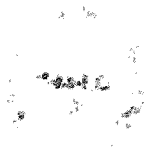


UNIVERSITY OF SOUTHAMPTON

School of Social Sciences

Economics Division

ESSAYS ON THE ECONOMETRIC ANALYSIS OF SPATIAL DATA

by

Federico Martellosio

Thesis for the degree of Doctor of Philosophy

March, 2006

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

SCHOOL OF SOCIAL SCIENCES

ECONOMICS DIVISION

Doctor of Philosophy

ESSAYS ON THE ECONOMETRIC ANALYSIS OF SPATIAL DATA

by Federico Martellosio

This thesis addresses issues in the econometric analysis of data observed over regular or irregular lattices, a lattice being a fixed set of observational units that have some spatial connotation, in a broad sense. As far as the case of data coming from irregular lattices is concerned, we investigate the correlation structure of spatial autoregressive models and we analyse the properties of tests for spatial autocorrelation. As for the case of regular lattices, we focus on statistical issues associated to some matrices, the so-called spatial design matrices, that arise naturally in many inferential problems in the context of isotropic spatial processes defined on uniform grids.

Chapter 1, titled 'The Correlation Structure of Spatial Autoregressions', proposes a novel method to study the properties of spatial autoregressive models defined over irregular lattices. A little graph theory provides simple interpretations of the correlations implied by such models in terms of the walks connecting two vertices, and reveals the statistical consequences of the presence of symmetries or regularities in the configuration of the observational units.

Chapter 2, titled 'Properties of Invariant Tests for Spatial Autocorrelation in the Linear Regression Model', sheds some new light on how the power of some popular tests for spatial autocorrelation in the errors of a linear model is affected by the matrix of regressors and by the assumed spatial structure. Conditions for unbiasedness and monotonicity of the power function of the tests are studied.

Chapter 3, titled 'Spatial Design Matrices and Associated Quadratic Forms: Structure and Properties', provides a complete characterization of the structure of spatial design matrices. The structural results are applied to study the statistical properties of statistics associated to the spatial design matrices, in particular of the classical variogram estimator, under several assumptions about the actual variogram.

Chapter 4, titled 'Circular Approximation to the Design Matrices of Isotropic Spatial Processes', develops an approximation to the spatial design matrices, with the aim of alleviating the computational effort required to obtain the cumulants of some of the statistics discussed in Chapter 3. The performance of the approximation is discussed in the case of independent data and of second-order stationary and isotropic processes.

Chapter 5 concludes by summarizing the main achievements of the thesis and suggesting directions for further research.

Contents

List of Figures

List of Tables

Acknowledgments

I am deeply indebted to my supervisor, Grant Hillier, for his guidance, help, encouragement, and for all the inspiring discussions.

I also thank all the members of the Economics Division of the University of Southampton, and I would like to mention in particular John Aldrich, Anurag Banerjee, Raymond O'Brien and Jean-Yves Pitarakis.

Thanks to the administrative staff of the School of Social Sciences, in particular to Gill Crumplin, Anne Edwards, and Fatima Fauzel for their kindness.

Thanks to Eleonora for her support and for reading the whole thing.

Financial support in the form of grants from ESRC (No. R42200134323) and the University of Southampton is gratefully acknowledged.

Preface

Consider a statistical process of the type $\{y_i, i \in U\}$, with y_i a univariate random variable and U a fixed and finite set of observational units. We call such a process *spatial* if U has some spatial, in a very broad sense, connotation. For instance, U could represent some physical space such as the set of regions of a country, or some abstract space such as a subset of a d-dimensional Euclidean space. For many of the problems considered in this thesis, however, what "space" is in question is not a fundamental issue. What is crucial is that: (i) the observational units are fixed, i.e., we are concerned with the joint distribution of the variables y_i given their locations, and not with the spatial variation of the locations themselves; (ii) it is assumed that the location in space of the observational units affects, in some way, the joint distribution of the variables y_i.

Data that can be assumed to be realizations of spatial processes of the type just described are often referred to as *lattice data*, where a lattice can be regular, for instance a uniform d-dimensional grid, or irregular. The development of techniques for drawing inference from lattice data has been a relevant activity in statistics since, at least, the fifties. Some milestone contributions are Moran (1950), Whittle (1954), Besag (1974) and Bartlett (1978). Conversely, in econometrics the attention to inferential procedures for spatial data has been long obfuscated by the dominating time-series perspective. Some early work in the field of "spatial econometrics" is represented by Paelinck and Klaassen (1979) and Anselin (1988). In very recent years, there has been an explosion of interest in spatial models in econometrics, with work in spatial econometrics being published by leading journals in empirical and theoretical econometrics (for instance, Conley, 1999, Kelejian and Prucha, 2001, Conley and Topa, 2002, Lee, 2002 and 2004). Such an explosion has been prompted not only by numerous empirical analyses showing the importance of adopting spatial statistical models in many fields of economics, but also by the development of economic theories in which some physical or social distance plays an explicit role in determining the interaction among agents (see, for

instance, the references in Anselin and Bera, 1998).

This thesis consists of four main chapters dealing with issues in the econometric, and more generally statistical, analysis of spatial data observed over regular or irregular lattices. Although they are related, the four chapters take the form of four papers that can be read independently. In the following, we summarize the focus of each paper and explain the relations among them.

