some authors (see Blommestein and Koper, 1997, and references therein).[12] Specification (b) is in a sense the most natural in the graph metric induced by the graph distance $d(\cdot, \cdot)$. The matrices A_h are just the binary version of the matrices P_h used in (c). Note that specification (c) is always different from (a), because A is symmetric (and it is not the zero matrix) and hence cannot be (similar to) a triangular matrix (if A were a triangular matrix, for instance representing temporal unilateral interaction, then we would have $P_h = A^h$, for each h). It is worth remarking (a), (b) and (c) imply the same conditional independence structure, both for CAR(p) and for SAR(p) models. Specifically, two variables y_i and y_j in a CAR(p) (resp., SAR(p)) model are conditionally independent (given all the remaining variables) if and only if $d(i, j) > p$ (resp., $d(i, j) > 2p$).

Regardless of how multi-parameter models are constructed, it is clear that the dependence between the characteristics of the weights matrices and the correlation properties of the models is more complicated than in one-parameter models. Nevertheless, extensions of the results and interpretations given in Section 2 are possible. In the rest of this section, we present the main steps necessary for such extensions, but, for the sake of brevity, we leave details to the reader.

Essentially, what is needed for the analysis of multi-parameter models is to generalize the concept of weight of a walk to the case where p different adjacency matrices are available. Denote by $e_h(i, j)$ an edge in the graph with adjacency matrix $\rho_h W_h$, for any $h = 1, ..., p$. For given $W_1, ..., W_p$, we define a *p-variate r-walk* from i to j as an alternating sequence of vertices and edges $[i, e_{h_1}(i, l_1), l_1, e_{h_2}(l_1, l_2), l_2, ...l_{r-1}, e_{h_r}(l_{r-1}, j), j]$, for some $h_1, ..., h_r = 1, ..., p$. Similarly to Definition 2.1, we define the weight of a p-variate r-walk as the product of the weights of its steps, that

[12] It should be noted, however, that the fact that the matrices A^h, $h = 1, ..., p$, commute has a number of advantages for the analysis of the models.

is, $\rho_{h_1} W_{h_1}(i, l_1) \rho_{h_2} W_{h_2}(l_1, l_2)$.... For simplicity, let us restrict attention to the case of symmetric weights matrices. Observe that, when $\rho_1, ..., \rho_p$ are such that the eigenvalues of $\sum_{h=1}^{p} \rho_h W_h$ are all smaller than one in absolute value, a covariance matrix of the form

$$\Omega = \sigma^2 \left(I - \sum_{h=1}^{p} \rho_h W_h \right)^{-q},$$

($q = 1$ for a CAR(p) model, $q = 2$ for a SAR(p) model) admits the representation

$$\Omega = \sigma^2 \sum_{r=0}^{\infty} \binom{r+q-1}{r} \left(\sum_{h=1}^{p} \rho_h W_h \right)^r . \tag{19}$$

Then, on expanding all r-th powers in (19) (with care, because the multinomial theorem can be used unless the W's commute), the following generalization of Property 2.4 becomes clear:

Property 3.1 *Let y follow a CAR(p) model with weights matrices $W_1, ..., W_p$. For $\rho_1, ..., \rho_p$ such that all eigenvalues of $\sum_{h=1}^{p} \rho_h W_h$ are smaller than one in absolute value, any p-variate walk from i to j contributes its weight to $cov(y_i, y_j)$.*

By expression (19), the contribution of any p-variate r-walk in a SAR(p) model with symmetric weights matrices equals $r + 1$ times its weight (if the W_h's are not symmetric modifications similar to those in Section 2.3 are necessary).

It is also possible to obtain an interpretation of the entries of the matrices $W_1, ..., W_p$ similar to that of Properties 2.7 and 2.24. For instance, letting Σ_p^* be the correlation matrix of a CAR(p) model it is easily shown that

$$(W_h)_{i,j} = \left. \frac{\partial (\Sigma_p^*)_{i,j}}{\partial \rho_h} \right|_{\rho_1 = ... = \rho_p = 0} .$$

Property 3.1 is useful, among other things, to understand some differences in the way correlations are formed in the tree specifications (a), (b), (c) of higher-order models.[13] We illustrate with a simple example.

Example 3.2 In Figure 5 a planar configuration of observational units is given together with a corresponding graph, where two vertices are joined by an edge if the corresponding units are contiguous. Consider specifications (a), (b) and (c) for a SAR(2) model on this graph (for (a) we impose that the diagonal elements of $W_2 = A^2$ are zero). Then the first few terms in the expansion of $cov(y_3, y_6)$ are, using (19),

$$\begin{aligned} &\text{(a) } 18\rho_1\rho_2 + 12\rho_2^2 + 12\rho_1^3 + 48\rho_1^2\rho_2 + 140\rho_1\rho_2^2 + 120\rho_2^3 + 20\rho_1^4; \\ &\text{(b) } 12\rho_1\rho_2 + 12\rho_1^3 + 16\rho_1^2\rho_2 + 20\rho_1\rho_2^2 + 20\rho_1^4; \\ &\text{(c) } 18\rho_1\rho_2 + 12\rho_1^3 + 16\rho_1^2\rho_2 + 60\rho_1\rho_2^2 + 20\rho_1^4. \end{aligned} \tag{20}$$

[13] Of course, this is not to say that the choice of specification should be based only on the correlations it implies. Other criteria, say based on the data or on computational considerations, may be more relevant in applications.

By a multivariate extension of (14), these are also the first few terms in the expansion of the correlation $corr(y_3, y_6)$. Using the notions developed above, all coefficients in (20) can be interpreted, which allows to clarify how the structure of the underlying graph affects the shape of the correlations implied by different SAR(2). For instance, the coefficient of ρ_2^2 (corresponding to $(r!)^{-1}\ \partial^2 corr(y_3, y_6)/\partial\rho_2^2\big|_{\rho_1=\rho_2=0} = 3W_2^2$) equals $r+1$ times the total weight of the 2-variate r-walks from i to j composed only by steps in the graph with adjacency matrix W_2, with $r = 2$. Such a total weight is 0 for specification (b) and (c) (because there are no 2-variate 2-walks joining 3 and 6 when $W_2 = A_2$ or $W_2 = P_2$), and is 4 for specification (a) (4 being given by the two walks $[3, e_2(3,2), 2, e_2(2,6), 6]$ and $[3, e_2(3,5), 5, e_2(5,6), 6]$ having weight 1 each, and the walk $[3, e_2(3,4), 4, e_2(4,6), 6]$ having weight 2).

Figure 5 about here

3.2 Spatial Autoregressions as Exponential Families

The choice of the parametrization of the matrices C and S in (17) and (18) obviously determines how the correlation structure of the models depend on the parameters. Parameterizations other than $\sum_{h=1}^{p} \rho_h W_h$ are certainly possible, but are rarely considered in the spatial econometric literature, and are not considered in this paper. A strong reason for restricting attention to a linear parametrization of C and S is that this is the only parametrization (up to a diffeomorphism) such that the models are *exponential families* (see Efron, 1978, Amari, 1990, Kass and Vos, 1997). Recognizing that CAR(p) and SAR(p) models are exponential families is important not only because of the many nice statistical properties satisfied by such families, but also because it emphasizes a crucial difference between conditional and simultaneous models (see Proposition 3.3 below), and it reveals an interesting consequence of using highly structured weights matrices in SAR(p) models (see Proposition 3.5 below).

We briefly remind the reader that a family of densities is said to be an exponential family if its elements are representable as

$$pdf(y;\theta) = \exp\left\{\sum_{h=1}^{k} \eta_h(\theta) s_h(y) - \kappa(\theta)\right\} b(y), \tag{21}$$

with respect to some dominating measure on the sample space. Assuming that k is the smallest integer such that (21) holds, $(s_1(y), ..., s_k(y))'$ is the minimal sufficient statistic for the parameter θ. The parameter $\eta = (\eta_1(\theta), ..., \eta_k(\theta))'$ is called the *canonical parameter* and the canonical parameter space $\Omega \subseteq \mathbb{R}^k$ is the set of canonical parameters such that the integral of $\exp\{\sum_{h=1}^{k} \eta_h(\theta) s_h(y) - \kappa(\theta)\}$ is finite. Subject to weak regularity conditions, an exponential family is said to be *full* if for each $\eta \in \Omega$ there exists a density $pdf(y;\theta)$ in the family, *curved* otherwise. If the family is full and, in addition, Ω is an open subset of $\mathbb{R}^k$, then the exponential family is said to be *regular*.

Calling a matrix *full* if it does not contain any zero entries, we have:

Proposition 3.3 *(a) Any CAR(p) model is a regular exponential family. (b) A SAR(p) model is generally a curved exponential family. For a SAR(p) model to be a regular exponential family, the matrix* $I - \sum_{h=1}^{p} \rho_h W_h$ *must be full.*

The term "generally" in the statement of Proposition 3.3 is intentionally vague. More details can be found in the proof of Proposition 3.3 and in Proposition 3.5 below, but the important point here is that, in typical applications of SAR(p) models, $I - \sum_{h=1}^{p} \rho_h W_h$ is not full and hence SAR(p) models are curved. In particular, a SAR(1) model is certainly curved if W has at least one zero off-diagonal entry (or, equivalently, the underlying graph is not complete).[14] We remark that the essential difference between CAR(p) and SAR(p) models leading to Proposition 3.3 is that in a CAR(p) model the inverse of the covariance matrix—the precision matrix—lies in a vector space of dimension equal to the number (p+1) of parameters, whereas it generally lies in a higher-dimensional space in the case of a SAR(p) model. A very informative discussion, from a time series perspective, of the properties of Gaussian models with precision matrix having linear structure can be found in Anderson (1971), Chapter 6.

The necessary condition in part (b) of Proposition 3.3 is by no means sufficient. Indeed, for a SAR(p) model to be a regular exponential family, $\sum_{h=1}^{p} \rho_h W_h$ must not only be full, but must also have a very special structure. This is particularly transparent in the special case when all weight matrices W_h are symmetric, and can be proved by referring to the concept of a quadratic subspace of symmetric matrices, first introduced by Seely (1971) and subsequently extensively used in the statistical literature on variance components. We denote by $\mathcal{L}(n)$ the vector space of all $n \times n$ real symmetric matrices.

Definition 3.4 *A subspace* $\mathcal{Q}$ *of* $\mathcal{L}(n)$ *is a quadratic subspace if* $Q \in \mathcal{Q} \Rightarrow Q^2 \in \mathcal{Q}$.

Letting Ψ_p denote the subspace of $\mathcal{L}(n)$ spanned by the matrices $W_0, W_1 ..., W_p$, we have:

Proposition 3.5 *A SAR(p) model with symmetric weights matrices* $W_1, ..., W_p$ *is a curved exponential family, except when* Ψ_p *is a quadratic subspace, in which case it is a regular exponential family.*

A quadratic subspace is a highly structured set of matrices. We will see in the next section that, in order to obtain a quadratic subspace, the spatial configuration of the observational units must satisfy a very high level of regularity, and a specific number of weights matrices must be used.

Remark 3.6 Proposition 3.5 gives a necessary and sufficient condition for the conditional and the simultaneous approaches to the specification of autoregressive models to be equivalent, because the

[14] A circular AR(1) model is a regular exponential model but does not constitute a counterexample to Proposition 3.3 (b), because the matrix W necessary to write it as a SAR(1) model would not satisfy assumption (ii) of Section 2.

covariance matrices of CAR(p) and SAR(p) models (constructed with the same matrices $W_1, ..., W_p$) belong to the same subspace Ψ_p if and only if Ψ_p is a quadratic subspace.

The practical importance of Propositions 3.3 and 3.5 lies in the fact that statistical curvature has well-known consequences on the properties of inferential procedures (e.g., Amari, 1990, Kass and Vos, 1997). For instance, any efficient estimator in the context of a curved exponential family involves a loss of information which should somehow be recovered, typically by conditioning on an approximately ancillary statistic.[15] As far as hypothesis testing is concerned, it is well known that large statistical curvature is generally associated to poor performance of tests in exponential families; for example, this has been shown formally for locally most powerful tests in one-parameter exponential families (e.g., Efron, 1978, Kallenberg, 1981). In the particular case of CAR(1) and SAR(1) models curvature should therefore play an important role in determining the power of tests of spatial autocorrelation. The simplest observation along these lines is that the Moran test for spatial autocorrelation (Moran, 1950) is uniformly most powerful in the class of similar (or invariant, with respect to a suitable group of transformations on the sample space) tests of $\rho = 0$ against the alternative of a (pure) CAR(1) model (which has zero curvature), for any W, but generally (and certainly if W has at least one zero off-diagonal entry, by Proposition 3.3) not against the alternative of a SAR(1) model (for details, we refer to Martellosio, 2006). The dependence of differential geometric measures of curvature, such as the Efron curvature (see Efron, 1978, and Amari, 1990), on W and ρ could certainly be studied, but this goes beyond the scope of the present paper.

We close this section by remarking that the implications of statistical curvature become even more serious when the mean of the models, instead of being zero, is parametrized as a regression function $X\beta$, because—as it is easily seen by writing out their density—this makes CAR(p) models curved (unless the column space of X is spanned by a set of linearly independent eigenvectors of W), and introduces further curvature in SAR(p) models (in the sense that it increases the dimension of a minimal sufficient statistic, with respect to the dimension of the parameter space).

4 Symmetries and Regularities

In this section we study properties of CAR(p) and SAR(p) models when the underlying graphs satisfy some symmetries (in a group theoretic sense) or some regularities (in a combinatorial sense). The results derived here, contrary to those obtained in Section 2, are probably not directly useful to the applied spatial econometrician, who typically faces data coming from very irregular configurations.[16] However, as we shall see, they are crucial to understand the connection between the second-order

[15] Finding suitable approximately ancillary statistics may be difficult in practice. In fact, to the best of our knowledge, no attempts to "recover information" have been made so far in the context of spatial autoregressive models. Future research in this area may prove fruitful.

[16] Of course, regular spatial configurations such as uniform grids are relevant in simulation experiments; e.g., Conley and Molinari (2006).

properties of spatial autoregressions and the structural properties of the underlying graphs.

For the sake of simplicity, in this section we focus on the models constructed by using the first p distance matrices of a graph as weights matrices, that is, according to specification (b) of Section 3. We refer to such models simply as *CAR(p) and SAR(p) models on a graph.*

An *automorphism* of a graph G with vertex set $V(G)$ and edge set $E(G)$ is a permutation g on $V(G)$ preserving adjacency (and non-adjacency), i.e. such that $(i,j) \in E(G)$ if and only if $(g(i), g(j)) \in E(G)$, for any $i,j \in V(G)$, where $g(i)$ denotes the action of g on i; Cvetković *et al.* (1980), Biggs (1993). The set of the automorphisms of a graph G, which is obviously a subgroup of the symmetric group of degree n, will be denoted by $Aut(G)$. The $n \times n$ permutation matrix representing a permutation g will be denoted by P_g. Note that the action of $Aut(G)$ on $V(G)$ induces naturally an action on $E(G)$ as well. Evidently, the covariance matrix of any CAR(p) or SAR(p) model on a graph G, denoted by Υ, is invariant under $Aut(G)$, in the sense that $P_g \Upsilon P_g' = \Upsilon$ for each $g \in Aut(G)$.[17] Hence, $\Upsilon_{i,j} = \Upsilon_{l,m}$ if there exists $g \in Aut(G)$ such that $g(i) = l$ and $g(j) = m$, for $i,j,l,m = 1,...,n$ (the same property holding, of course, also for the correlations).