Chapter 1 is devoted to the analysis of properties of the most popular models for lattice data, the so-called spatial autoregressive models. In particular we study the correlation structure of these models when the set U of observational units is completely unrestricted. An interpretation of the correlations in terms of walks between the vertices of a graph is developed. The interpretation enables us to explain some known peculiarities of the models and to prove some new results. Since in this chapter we are concerned with correlation properties, we consider only zero-mean models.

Chapter 2 studies power properties of invariant tests for spatial autocorrelation in the context of the linear regression model. The alternative hypothesis of spatial autocorrelation is provided by the models studied in the previous chapter. The main purpose of the chapter is to clarify the role played by the spatial structure of the set U of observational units and by the matrix of regressors in affecting the power of invariant tests for autocorrelation.

While in the first two chapters the focus is on irregular lattices, the following two chapters are concerned with spatial data observed over regular lattices, namely d-dimensional rectangular uniform grids.

Chapter 3 analyses the structure of the so-called spatial design matrices. The quadratic forms associated to these matrices are recurrent objects in inferential procedures in the context of isotropic processes defined on uniform grids. We give generating functions for the cumulants of the quadratic forms associated to the spatial design matrices. An application to the study of the exact properties of the classical variogram estimator is also discussed.

Chapter 4 is intimately related to the material of Chapter 3. In this chapter we provide an approximation to the spatial design matrices, with the aim of reducing the computational effort required to obtain the cumulants of the quadratic forms associated to the spatial design matrices. The performance of the approximation is discussed under several assumptions.

Chapter 5 summarizes the main findings of the thesis and discusses possible extensions.

One theme that unifies the four main chapters is the attempt to understand how the spatial structure of the set U of observational units affects the properties of statistical entities

relative to data observed over U. The statistical entities are spatial autoregressive models in Chapter 1; invariant tests for spatial autocorrelation in Chapter 2; quadratic forms in the spatial design matrices and approximations to them in, respectively, Chapters 3 and 4.

Chapter 3 is a slightly modified version of the paper Hillier and Martellosio (2006). One significant modification is the replacement of Lemma 5 of Hillier and Martellosio (2006) by the more general Lemma 3.3.4. Chapter 4 is also derived from joint work with Grant Hillier.

Throughout the thesis, items such as theorems, propositions, examples, remarks, etc. are numbered consecutively within each section, by chapter and section, so that Theorem 1.2.3, for instance, denotes the third of those items in Section 2 of Chapter 1. Equations are numbered consecutively within each chapter; for instance, (2.11) refers to the eleventh equation of Chapter 2.

Chapter 1

The Correlation Structure of Spatial Autoregressions

Abstract

Conditional Autoregressive (CAR) and Simultaneous Autoregressive (SAR) models are used in many scientific fields to represent interaction among spatial data. They are both usually specified on the basis of p weights matrices, which are chosen to reflect a priori information on the associations among the random variables under analysis. This paper studies correlation properties of CAR and SAR models for an arbitrary spatial configuration of the units where the data are observed. We derive new results and interpretations concerning the one-parameter models ($p = 1$). In particular, a little graph theory provides simple interpretations of the correlations in terms of the walks connecting two vertices, and reveals the statistical consequences of the presence of symmetries or regularities in the configuration of the observational units. Extensions to multi-parameter models ($p > 1$) are also considered.

Keywords: spatial autoregressions; weighted graphs; exponential families; quadratic subspaces; association schemes.

1.1 Introduction

As originally explained by Brook (1964), time series autoregressive processes can be generalized to a spatial setting in two different ways. The two approaches give rise to Conditional Autoregressions (CAR) and Simultaneous Autoregressions (SAR). Since their first formulation—attributable, respectively, to Besag (1974) and Whittle (1954)—these models have been used extensively in many scientific fields. Some of the most common applications are in disease mapping and image analysis for the CAR models, in econometrics and regional science for the SAR models.

The main purpose of this paper is to analyze the correlation structure of CAR and SAR models defined over an arbitrary set of observational units. Both classes of spatial autoregressive models were first developed to analyze data on *regular lattices*, which arise for instance in agricultural field experiments or when decomposing an image in pixels. Nevertheless, they soon started to be used also for the *irregular lattices* typical of non-experimental contexts (e.g., Besag, 1975). Recent literature has pointed out that in the latter case the models exhibit some undesirable or unexpected properties (e.g., Besag and Kooperberg, 1995, and Wall, 2004). In fact, the theoretical analysis of the models defined on irregular lattices can be complicated and it is fair to say that some of the properties of spatial autoregressions are not completely understood by practitioners. This is unfortunate, because specifying the models may be difficult if their properties are not clear.

The correlation structure of spatial autoregressions is well-understood in the case of regular lattices (e.g., Besag, 1972 and 1981). Here, we largely focus on the case of an unrestricted spatial configuration of the observational units, but we also consider the effects of regularities in the neighborhood structure of the units. The reason why, in general, the correlation structure of spatial autoregressions does not lend itself to a straightforward analysis is that such models—in their most common parametrization—imply a simple structure for the inverse of a covariance matrix, and not for the covariance matrix directly. Yet, these models have important advantages—for instance, they are exponential families, they use a parsimonious parametrization, they have a clear conditional independence structure and, for the CAR models, they lead to simple Markov Chain Monte Carlo analysis—and are so popular that a detailed understanding of the behavior of the correlations they imply on irregular lattices seems to be needed.