Example 4.1 When y follows a CAR(p) or SAR(p) model on the graph of Figure 5, $cov(y_i, y_2) = cov(y_j, y_5)$ for $(i,j) = (2,5), (1,1), (3,3), (4,4), (6,6)$.

Clearly, models defined on graphs having large automorphism group have simple covariance properties. The following definition can be found, for instance, in Biggs (1993).

Definition 4.2 *A graph G is said to be: vertex-transitive if $Aut(G)$ acts transitively on $V(G)$; edge-transitive if $Aut(G)$ acts transitively on $E(G)$; distance-transitive if for all $i,j,l,m \in E(G)$ such that $d(i,j) = d(l,m)$, there exists $g \in Aut(G)$ such that $g(i) = l$ and $g(j) = m$.*

It follows from the definition that if G is vertex-transitive, then the variances $\Upsilon_{i,i}$ of any CAR(p) or SAR(p) model on G do not depend on i, whereas if G is edge-transitive, then the covariances $\Upsilon_{i,j}$, with i and j neighbors, do not depend on the pair i,j. For the reader's convenience, examples of a small vertex transitive graph and of a small edge-transitive graph are given in Figure 6.

Figure 6 about here

It is of interest, at this point, to characterize a notion of stationarity for our models on arbitrary, and finite, graphs. We do not distinguish between weak and strong stationarity, since in this paper we are concerned only with Gaussian distributions.

[17] In group theoretic terminology, this is equivalent to saying that Σ belongs to the commuting algebra (centralizer) $\mathcal{C}(G)$ of the linear representation of $Aut(G)$ in $\mathbb{R}^n$. A basis of $\mathcal{C}(G)$ is given by the relationship matrices R_l, $l = 1,...,t$, of the orbits of the action of $Aut(G)$ on $V(G) \times V(G)$; see James (1957) for an application of these notions in statistics.

Definition 4.3 *A model for the random vector y, whose index set is in a one to one correspondence with the vertex set of a graph G, is said to be G-stationary if the covariance between any two variables y_i and y_j depends on i, j only through $d(i, j)$.*

It should be noted that G-stationarity is a very strong, and in a sense impractical, property of a statistical model. The reasons for introducing such a notion are: (i) it is the most natural concept of stationarity on arbitrary graphs endowed with the metric induced by the distance $d(\cdot, \cdot)$; (ii) it helps to clarify the relations between the correlation structure of spatial autoregressions and the properties of the underlying graphs. We remark that, in the particular case of regular lattices, weaker notions of stationarity are usually formulated in terms of invariances of the distribution with respect to a set of translations, and are based on the Euclidean metric, rather than a graph metric.

A sufficient condition for CAR(p) or SAR(p) models on a graph G to be G-stationary is, evidently, distance-transitivity of G. However, the condition is not necessary, in the same way as vertex-transitivity is not necessary for the models to be homoskedastic. Interestingly, it turns out that necessary and sufficient conditions for homoskedasticity and G-stationarity of CAR(p) or SAR(p) models on a graph G are given in terms of combinatorial regularities, rather than group theoretic symmetries, of G. To show that this is the case, we need to generalize the concepts of vertex-transitivity and distance-transitivity to the concepts of walk-regularity and distance-regularity, respectively.

A graph is said to be *walk-regular* if the number of closed r-walks starting at a vertex i does not depend on i, for each $r = 2, 3, ...$ (Godsil and McKay, 1980). Observe that walk-regularity implies degree-regularity. When $p = 1$, we have:

Proposition 4.4 *Any CAR(1) or SAR(1) model with $W = A$ is homoskedastic if and only if G is walk-regular.*

Certainly, it is well-understood that CAR(1) and SAR(1) models are in general heteroskedastic. The novelty of Proposition 4.4 lies in identifying precisely the circumstances, expressed in terms of the structure of the underlying graphs, in which they are homoskedastic.

Remark 4.5 One might have expected degree-regularity of a graph to be sufficient for homoskedasticity of CAR(1) and SAR(1) models with $W = A$. According to Proposition 4.4, this is not the case, because a degree-regular graph needs not be walk-regular. However, from the proof of Proposition 4.4 it can be deduced that the variances of a model on a degree-regular graph are, in general, all very similar, at least when ρ is not too large and the *girth* (i.e., the length of the shortest cycle) of the graph is large. This is because, if G is a degree-regular graph with girth g, then $(A^r)_{i,i}$ does not depend on i, for $r = 0, ..., g - 1$; see Fiol (1997), p. 12.

Remark 4.6 Proposition 4.4 does not, in general, extend to higher-order models on graphs (with $W_h = A_h$, $h = 1, ..., p$). This is because walk-regularity does not necessarily imply that the powers of

A_h, $h \geq 2$, have constant diagonal (and hence, using the same notation as in the proof of Proposition 4.4, the variances $\sum_{r=0}^{\infty} \alpha_r \left(\sum_{h=1}^{p} \rho_h A_h\right)^r$ depend on i). In contrast, higher-order models with $W_h = A^h$, $h = 1, ..., p$, are always homoskedastic when the underlying graph is walk-regular.

A graph is said to be *distance-regular* if the number of vertices at any given graph distances d_1 from a vertex i and d_2 from a vertex j only depends on $d(i, j)$; see, e.g., Brouwer *et al.* (1989). This class of graphs, which contains the class of distance-transitive graphs and is contained in the class of walk-regular graphs, has attracted a great deal of attention in both the mathematical and the statistical literature. The importance of distance-regular graphs to our treatment of spatial autoregressions is emphasized by the next result.

Theorem 4.7 *Any CAR(p) or SAR(p) model on a graph G is G-stationary if and only if G is distance-regular.*

The notion of distance-regular graph also enables us to further elucidate the relationship between the curvature of a SAR(p) model and the regularities of the underlying graph (see Proposition 3.5).

Theorem 4.8 *A SAR(p) model on a graph G is a regular exponential family if G is distance-regular with diameter p, a curved exponential family otherwise.*

Simple examples of distance-regular graphs are a complete graph (i.e., a graph where any two vertices are neighbors), which has diameter 1, and a d-dimensional cube, which has diameter d. For many other examples, see Brouwer *et al.* (1989).

Remark 4.9 Theorems 4.7 and Theorem 4.8 hold also for SAR(p) models constructed according to specification (a) of Section 3 rather than (b) (or for models with covariance matrix $\left(I - \sum_{h=1}^{p} \rho_h A^h\right)^{-1}$, which are not CAR models because they cannot be decomposed as in (16), due to the fact that generally $(A^h)_{i,i} \neq 0$). This can be easily deduced from the proof of the theorems, because, when G is distance-regular with diameter ξ, the matrices $A^0, A^1, ..., A^\xi$ span the same (Bose-Mesner) algebra as $A_0, A_1, ..., A_\xi$ (Bannai and Ito, 1984).

5 Concluding Remarks

The purpose of this paper was to analyze the correlation structure of spatial autoregressions on arbitrary graphs. We have provided an interpretation of how the correlations implied by CAR(1) and SAR(1) models depend on the properties of the underlying graph. The basic observation is that, when the random vector y is modeled as a CAR(1) or SAR(1) process, $cov(y_i, y_j)$ can be viewed as a sum of contributions coming from *all* walks between the observational units i to j, with each contribution being the weight of a walk. We have shown that such an interpretation helps to explain some properties and peculiarities of the models. Extensions to multi-parameter

spatial autoregressions are more complicated, but can be based on the results pertaining the one–parameter models. Our analysis has also shed some light on how different specifications of the weights matrices affect the correlation properties of spatial autoregressions. This is important, because the specification of weights matrices may be very difficult and uncertain when the models are defined on irregular lattices. We have also studied the stronger properties enjoyed by the models when the underlying graphs satisfy symmetries or regularities.