Large part of our analysis will be 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 which are Markovian with respect to a given graph; Besag, 1974), but is less common to discuss the properties of the covariance matrices of CAR and SAR models. There are several advantages to adopting a graph theoretic perspective in the analysis of spatial autoregressions. 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 in clarifying 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 divided into three main sections and a discussion. In Section 1.2 we first briefly review the general definitions of CAR and SAR models, and then discuss some issues concerning their parameterization. Obviously, the choice of parametrization is crucial when discussing the correlation structure of a family of distributions. We show, in particular, that the parameterizations commonly employed for CAR and SAR models, and that we also employ in this paper, have the advantage of making CAR and SAR models exponential families with low-dimensional minimal sufficient statistic. While, in their common parametrization, CAR models are always regular exponential families, SAR models are generally curved (in the sense of Efron, 1978), and are regular exponential families only if the assumed spatial structure is highly regular. Section 1.3 is the central part of the paper and is mainly devoted to a detailed study of the correlation structure of the one–parameter models defined on general graphs. We consider more briefly multi-parameter models, that are much less popular in applications. Extending the parameter space of the first-order models is a natural way to try to overcome some of the limitations of the first-order models, yet maintaining some of their advantages. After all, it would be surprising if a family of distributions indexed by a single parameter (or two, including one scaling the covariance matrix) were particularly useful to describe the interaction structure of variables observed over an highly irregular lattice! In Section 1.4 the special features of spatial autoregressive models on graphs exhibiting some regularities or symmetries are discussed. In particular, we give the conditions for two pairs of variables to have the same correlations. 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). We conclude the paper with a summary of the main results. All proofs are relegated to the Appendix.

1.2 Spatial Autoregressions and Their Parametrization

This section presents the theoretical framework of the paper. Section 1.2.1 defines the models and is mostly review. Section 1.2.2 discusses some issues, that are important in the rest of the paper, concerning the most common parametrization of the models.

1.2.1 The Models

Consider a *fixed* and *finite* set of n observational units, for instance the set of regions of a country. For the sake of 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 and SAR models are families of distributions for the n-dimensional random vector $y = (y_1, ..., y_n)'$, where y_i is the random variable associated to the i-th observational unit. They are specified on the basis of two $n \times n$ real matrices S and C chosen to reflect a priori information on relations among the n observations (see, for instance, Cressie, 1993). For instance, the (i,j)-th entry of C or S may be taken to be a certain function of some distance between the i-th and the j-th observational units. Usually, such matrices are sparse, which, in the example just mentioned, requires that the (i,j)-th entry of C or S is set to zero when the distance between i and j is larger than some threshold. The matrices S and C depend on known constants and unknown parameters, and are such that the matrices $I - S$ and $I - C$ are nonsingular, with I denoting the $n \times n$ identity matrix. In this paper, as in most of the theoretical and empirical literature on spatial autoregressive models, we confine ourselves to Gaussian models ($N_v(\cdot, \cdot)$ will denote the v-variate Gaussian distribution, dropping the subscript in the univariate case). Only the zero-mean case is considered, because our focus is on the second-order structure of the models.

Let M be an $n \times n$ diagonal matrix with positive diagonal entries. A CAR model, also known as *auto-normal* scheme, is specified through the n conditional distributions

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

which, provided that $(I - C)^{-1}M$ is symmetric and positive definite (hereafter abbreviated

to p.d.), yield the joint distribution (see Besag, 1974)

$$y \sim N_n\left(0, (I - C)^{-1}M\right) \tag{1.1}$$

(here and throughout the i-th entry of a vector r is denoted by r_i, the (i,j)-th entry of a matrix R by $R(i,j)$; all vectors and matrices considered in this paper are real-valued).

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

$$y = Sy + \varepsilon,$$

where ε is an n-dimensional vector with $\varepsilon \sim N_n(0, \Lambda)$, Λ being a diagonal matrix with positive diagonal entries. The resulting joint distribution is

$$y \sim N_n\left(0, (I - S)^{-1}\Lambda(I - S')^{-1}\right). \tag{1.2}$$

Note that, for the CAR specification to be valid, the matrix C must have only zero entries on its main diagonal and must satisfy the property that $C(i,j) = 0$ if and only if $C(j,i) = 0$. It is common, although not necessary, to impose these two restrictions on the matrix S as well, and this is what we do in the present paper, unless otherwise specified. Note that the symmetry of the zero entries of S entails that time series unilateral autoregressive models are not in the class of SAR models here considered. By the symmetry of their zero entries, C and S can be assumed to be *irreducible* (see, for instance, Gantmacher, 1974, Ch. 13) without loss of generality, because otherwise there would exist a permutation of the index set $\{1, 2, ..., n\}$ such that the covariance matrices of the models are block diagonal, i.e., the models could be decomposed into the product of at least two models.

It is worth stressing that we are only assuming that the zero entries of C and S are symmetric and not that C and S are symmetric. Indeed, the latter assumption would be too restrictive in some applications, especially in the case of irregular lattices. For the CAR model, this is because if C is symmetric and irreducible then M in (1.1) is a scalar multiple of I. This is because if C is symmetric it must hold that $(I - C)^{-1}M = M(I - C)^{-1}$; but if C is irreducible then $(I - C)^{-1}$ is entrywise positive (e.g., Gantmacher, 1974, p. 69) and the only diagonal matrix with positive diagonal entries that commutes with a positive matrix is a scalar multiple of I. Therefore, symmetry of C implies that the conditional variances are constant, which may not be desirable in applications (cf. Section 1.3.2). For the SAR model, symmetry of S is even more restrictive, as the covariance matrix of a SAR model is symmetric for any S and any Λ.

By far, the most common way of parametrizing the models is to take the matrices C and S to be linear in $p \geq 1$ parameters and M and Λ linear in one parameter (we will see below that there are strong reasons for adopting such a parametrization). More precisely, a CAR model is commonly specified by setting $C = \sum_{l=1}^{p} \rho_l W_l$ and $M = \tau^2 L$, where τ^2, $\rho_1, ..., \rho_p$ are functionally independent unknown parameters, L is a known diagonal matrix with positive diagonal entries and $W_1, ..., W_p$ are non-zero linearly independent known matrices. The parameter space of the model is defined by $\tau^2 \in \mathbb{R}^+$ and $\rho_1, ..., \rho_p$ belonging to the subset of $\mathbb{R}^p$ where the covariance matrix of the CAR model is p.d. Similarly, for a SAR model, $S = \sum_{l=1}^{p} \rho_l W_l$ and $\Lambda = \sigma^2 V$ (using the same notation as for C, for the sake of convenience), where σ^2, $\rho_1, ..., \rho_p$ are functionally independent unknown parameters, V is a known diagonal matrix with positive diagonal entries and $W_1, ..., W_p$ are non-zero linearly independent known matrices. The parameter space of the model is given by $\sigma^2 \in \mathbb{R}^+$ and $\rho_1, ..., \rho_p$ belonging to the subset of $\mathbb{R}^p$ where the covariance matrix of the SAR model is p.d. The matrices W_l are often called *(spatial) weights matrices*. Clearly, the weights matrices inherit the restrictions imposed on C and S, and therefore each W_l must have zero diagonal and must satisfy the property that $W_l(i,j) = 0$ if and only if $W_l(j,i) = 0$. Moreover, when used in the CAR specification, each product $L^{-1}W_l$ must be symmetric.

We make the additional assumption that the weights matrices are (entrywise) nonnegative. This 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.

From now on and unless otherwise specified, we will reserve the terms "CAR models" and "SAR models" for the families of distributions (1.1) and (1.2), when they are indexed by the parameters $\rho_1, ..., \rho_p$ and τ^2 or σ^2 as described above. Their covariance matrices will be denoted by $\Sigma_{c,p}$ and $\Sigma_{s,p}$ respectively.

1.2.2 Consequences of the Parametrization

In this section we discuss some consequences of the parametrization of CAR and SAR models. In particular, we are interested in how CAR and SAR models fit in the theory of *exponential families*. This is important, because of the many nice statistical properties satisfied by such families and because it allows us to start discussing the consequences of using highly

structured weights matrices. For details concerning exponential families, see Amari (1990) and Kass and Vos (1997). Here we only remind the reader that a family of densities is said to be an exponential family if its elements are representable as

$$pdf(y;\theta) = \exp\left\{\sum_{l=1}^{k} \eta_l(\theta) s_l(y) - \kappa(\theta)\right\} h(y), \tag{1.3}$$

with respect to some dominating measure on the sample space. Assuming that k is the smallest integer such that (1.3) 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 $\exp\{\sum_{l=1}^{k} \eta_l(\theta) s_l(y) - \kappa(\theta)\}$ is finite. Under some 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*.

CAR models are families of Gaussian distributions where the inverse of the covariance matrix—the precision matrix—has a linear structure (in the parameters $1/\tau^2, \rho_1/\tau^2, ..., \rho_p/\tau^2$). A discussion of these general models from a time series perspective can be found in Anderson (1971), Chapter 6. One important consequence of the linearity of the precision matrix is stated in the following lemma.

Lemma 1.2.1 *A CAR model is a regular exponential family, with minimal sufficient statistics* $y'L^{-1}W_l y$, $l = 0, 1, ..., p$.

Henceforth, $\mathcal{L}(n)$ denotes the vector space of all $n \times n$ real symmetric matrices. Since for the general model (1.1) to admit a representation (1.3) it is necessary that $M^{-1}(I - C)$ belongs to a linear subspace of $\mathcal{L}(n)$ of dimension equal to the number of parameters of the model, it is easily seen that the parametrization of the models in terms of $\rho_1, ..., \rho_p, \tau$ is virtually the only one—up to a diffeomorphism (see Kass and Vos, 1997)—such that CAR models are regular exponential families (assuming that the parameters in M are functionally independent of those in C).