Some comments concerning the assumption of Gaussianity as the joint distribution of a spatial process are in order. In particular, it is natural to ask how Gaussianity affects the interaction structure of a process defined on a graph. The question is best answered in the context of Markov random fields, of which CAR models are an example. A random field is said to be a *pairwise interaction process* if its log-density does not depend on the data through interaction terms $y_{i_1} y_{i_2} ... y_{i_m}$ of order m higher than two (for details, see Besag, 1974, or Section 6.4 of Cressie, 1993). Now, by the Hammersley-Clifford theorem, Gaussianity implies that a Markov random field is a pairwise interaction processes. But, again by the Hammersley-Clifford theorem, the same restriction is implied by the absence in the graphs of *cliques* (i.e., sets of vertices which are all neighbors of each other) of size greater than two. We therefore conclude that Gaussianity does not restrict (in the above sense) the interaction structure of Markov random fields defined on graphs that do not contain cliques of size greater than two, such as bipartite graphs (see Section 2.1). On the other hand, it may well be possible that pairwise interaction is too restrictive for data observed over graphs with cliques of size greater than two; see Besag and Tjelmeland (1998) and Lee *et al.* (2001). This is an interesting aspect to consider in future work, since the graphs of irregular lattices do generally contain cliques of size greater than two.

Finally, in this paper we have not discussed, if not very briefly, advantages and disadvantages of CAR and SAR models, with respect to each other or with respect to alternative spatial models (e.g., models that parametrize directly a covariance matrix). We hope these issues will receive more attention in future econometric research.

Acknowledgements

This paper is based on material from my Ph.D. dissertation at the University of Southampton. I am indebted to Grant Hillier for his guidance, and I thank Paolo Paruolo, Peter Robinson, Kees Jan van Garderen for discussion and encouragement. I appreciate comments from participants at the Third Workshop in Spatial Econometrics, Strasbourg, June 2004, the Cemmap Workshop in Spatial Econometrics, London, November 2004, and the First Italian Congress of Econometrics and Empirical Economics, Venice, January 2005.

Appendix. Proofs

Proof of Lemma 2.6: Let the r-th derivative of a function $f(x)$ be denoted by $D^r f(x)$, and $D^r f(x)$ evaluated at $\rho = 0$ be denoted by $D^r f(0)$. Representation (8) is derived from the MacLaurin expansion $\Sigma^*_{i,j}(\rho) = \sum_{r=0}^{\infty} \frac{1}{r!} D^r \Sigma^*_{i,j}(0)\rho^r$, by expressing each element $D^r \Sigma^*_{i,j}(0)/r!$ in terms of entries of powers of W. This can be achieved as follows. For $|\rho| < \lambda_{\max}^{-1}$ and any $i, j = 1, ..., n$, write $\Sigma^*_{i,j}(\rho) = \Sigma_{i,j}(\rho)\eta(v(\rho))$, with $\eta(z) = z^{-1/2}$ and $v(\rho) = \Sigma_{i,i}(\rho)\Sigma_{j,j}(\rho)$. By Leibniz's formula

$$\frac{1}{r!} D^r \Sigma^*_{i,j}(0) = \sum_{s=0}^{r} \frac{1}{s!(r-s)!} D^{r-s}\Sigma_{i,j}(0) D^s \eta(v(0)). \tag{22}$$

The term $D^s\eta(v(0))$ can be obtained by Faá di Bruno's formula (e.g., Albramowitz and Stegun, 1979) as

$$D^s\eta(v(0)) = \sum \frac{s!}{k_1!...k_s!} (D^k\eta)(v(0)) \prod_{t=1}^{s} \left(\frac{D^t v(0)}{t!} \right)^{k_t}, \tag{23}$$

where $k = \sum_{t=1}^{s} k_t$, the summation is over all s-tuples of non-negative integers $(k_1, ..., k_s)$ such that $\sum_{t=1}^{s} tk_t = s$. We therefore need the terms $(D^k\eta)(v(0))$ (i.e., the k-th derivative of $\eta(z)$ evaluated at $v(0)$) and $D^t v(0)/t!$. Since $D^k\eta(z) = \alpha_k z^{-1/2-k}$, with $\alpha_k = \prod_{l=0}^{k-1} \left(-\frac{1}{2} - l\right)$, and $v(0) = 1$, $(D^k\eta)(v(0)) = \alpha_k$. From (6), we have

$$v(\rho) = \sum_{r,s=0}^{\infty} \rho^{r+s} (W^r)_{i,i} (W^r)_{j,j} = \sum_{r=0}^{\infty} \rho^r \sum_{u=0}^{r} (W^{r-u})_{i,i} (W^u)_{j,j},$$

and hence

$$\frac{D^t v(0)}{t!} = \sum_{u=0}^{t} (W^{t-u})_{i,i} (W^u)_{j,j}.$$

Thus, since $D^1 v(0) = 0$, (23) yields $D^s\eta(v(0)) = 0$ if $s = 1$, and

$$D^s\eta(v(0)) = \sum \frac{s!\alpha_k}{k_2!...k_s!} \prod_{t=2}^{s} \left(\sum_{u=0}^{t} (W^{t-u})_{i,i} (W^u)_{j,j} \right)^{k_t}, \tag{24}$$

if $s > 1$, where $\sum_{t=2}^{s} k_t = k$ and the first summation is over all $(s-1)$-tuples of non-negative integers $(k_2, ..., k_s)$ such that $\sum_{t=2}^{s} tk_t = s$. Substituting $D^{r-s}\Sigma_{i,j}(0) = (r-s)!(W^{r-s})_{i,j}$ and (24) into (22) gives

$$\frac{1}{r!} D^r \Sigma^*_{i,j}(0) = (W^r)_{i,j} - \sum_{s=2}^{r} (W^{r-s})_{i,j} \sum \frac{\alpha_k}{k_2!...k_s!} \prod_{t=2}^{s} \left(\sum_{u=0}^{t} (W^{t-u})_{i,i} (W^u)_{j,j} \right)^{k_t},$$

which completes the proof of the lemma.

Proof of Proposition 2.10: When $W = A$, $W_{i,j} = 1$ and $(W^2)_{i,j} = \sum_{l_1=1}^{n} A_{i,l_1} A_{l_2,j}$ is equal to the number of common neighbors of i and j, for any pair of neighbors i, j. On the other hand, when $W = D^{-1/2} A D^{-1/2}$, $W_{i,j} = (\delta_i \delta_j)^{-1/2}$, for any pair of neigbors i, j. The proposition then follows easily from expression (10).

Proof of Proposition 2.14: Since $(W^{d(i,j)})_{i,j} > 0$ for any W and for any $i,j = 1,...,n$, expression (10) establishes that the larger $d(i,j)$ is, the slower $\Sigma^*_{i,j}(\rho)$ goes to zero as $\rho \to 0$, proving the desired result.

Proof of Proposition 2.15. Nonnegativity and irreducibility of W imply that at least one term in (6) is positive when $\rho > 0$, proving the first part of the proposition. Representation (6) also shows that if $\rho < 0$ and $d(i,j)$ is odd (resp. even) then $\Sigma^*_{i,j}(\rho) < 0$ (resp. $\Sigma^*_{i,j}(\rho) > 0$) for small enough ρ, since $(W^{d(i,j)})_{i,j} > 0$, for any W and any any $i,j = 1,...,n$.