Conversely, SAR models are regular exponential families only in very special circumstances. Setting, for notational convenience, $\rho_0 = -1$ and $W_0 = I$, it is easily seen that a SAR model is embedded in a regular exponential family with canonical parameters

$$-\frac{\rho_{l_1}\rho_{l_2}}{2\sigma^2}; \quad l_1 = 0, 1, ..., p, \; l_2 = 0, 1, ..., l_1,$$

and sufficient statistics

$$y'B_{l_1,l_2}y; \quad l_1 = 0,1,...,p, \; l_2 = 0,1,...,l_1,$$

where

$$B_{l_1,l_2} = W'_{l_1}V^{-1}W_{l_2} + (1-\delta_{l_1,l_2})W'_{l_2}V^{-1}W_{l_1},$$

δ_{l_1,l_2} being the Kronecker delta. Note that such an embedding family is not quite a CAR model (in the parametrization we have adopted for CAR models) because the matrices B_{l_1,l_2}, $l_1 = l_2$, do not have zero diagonals.

Let Δ_p be the subspace of $\mathcal{L}(n)$ spanned by the matrices B_{l_1,l_2}, for $l_1 = 0,1,...,p$ and $l_2 = 0,1,...,l_1$. The dimension of Δ_p, which we denote by $\dim(\Delta_p)$, is the dimension of a minimal sufficient statistic for a SAR model. Since the matrices $B_{l,0}$, $l = 0,1,...,p$, are linearly independent (because the matrices W are), the model is a curved exponential family if $\dim(\Delta_p) > p+1$, a full exponential family if $\dim(\Delta_p) = p+1$ (see Amari, 1990, p. 109).

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

Lemma 1.2.2 *For a SAR model to be a full exponential family, the matrix* $S = I - \sum_{l=1}^{p} \rho_l W_l$ *must be full.*

(Note that if a SAR model is a full exponential family than it is a regular exponential family by the same argument as the one used in the proof of Lemma 1.2.1.) Since the SAR models used in applications are constructed by specifying a sparse matrix S, Lemma 1.2.2 implies that such models are curved. The curvature of a statistical model has well-known consequences for inferential procedures. 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 (Kass and Vos, 1997, Chapter 3).

The next natural step would be to study measures of curvature of SAR models. Differential geometric measures of curvature (such as the Efron curvature, see Amari, 1990) can certainly be analyzed, but this goes beyond the scope of the present paper. Here, we limit ourselves to the following lemma, where the notation (p_1, p_2)-curved exponential family is used to indicate that an exponential family has p_1-dimensional parameter space and p_2-dimensional minimal sufficient statistic.

Lemma 1.2.3 *Suppose that* $I - \rho W$ *is not full. Then, a first-order SAR model with weights matrix* W *is a* $(3,2)$*-curved exponential family.*

Note that Lemma 1.2.3 does not hold for unilateral time series autoregressive models, which, as we have already mentioned, are not in the class of SAR models considered in the present paper. Such models are regular exponential families when, for instance, (the analog of) W is an orthogonal matrix.

The condition in Lemma 1.2.2 is necessary, but by no means sufficient. Indeed, for a SAR model to be a regular exponential family, S must not only be full, but it must also have a very special structure. This is particularly transparent in the special case when the spatial weights matrices are symmetric and $V = I$. Before stating the next result, we need to define the concept of quadratic subspace of symmetric matrices, first introduced by Seely (1971) and subsequently extensively used in the literature on the analysis of variance components.

Definition 1.2.4 *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 be the subspace of $\mathcal{L}(n)$ spanned by the matrices $I, W_1 ..., W_p$, we have:

Lemma 1.2.5 *Assume that $W_1, ..., W_p$ are symmetric and $V = I$. Then, a SAR model is a curved exponential family, unless 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 Section 1.4 that in order to obtain a quadratic subspace, a very high level of regularity must be present in the spatial configuration of the observational units and a specific number of weights matrices must be used. Note that Lemma 1.2.5 gives a necessary and sufficient condition for the simultaneous and the conditional approaches to the specification of autoregressive models to be equivalent when $W_1, ..., W_p$ are symmetric and $V = I$. This is because, under the stated condition, $\Sigma_{s,p}$ and $\Sigma_{c,p}$ belong to the same subspace Ψ_p if and only if Ψ_p is a quadratic subspace.

1.3 The Correlation Structure of Spatial Autoregressions

In the context of spatial autoregressions, the neighborhood structure of the set of the n observational units can be conveniently represented by a graph. This corresponds to identifying the index set of the random vector y with the vertex set of a graph and to defining the adjacency matrix (as defined below) using a priori information about relationships among the random variables in y.

The precise definition of the graphs considered in the present paper is given in Section 1.3.1, where we also summarize a series of well-known graph theoretic definitions and results. In Sections 1.3.2 and 1.3.3 we study in detail the correlation properties of first-order spatial autoregressive models with arbitrary weights matrices. We denote the first-order models by CAR(1) and SAR(1) models and, for convenience, we drop the subscript from the single weights matrix W_1 of CAR(1) and SAR(1) models. In Section 1.3.4 we consider extensions of one-parameter models on graphs to multi-parameter models.