Proof of Proposition 2.16: If the graph underlying a CAR(1) model is bipartite, then $\lambda_1 = \lambda_{\max}$ (because of Lemma 2.3 and the fact that $|\lambda_1| \leq \lambda_{\max}$ by the Perron-Frobenius theorem), and consequently representation (6) holds over the whole parameter space of the model. By the same representation, $\Sigma_{i,j}(\rho)$, for any $|\rho| < \lambda_{\max}^{-1}$ and any $i,j = 1,...,n$, is an odd function if $(W^r)_{i,j} = 0$ for even r, whereas it an even function if $(W^r)_{i,j} = 0$ for odd r. We now show that when the underlying graph is bipartite, $(W^r)_{i,j} = 0$ if r and $d(i,j)$ have different parity. If $(W^r)_{i,j}$ were different from zero for r and $d(i,j)$ with different parity, then there would be a closed walk of length $r + d$, which would be odd. But this is impossible, because a closed walk of odd length always contains a cycle of odd length, and a bipartite graph does not contain any cycles of odd length. We can therefore conclude that $\Sigma_{i,j}(\rho)$, $i,j = 1,...,n$, is an odd function if $d(i,j)$ is odd, whereas it is an even function if $d(i,j)$ is even. The proof is completed on noting that, under bipartiteness, $\Sigma_{i,i}(\rho)$, $i = 1,...,n$, is an even function, since $d(i,i) = 0$.

Proof of Proposition 2.17: Consider the spectral decomposition $W = \sum_{l=1}^n \lambda_l q_l q_l'$, where q_l is an eigenvector of W associated to λ_l. Then, $(I - \rho W)^{-1} = \sum_{l=1}^n (1 - \rho\lambda_l)^{-1} q_l q_l'$. According to the Perron-Frobenius theorem, $\lambda_{\max}$ is a simple eigenvalue and q_n can be taken to be entrywise positive. It follows that the (eigenprojection) matrix $q_n q_n'$ is entrywise positive, from which it is easily seen that, as $\rho \to \lambda_{\max}^{-1}$, $\Sigma^*_{i,j}(\rho) \to (q_n)_i (q_n)_j / [(q_n)_i^2 (q_n)_j^2]^{1/2} = 1$, for any $i,j = 1,...,n$, proving the first part of the proposition. Similarly, as $\rho \to \lambda_1^{-1}$, $\Sigma^*_{i,j}(\rho) \to \sum_{l=1}^{m_1} (q_l)_i (q_l)_j / [\sum_{l=1}^{m_1} (q_l)_i^2 \sum_{l=1}^{m_1} (q_l)_j^2]^{1/2}$. Depending on q, this limit can take any value in $[-1, 1]$, and the particular cases in (11) are easily established.

Proof of Proposition 2.25: From (12), write

$$\Gamma_{i,j}(\rho) = \sum_{r,s=0}^{\infty} \left\{ \rho^{r+s} \sum_{l=1}^{n} (W^r)_{i,l} (W^s)_{j,l} \right\},$$

for $|\rho| < \lambda_{\max}^{-1}$. If the graph underlying the SAR(1) model is bipartite, then no l exists such that $(W^r)_{i,l}(W^s)_{j,l} > 0$ for $r + s$ and $d(i,j)$ having different parity. This is because if such an l existed, then there would be a closed walk of length $d(i,j) + r + s$, which would be odd, and hence the graph could not bipartite, for a closed walk of odd length always contains a cycle of odd length. The proposition then follows using the same arguments as in Proposition 2.16.

Proof of Proposition 2.26: Let $\eta_1, ..., \eta_n$ denote the eigenvalues of $\Gamma(\rho)$ in non-decreasing order, and let $v_1, ..., v_n$ denote a set of orthonormal eigenvectors of $\Gamma(\rho)$, with v_l associated to η_l,

$l = 1, ..., n$. Since W is nonnegative and irreducible, the matrix $(I - \rho W)^{-1}$ is entrywise positive for any $0 < \rho < \lambda_{\max}^{-1}$ (e.g., Gantmacher 1974, p. 69). It follows that, for any $0 < \rho < \lambda_{\max}^{-1}$, $\Gamma(\rho)$ is entrywise positive and hence, by Perron's theorem (e.g., Horn and Johnson, 1985, Theorem 8.2.11), that η_n is simple. Observe also that η_n is the only eigenvalue of $\Gamma(\rho)$ that does not have a finite limit as $\rho \to \lambda_{\max}^{-1}$, because the eigenvalues of $\Gamma(\rho)$ are continuous in ρ, and $rank[(I - \lambda_{\max}^{-1} W')(I - \lambda_{\max}^{-1} W)] = rank[(I - \lambda_{\max}^{-1} W)] = n-1$, for $\lambda_{\max}$ is a simple eigenvalue of W by the Perron-Frobenius theorem. Next, observe that, as $\rho \to \lambda_{\max}^{-1}$, $v_n \to q_n$, where q_n is an eigenvector of W associated to $\lambda_{\max}$ (because when $\rho = \lambda_{\max}^{-1}$, $\Gamma^{-1}(\rho)$ has an eigenvector q_n corresponding to its smallest eigenvalue 0). Using the spectral decomposition $\sum_{l=1}^{n} \eta_l v_l v_l'$ of $\Gamma(\rho)$ in $\Gamma_{i,j}^*(\rho) = \Gamma_{i,j}(\rho)/(\Gamma_{i,i}(\rho)\Gamma_{j,j}(\rho))^{1/2}$, it is easily seen that as $\rho \to \lambda_{\max}^{-1}$, $\Gamma_{i,j}^*(\rho) \to (v_n)_i (v_n)_j / [(v_n)_i^2 (v_n)_j^2]^{1/2} = 1$, for $i, j = 1, ..., n$, establishing the first part of the proposition. Similarly to above, η_1 is the only eigenvalue of $\Gamma(\rho)$ that does not have finite limit as $\rho \to \lambda_{\min}^{-1}$, because the eigenvalues of $\Gamma(\rho)$ are continuous in ρ, and $rank[(I - \lambda_{\min}^{-1} W')(I - \lambda_{\min}^{-1} W)] = rank[(I - \lambda_{\min}^{-1} W)] = m_{\min}$. Then, as $\rho \to \lambda_{\min}^{-1}$, $\Gamma_{i,j}^*(\rho) \to \sum_{l=1}^{m_{\min}} (v_l)_i (v_l)_j / [\sum_{l=1}^{m_{\min}} (v_l)_i^2 \sum_{l=1}^{m_{\min}} (v_l)_j^2]^{1/2}$. The particular cases in (15) follow easily from the fact that when $m_{\min} = 1$, $v_1 \to q_{\min}$, because $\Gamma^{-1}(\lambda_{\min}^{-1})$ has an eigenvector q_1 corresponding to its smallest eigenvalue 0. In the remaining cases, the limit can take any value in $[-1, 1]$, depending on $v_1, ..., v_{m_{\min}}$.