1.3.1 Graph Theoretic Notions

Throughout the paper and unless otherwise specified, a *graph* $G = (V(G), E(G))$ is a pair of a finite non-empty set $V(G)$ of *vertices* and a non-empty set $E(G)$ of two-element subsets, called *edges*, of $V(G)$. A graph has associated with it a weight function $z : V(G) \times V(G) \to \mathbb{R}$ such that $z(i,j) = 0$ if and only if $(i,j) \notin E(G)$.

This definition, together with the assumption—made to avoid trivial cases—that a graph G is *connected* but not *complete* (i.e. that every vertex appears in at least one edge and that $E(G)$ does not contain all the two-element subsets of $V(G)$), completely characterizes the graphs considered in the paper. In graph theoretic terminology, we would say that such graphs are *finite*, *undirected* (a graph is said to be *directed* if its edges are defined as ordered pairs of vertices), *weighted* and do not contain *loops* or *multiple edges*. An exposition of most of the graph theoretic concepts presented in this section can be found in Biggs (1993).

The *adjacency matrix* of a graph G is the $|V(G)| \times |V(G)|$ matrix $Z = Z(G)$ with entries $Z(i,j) = z(i,j)$. The assumption of connectivity of a graph implies that its adjacency matrix is irreducible. Note that the the matrices C and S, as defined in Section 1.2.1, or the spatial weights matrices W, can always be regarded as the adjacency matrix of a graph.

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

A *walk* is a sequence of vertices $(i_0, i_1, ..., i_r)$ such that two consecutive vertices in the sequence form an edge. Such a walk is sometimes referred to as an r-walk, r being the *length* of the walk. A walk is said to be *closed* if $i_0 = i_r$. 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. Note that if $(i_0, i_1, ..., i_r)$ is a walk, then the reverse sequence $(i_r, i_{r-1}, ..., i_0)$ is a walk too. When we speak of walks between (or connecting) two vertices, it is with the

understanding that we are not double-counting a walk and its reverse, so that, for instance, the number of r-walks between two vertices i and j is equal to the number of r-walks from i to j (or from j to i).

We can now define the *distance* $d(i,j) = d_G(i,j)$ between any two vertices i and j of G to be the length of the shortest path in G joining i and j. This is the distance that, as we shall see shortly, turns out to be the relevant one in interpreting CAR and SAR models. The largest distance between any two vertices of a graph G is called the *diameter* of G, and will be denoted by $d = d(G)$. It is then natural to define the *distance matrices* $A_l = A_l(G)$, for $l = 0, 1, ..., d$, such that $A_l(i,j)$ equals 1 if $d(i,j) = l$, 0 otherwise, for each $i, j \in V(G)$. Note that the distance matrices are disjoint symmetric $(0,1)$ matrices summing to the $n \times n$ matrix of all ones. Also, note that $A_0 = I$ and that A_1—which we will usually denote simply by A—is the binary version of Z. In the special case of an *unweighted* graph, i.e. a graph having all weights 0 or 1, $Z = A$.

The notion of bipartite graph is critical to understanding some aspects of our analysis of spatial autoregressions. 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 (see, e.g.,). A necessary and sufficient condition which is useful to check for bipartiteness is the absence of cycles of odd length. This suggests that "regular" graphs such as those of any rectangular lattice are bipartite, whereas "irregular" graphs such as those of geographical maps of regions are generally not bipartite.

The *spectrum* of a graph G is the set of eigenvalues, repeated according to their multiplicities, of A. Saying that the spectrum of a graph is symmetric about zero if it contains an eigenvalue λ whenever it contains $-\lambda$ with the same multiplicity as λ, we have (e.g., Cvetković *et al.* (1980), Theorem 3.11):

Lemma 1.3.1 *A graph is bipartite if and only if its spectrum is symmetric about zero.*

Next, we consider the subclass of graphs which are implicitly defined by the CAR specification. These are the graphs whose adjacency matrix becomes symmetric after premultiplication by a diagonal matrix with positive diagonal entries (note that such a condition does not pose any restriction on the choice of an edge set, but it does pose restrictions on the choice of the weight function). For a diagonal matrix with positive diagonal entries M, we denote by $\Gamma(M)$ the class of graphs having adjacency matrix Z such that $M^{-1}Z$ is symmetric. For a graph $G \in \Gamma(M)$, we define the *weight of a walk* $(i_0, i_1, ..., i_r)$ as the product $\Pi_{i=0}^{r-1} Z(i_i, i_{i+1})$

of the weights of its steps, times $M(i_r, i_r)$. The factor $M(i_r, i_r)$ serves to ensure that a walk $(i_0, i_1, ..., i_r)$ in a graph $G \in \Gamma(M)$ has the same weight as the reverse walk $(i_r, i_{r-1}, ..., i_0)$, which is convenient later. Given the notion of weight of a walk we have just given, it is easy to check that, for a graph $G \in \Gamma(M)$, the (i,j)-th entry of $Z^r M$ represents the sum of the weights of all the r-walks connecting i and j. In the particular case of an unweighted graph, $Z^r(i,j)$ simply represents the number of r-walks from i to j.

Having reviewed the necessary graph theoretic tools, we can now move to consider the classes of CAR and SAR models defined on general graphs, or—in other words—the classes of CAR and SAR models with arbitrary weights matrices. Models on graphs satisfying some regularities or symmetries will be considered in Section 1.4.

1.3.2 CAR(1) Model

In this section we study the covariance structure of the CAR(1) model, that is, of the family of n-variate normal distributions with covariance matrix $\Sigma_{c,1} = \tau^2(I - \rho W)^{-1}L$. To such a specification we associate, as discussed above, the graph with adjacency matrix ρW. Without loss of generality, we fix $\tau^2 = 1$. To underline the dependence on ρ, the covariance between any two variables y_i and y_j in a CAR(1) model will be denoted by $\gamma_{i,j}(\rho)$.

Two subfamilies of the general CAR(1) model, corresponding to two simple choices of the weighting function of a graph, are particularly popular in applications (cf. Sun *et al.*, 1999, where the two subfamilies are called Model 1A and Model 2).

Model 1. The weights matrix W is obtained by row-standardization of the first distance matrix A. That is, if D denotes the diagonal matrix with i-th diagonal entry equal to the degree d_i of the vertex i, then Model 1 is obtained by setting $W = D^{-1}A$ and $L = D^{-1}$, which results in the covariance matrix $\Sigma_{c,1} = (D - \rho A)^{-1}$.

Model 2. The weights matrix is $W = A$ (or, equivalently, the graph is unweighted), which results in $\Sigma_{c,1} = (I - \rho A)^{-1}$.

It should be noted that there are advantages and disadvantages to both models. A relative advantage of Model 1 is that such a specification entails that the conditional variances $M(i,i)$ are not constant, but are inversely related to the number of neighbors d_i, which may be appropriate in some applications (Martin, 1990). Conversely, a relative advantage of Model 2 is that it reduces to a spherical distribution when ρ vanishes, i.e., the model includes the null hypothesis of i.i.d. data. Note that the two models are equivalent—up to a reparametrization—if and only if G is degree-regular.

Also, it is worth clarifying from the outset the interpretation of the parameter ρ in a CAR(1) model. Denoting by ρ^{ij} the partial correlation coefficient between y_i and y_j, it is easily seen that $\rho = \rho^{ij}[W(i,j)W(j,i)]^{-1/2}$, for any pair i, j of neighbors. Thus, in a CAR(1) model ρ is equal to a standardized version of the correlation coefficient in the conditional distribution of any pair of neighbors (given the rest of the variables). In particular, for any pair i, j of neighbors, we have $\rho = (d_i d_j)^{1/2}\rho^{ij}$ in Model 1 and $\rho = \rho^{ij}$ in Model 2.

We now present a graph theoretic interpretation of the second order structure of CAR(1) models. The restriction that $L^{-1}W$ is symmetric entails that the eigenvalues of W are real, because $LL^{-1}W = W$ is similar to the symmetric matrix $L^{1/2}L^{-1}WL^{1/2}$. We denote such eigenvalues by $\lambda_1, ..., \lambda_n$, labelled in non-decreasing order of magnitude. Note that $\lambda_1 < 0$ and $\lambda_n > 0$, because $tr(W) = \sum_{i=1}^{n} \lambda_i = 0$ and at least one eigenvalue does not vanish (since W is irreducible). It follows that the parameter space of the model, i.e. the set of parameters values such that $\Sigma_{c,1}$ is p.d., is defined by $\lambda_1^{-1} < \rho < \lambda_n^{-1}$. By the Perron-Frobenius theorem for nonnegative irreducible matrices, $\lambda_n \geq |\lambda_i|$, $i = 1, ..., n-1$. Now, for a square matrix Q, $\sum_{r=0}^{\infty} Q^r = (I-Q)^{-1}$ if and only if all the eigenvalues of Q are smaller than one in absolute value (here, as usual, $Q^0 = I$). In our case, when $|\rho| < \lambda_n^{-1}$, $(I-\rho W)^{-1}$ admits the convergent power series representation $\sum_{r=0}^{\infty} (\rho W)^r$ and thus, on noting that $W^r(i,j) = 0$ for $r < d(i,j)$, we have

$$\gamma_{i,j}(\rho) = L(j,j) \sum_{r=d(i,j)}^{\infty} \rho^r W^r(i,j), \tag{1.4}$$

for $i, j = 1, ..., n$. Recall from the previous subsection that for a graph G with adjacency matrix C (C being the matrix in equation (1.1)), $L(j,j)C^r(i,j)$ represents the sum of the weights of all the r-walks between i and j. Representation (1.4) makes clear how the covariance structure of the CAR(1) model is determined by the connectivity properties of the underlying graph, in the region $|\rho| < \lambda_n^{-1}$.[1] Namely, in a CAR(1) model the covariance $\gamma_{i,j}(\rho)$ is determined by the total weight of all the walks connecting i and j, or, equivalently, *any walk connecting i and j contributes its weight to $\gamma_{i,j}(\rho)$*. (Such an interpretation holds for any matrix C, but some of the results to be derived below are specific to the parametrization $C = \rho W$.) We will shortly see that the fact that all the walks, rather than only the paths or some other sequences of vertices, between i and j contribute to $\gamma_{i,j}(\rho)$ has important consequences on the correlation properties of CAR(1) models.