Proof of Proposition 3.3: Setting $\rho_0 = -1$ and $W_0 = I$, it is readily verified that a CAR model is a family of densities (21) with $\eta_h = \rho_h/(2\tau^2)$ and $s_h(y) = y' L^{-1} W_h y$, for $h = 0, 1, ..., p$. Since the matrices W_h are assumed to be linearly independent, the $(p+1)$-dimensional statistic $s(y)$ is minimal sufficient. The canonical parameter space Ω is the set of parameters η_h such that $\Sigma_{c,p} = -(\sum_{h=0}^{p} \eta_h W_h)^{-1} L$ is p.d. For any p and any $W_1, ..., W_p$, $|\Sigma_{c,p}|$ is a continuous function of $\eta_0, ..., \eta_p$, by the definition of determinant of a square matrix. Thus, $|\Sigma_{c,p}| \to |L|$ as $(\eta_1, ..., \eta_p)$ approaches the p-dimensional zero vector, with the consequence that there is always a non-empty p-ball centered at the p-dimensional zero vector where $\Sigma_{c,p}$ is p.d. That is, the parameter space of the model is non-empty for any p and any choice of the matrices $W_1, ..., W_p$. Continuity of $|\Sigma_{c,p}|$ also implies that Ω is open, which completes the proof of part (a) of the proposition. We now move to part (b). It is easily seen that a SAR(p) model is embedded in a regular exponential family with sufficient statistics $y' B_{h_1,h_2} y$, $h_1 = 0, 1, ..., p$, $h_2 = 0, 1, ..., h_1$, where $B_{h_1,h_2} = W_{h_1}' W_{h_2} + (1 - \delta_{h_1,h_2}) W_{h_2}' W_{h_1}$, δ_{h_1,h_2} being the Kronecker delta, equal to 1 if $h_1 = h_2$, to 0 otherwise. Let Δ_p be the subspace (of the vector space of all $n \times n$ real symmetric matrices) spanned by the matrices B_{h_1,h_2}, for $h_1 = 0, 1, ..., p$ and $h_2 = 0, 1, ..., h_1$. The dimension of Δ_p, to be denoted by $\dim(\Delta_p)$, is the dimension of a minimal sufficient statistic in a SAR(p) model. Since the matrices $B_{h,0}$, $h = 0, 1, ..., p$, are linearly independent (because the matrices W_h are linearly independent by assumption), $\dim(\Delta_p) \geq p + 1$, and a SAR($p$) model is a regular exponential family if $\dim(\Delta_p) = p+1$, a curved exponential family if if $\dim(\Delta_p) > p + 1$ (see, e.g., Amari, 1990, p. 109, and note that openness of the canonical parameter space follows by the same argument as the one used above for a CAR(p) model). To complete the proof it remains to show that $\dim(\Delta_p) > p + 1$ if $I - \sum_{h=1}^{p} \rho_h W_h$ is not full. To

see this, consider two units i^*, j^* at distance 2 in the graph having adjacency matrix $\sum_{h=1}^{p} \rho_h W_h$. When $I - \sum_{h=1}^{p} \rho_h W_h$ is not full, existence of two such units is guaranteed by the assumption of irreducibility of $\sum_{h=1}^{p} \rho_h W_h$, which implies that the underlying graph is connected and $n > 2$. Observe that $W_h(i^*, j^*) = 0$, for any $h = 1, ..., p$, and that there exist some $h_1^*, h_2^* = 1, ..., p$ such that $W'_{h_1^*} W_{h_2^*}(i^*, j^*) > 0$. It follows, by assumption (ii) of Section 2, that $B_{h_1^*, h_2^*}$ is linearly independent of $B_{h,0}$, for any $h = 0, 1, ..., p$, which implies $\dim(\Delta_p) > p + 1$.

Proof of Proposition 3.5: Let Γ_p denote the covariance matrix of a SAR(p) model. The matrix Γ_p^{-1} is the square of a matrix belonging to the $(p+1)$-dimensional subspace Ψ_p. Therefore, if Ψ_p is a quadratic subspace, $\Gamma_p^{-1} \in \Psi_p$, and hence the SAR(p) model is a regular exponential family. Conversely, assume that a SAR(p) model is a regular exponential family. Then, Γ_p^{-1} must belong to a $(p+1)$-dimensional subspace, say Θ_p, of $\mathcal{L}(n)$. It follows that all the matrices $W_0, ..., W_p$ belong to Θ_p and, since they are linearly independent by assumption, that they span Θ_p. Hence $\Theta_p = \Psi_p$, which implies that Ψ_p is a quadratic subspace, Definition 3.4 being satisfied with $Q = \Gamma_p^{-1/2}$.

Proof of Proposition 4.4: When $W = A$ in a CAR(1) or SAR(1) model, and for $|\rho| < \lambda_n^{-1}$ and any $i = 1, ..., n$, $var(y_i) = \sum_{r=0}^{\infty} \alpha_r \rho^r (A^r)_{i,i}$, with $\alpha_r = 1$ for a CAR(1) model, $\alpha_r = r + 1$ for a SAR(1) model (having fixed $\sigma^2 = \tau^2 = 1$). Since $(A^r)_{i,i}$ is equal to the number of closed r-walks starting at i, it follows that, for $|\rho| < \lambda_n^{-1}$, $var(y_i)$ is independent of i if and only if G is walk-regular. The condition extends to all values of ρ such that the covariance matrices of the models are p.d., because $var(y_i)$, $i = 1, ..., n$, is an analytic function of ρ and as such completely determined by the values they take on any non-empty open interval contained in its domain.

Proof of Theorem 4.7: Assume that the graph G with diameter ξ is distance-regular. Then, the distance matrices $A_0, A_1, ..., A_\xi$ of G span a Bose-Mesner algebra; see, for instance, Bannai and Ito (1984). Since a Bose-Mesner algebra is closed under matrix generalized inversion, it follows that for any $1 \leq p \leq \xi$, the covariance matrix of a CAR(p) model on G is a linear combination of $A_0, A_1, ..., A_\xi$, i.e., according to Definition 4.3, the CAR(p) model is G-stationary. Closure under matrix multiplication of a Bose-Mesner algebra entails that the same conclusion holds for a SAR(p) model on G, proving the sufficiency of the condition in the proposition. Conversely, assume that a CAR(p) or SAR(p) model on a graph G with diameter ξ is G-stationary, i.e., that its covariance matrix, say Σ_p, belongs to $Span\{A_0, ..., A_\xi\}$. Then, $A_1^h \in Span\{A_0, ..., A_\xi\}$, for $h = 0, 1, ...$, because there always exists an open p-ball Θ_p centered at the p-dimensional zero vector such that for $\rho_1, ..., \rho_p \in \Theta_p$, $\Sigma_p = \sum_{r=0}^{\infty} \alpha_r \left(\sum_{h=1}^{p} \rho_h A_h\right)^r$, with $\alpha_r = 1$ for a CAR(p) model, $\alpha_r = r + 1$ for a SAR(p) model. Since $A_1^0, ..., A_1^\xi$ are linearly independent (because, if $A_1^t(i,j) \neq 0$ for some $t = 1, ..., d$, then $A_1^h(i,j) = 0$, for any $h = 0, ..., t-1$), it follows that they form a basis for $Span\{A_0, ..., A_\xi\}$. Thus, $A_0, ..., A_\xi$ are polynomials of maximum degree ξ in A_1, which is a sufficient condition for G to be distance-regular (Bannai and Ito, 1984, pp. 190-192).

Proof of Theorem 4.8: If G is distance-regular with diameter p, then $A_0, ..., A_p$ span an algebra of symmetric matrices; e.g., Bannai and Ito (1984). Since an algebra of symmetric matrices is a quadratic subspace, the sufficiency of the condition in the proposition is a consequence of Proposition

3.5. Conversely, assume that a SAR(p) model on a graph G is a regular exponential family, and let Γ_p denote its covariance matrix. Then, Γ_p^{-1} belongs to a $(p+1)$-dimensional subspace, say Φ_p, of $\mathcal{L}(n)$. It follows that $A_1^h \in \Phi_p$, for any $h = 0, 1, ...$, because there is always an open p-ball Θ_p centered at the p-dimensional zero vector such that, for any $\rho_1, ..., \rho_p \in \Theta_p$, $\Gamma_p^{-1} = \sum_{r=0}^{\infty}(r+1)\left(\sum_{h=1}^{p}\rho_h A_h\right)^r$. But, since the matrices $A_1^0, ..., A_1^p$ are linearly independent, they form a basis for Φ_p, and hence G is distance-regular, by the final argument used in the proof of Theorem 4.7.

References

Albramowitz, A., Stegun, S., 1979. Handbook of Mathematical Functions. Nauka, Moskow.

Amari, S., 1990. Differential-Geometrical Methods in Statistics. Springer, New York.

Anderson, T.W., 1971. The Statistical Analysis of Time Series. Wiley, New York.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer, Dordrecht.

Anselin, L., Smirnov, O., 1996. Efficient algorithms for constructing proper higher order spatial lag operators. Journal of Regional Science 36, 67-89.

Baltagi, B.H., Song, S.H., Koh, W., 2003. Testing panel data regression models with spatial error correlation. Journal of Econometrics 117, 123-150.

Bannai, E., Ito, T., 1984. Algebraic Combinatorics I: Association Schemes. Benjamin/Cummings, London.

Bao, Y., Ullah, A., 2006. Finite sample properties of maximum likelihood estimator in spatial models. Journal of Econometrics, forthcoming.

Besag, J.E., 1972. On the correlation structure of some two-dimensional stationary processes. Biometrika 59, 43-48.

Besag, J.E., 1974. Spatial interaction and the statistical analysis of lattice data. Journal of the Royal Statistical Society B 36, 192-236.

Besag, J.E., Kooperberg, C., 1995. On conditional and intrinsic autoregression. Biometrika 82, 733-746.

Besag, J.E., Tjelmeland, H., 1998. Markov random fields with higher-order interactions. Scandinavian Journal of Statistics 25, 415-433.

Biggs, N.L., 1993. Algebraic Graph Theory, second edition. Cambridge University Press, Cambridge.

Blommestein, H.J., Koper, N.A.M., 1997. The influence of sample size on the degree of redundancy in spatial lag operators. Journal of Econometrics 82, 317-333.

Brouwer, A. E., Cohen, A. M., Neumaier, A., 1989. Distance Regular Graphs. Springer-Verlag, New York.

Champagnat, F., Idier, J., Goussard, Y., 1998. Stationary Markov random fields on a finite rectangular lattice. IEEE Transactions on Information Theory 44, 2901-1916.

Cliff, A. D., Ord, J.K., 1981. Spatial processes: Models and applications. Pion, London.

Conley, T.G., Molinari, F., 2005. Spatial correlation robust inference with errors in location and distance. Journal of Econometrics, forthcoming.

Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.

Cressie, N., Chan, N.H., 1989. Spatial Modeling of Regional Variables. Journal of the American Statistical Association 84, 393-401.

Cvetković, D.M., Doob, M., Sachs, H., 1980. Spectra of Graphs. Academic Press, New York.

Efron, B., 1975. Defining the curvature of a statistical problem (with applications to second order efficiency). Annals of Statistics 3, 1189-1242.

Fiol, M.A., 1997. Some applications of the proper and adjacency polynomials in the theory of graph spectra. The Electronic Journal of Combinatorics 4, #R21.

Gantmacher, F.R., 1974. The Theory of Matrices, vol. II. Chelsea, New York.

Giacomini, R., Granger, C.W.J., 2004. Aggregation of space-time processes. Journal of Econometrics 118, 7-26.

Godsil, C., 1993. Algebraic Combinatorics. Chapman and Hall, New York.

Godsil, C.D., McKay, B.D., 1993. Feasibility conditions for the existence of walk-regular graphs. Linear Algebra and its Applications 30, 51-61.

Huang, J.S., 1984. The autoregressive moving average model for spatial analysis. Australin Journal of Statistics 2, 169-178.

James, A.T., 1957. The relationship algebra of an experimental design, Annals of Mathematical Statistics 28, 993-1002.

Kallenberg, W.C.M., 1981. The shortcoming of locally most powerful tests in curved exponential families. The Annals of Statistics 9, 673-677.

Kass, R.E., Vos, P.W., 1997. Geometrical Foundations of Asymptotic Inference. Wiley, New York.

Kelejian, H.H., Robinson, D.P., 1995. Spatial correlation: a suggested alternative to the autoregressive model, in: L. Anselin, R. Florax (Eds.), New Directions in Spatial Econometrics. Springer Verlag, Berlin.

Kelejian, H.H., Prucha, I.R., 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, International Economic Review 40, 509-533.

Kelejian, H.H., Prucha, I.R., 2004. Estimation of simultaneous systems of spatially interrelated cross sectional equations. Journal of Econometrics 118, 27-50.

Lee, L.F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252-277.

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. Econometrica 72, 1899-1926.

Lee, L.F., 2006. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics, forthcoming.

Lee, J., Kaiser, M.S., Cressie, N., 2001. Multiway dependence in exponential family conditional distributions. Journal of Multivariate Analysis 79, 171-190.

Martellosio, F., 2006. Power properties of invariant tests for spatial autocorrelation in linear regression. University of Amsterdam working paper.

Martin, R.J., 1990. The role of spatial statistical processes in geographical modelling, in Griffith, D.A., (ed.), Spatial Statistics: Past, Present, and Future, Ann Arbor, Michigan, Institute of Mathematical Geography, 109-127.

Moran, P.A.P., 1950. Notes on continuos stochastic phenomena. Biometrika 37, 17-23.

Robinson, P.M., 2006. Efficient estimation of the semiparametric spatial autoregressive model, Cemmap Working Paper.

Seely, J., 1971. Quadratic subspaces and completeness. Annals of Mathematical Statistics 42, 710-721.

Speed, T.P., Kiiveri, H.T., 1986. Gaussian distributions over finite graphs. Annals of Statistics 14, 138-150.

Wall, M.M., 2004. A close look at the spatial structure implied by the CAR and SAR models. Journal of Statistical Planning and Inference 121, 311-324.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 41, 434-449.

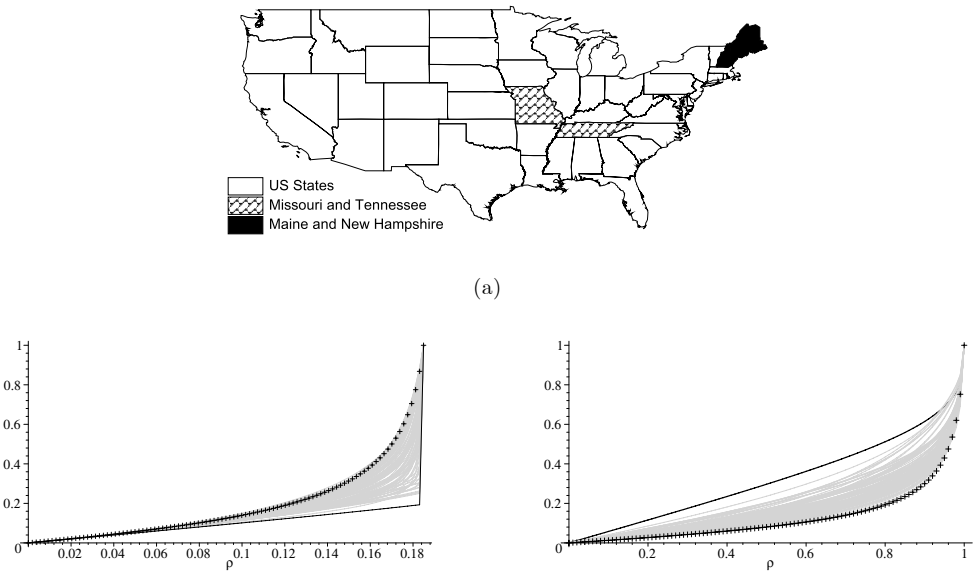


Figure 1: The correlations, as a function of ρ, implied by a CAR(1) model with (b) $W = A$, (c) $W = D^{-1/2}AD^{-1/2}$ between the 107 pairs of contiguous US states, with emphasis on Missouri and Tennessee (crosses) and Maine and New Hampshire (dark solid line).

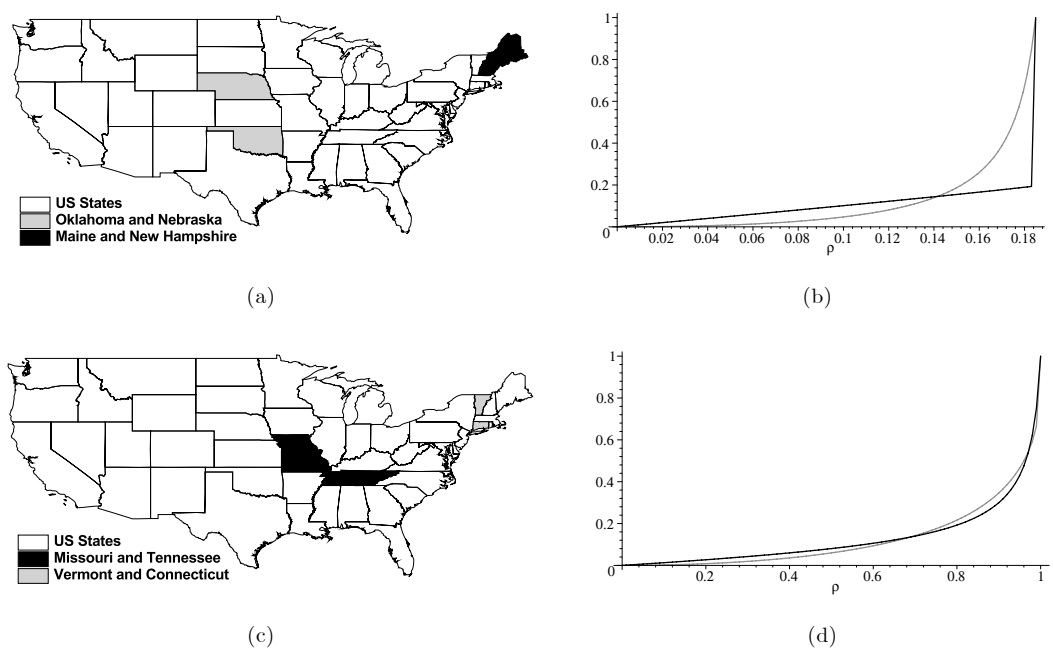
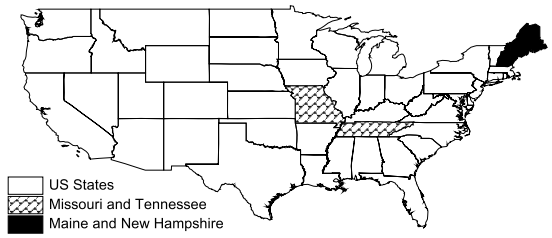
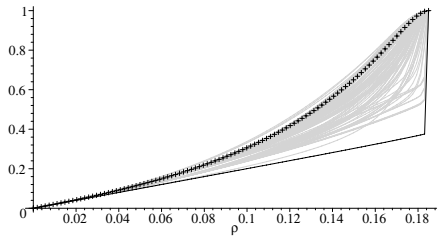


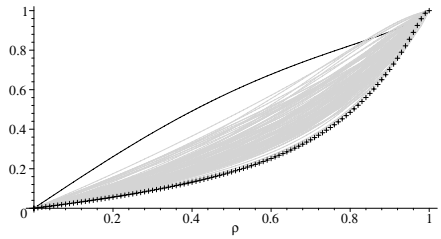
Figure 2: (b) The correlations, as a function of ρ, implied by a CAR(1) model with $W = A$ for a pair of neighbors (darker line; Maine and New Hampshire) and for a pair of non-neighbors (lighter line; Oklahoma and Nebraska). (d) The correlations, as a function of ρ, implied by a CAR(1) model with $W = D^{-1/2}AD^{-1/2}$ for a pair of neighbors (darker line; Missouri and Tennessee) and for a pair of non-neighbors (lighter line; Vermont and Connecticut).



(a)

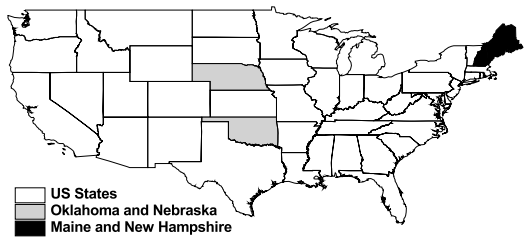


(b)

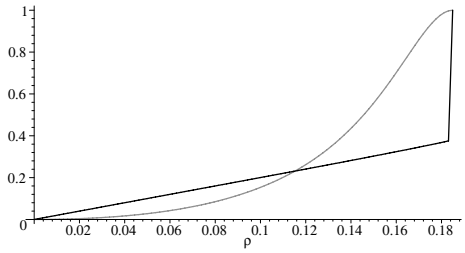


(c)

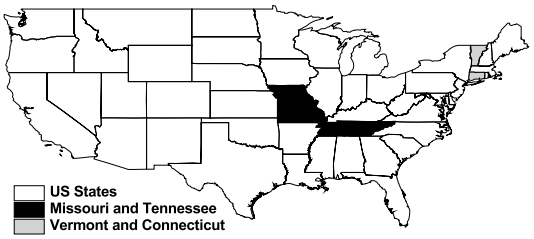
Figure 3: The correlations, as a function of ρ, implied by a SAR(1) model with (b) $W = A$, (c) $W = D^{-1}A$ between the 107 pairs of contiguous US states, with emphasis on Missouri and Tennessee (crosses) and Maine and New Hampshire (dark solid line).



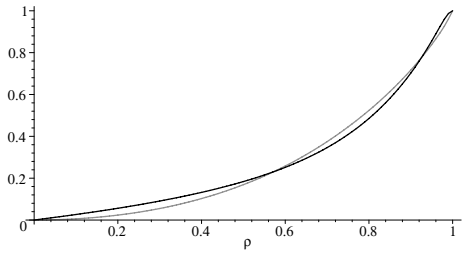
(a)



(b)



(c)



(d)

Figure 4: (b) The correlations, as a function of ρ, implied by a SAR(1) model with $W = A$ for a pair of neighbors (darker line; Maine and New Hampshire) and for a pair of non-neighbors (lighter line; Oklahoma and Nebraska). (d) The correlations, as a function of ρ, implied by a SAR(1) model with $W = D^{-1}A$ for a pair of neighbors (darker line; Missouri and Tennessee) and for a pair of non-neighbors (lighter line; Vermont and Connecticut).

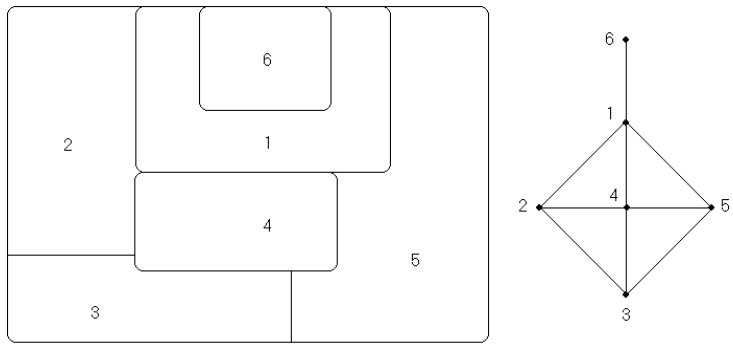


Figure 5: A planar configuration of observational units and a corresponding graph.

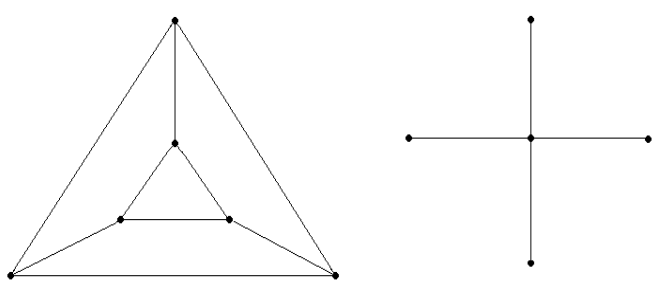


Figure 6: A vertex-transitive graph (left) and an edge-transitive graph (right).