Because of the simple forms of the weights of the walks implied by Models 1 and 2, the

[1] When the graph is non-bipartite, that is when $|\lambda_1| < |\lambda_n|$, there is a region of the parameter space where the interpretation just given does not hold. However, the representation (1.4) is always valid for positive ρ.

interpretation of the covariance structure is particularly simple in such models. In Model 2, the contribution to $\gamma_{i,j}(\rho)$ of any r-walk between i and j is ρ^r. Such a contribution is strictly decreasing in r, because $|\rho| < \lambda_n^{-1}$ for representation (1.4) to exist and λ_n cannot be smaller than the minimum row sum of A (e.g., Gantmacher, 1974, p. 63). In Model 1, conversely, the contribution of a walk depends not only on the length of the walk, but also on the vertices that the walk visits. More precisely, the contribution to $\gamma_{i_0,i_r}(\rho)$ of a walk $(i_0, i_1, ..., i_r)$ is $\rho^r / \prod_{l=0}^{r} d_{i_l}$. For a fixed r, such a contribution is large if the walk passes through vertices with a small number of neighbors. When r increases, the contribution of an r-walks tend to decrease, because $|\rho| < 1$ (as $|\rho| < \lambda_n^{-1}$ for the representation (1.4) to exist and $\lambda_n = 1$ for a row-stochastic matrix). The main distinction between Model 1 and Model 2, in terms of their graph theoretic interpretation, is perhaps clearer when considering the contribution to $\gamma_{i,j}(\rho)$ of an edge (i.e. a 1-walk) between two first-order neighbors i and j. Such a contribution is constant in Model 2, whereas it is inversely related to the number of neighbors of i and j in Model 1. This suggests that Model 1 is appropriate in cases when the resource determining the influence amongst is a limited one; for example, transportation amongst geographical units.

Representation (1.4) also offers an intuitive explanation of the following well-known properties, or "peculiarities", of the covariances, and hence the correlations, in a CAR(1) model:

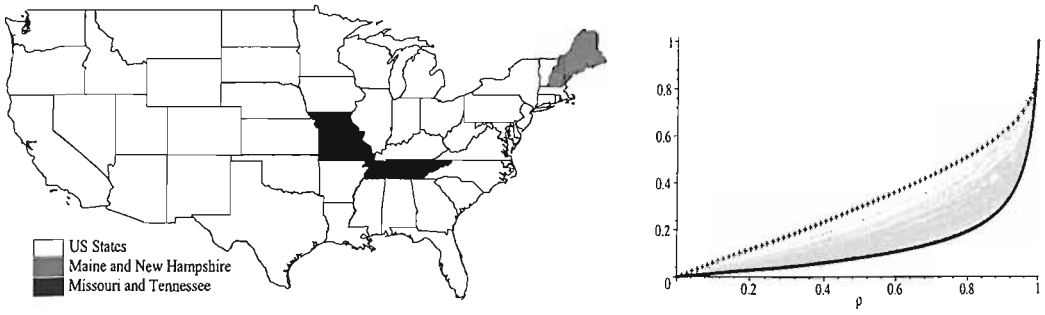
(i) the covariances between variables observed at first-order neighbors are generally different functions of ρ (e.g., Besag and Kooperberg, 1995);

(ii) the ranking of first-order neighbors in terms of their degree of correlation may vary across the parameter space (e.g., Wall, 2004, p. 320-1);

(iii) the behavior of the covariances in a CAR(1) model can be irregular for negative ρ (e.g., Wall, 2004, p. 321).

We next discuss in turn each of the three properties, with the aim of showing how the adoption of a graph theoretic perspective improves the understanding of the correlation structure of spatial autoregressions.

As for property (i), observe that, according to representation (1.4), the covariances between variables observed at first-order neighbors are constant if and only if the total weight of the r-walks connecting a pair of neighbors is the same for all the pairs of neighbors, for each $r = 1, 2, ...$ Clearly, such a condition is not met on general graphs, firstly because different



(a) (b)

Figure 1.1: The correlations, as a function of ρ, implied by Model 1 between the 107 pairs of contiguous US states, with emphasis on Missouri and Tennessee (thick line) and Maine and New Hampshire (crosses).

pairs of first-order neighbors are generally linked by a different number of walks of a certain length (unless strong regularities are imposed on the graph; cf. Section 1.4), and secondly because the weights associated to r-walks connecting different pairs of first-order neighbors are generally different (except for simple weight functions as that implied by Model 2). For illustration, let us consider the cases of Models 1 and 2. Given the weights of walks implied by such models (already discussed above), we have that in Model 1 $\gamma_{i,j}(\rho)$ is large if the short walks between i and j pass through sites with a small number of neighbors, while in Model 2 $\gamma_{i,j}(\rho)$ is large if there is a large number of short walks between i and j. Therefore, Model 1 tends to give large covariance (and correlation) to pairs of neighbors close to the "edges" of the underlying graph (because vertices close to the edges have generally less neighbors than "internal" vertices, at least on common planar graphs), and, on the contrary, Model 2 tends to give large covariance (correlation) to "internal" pairs of neighbors. An example may help to emphasize this fundamental difference between the two models. For ease of comparison, we use the same example as in Wall (2004).

Example 1.3.2 Consider the graph having as vertices the 48 continental US states, and let two states be connected by an edge if and only if they are contiguous. Then, Model 1 implies that, for almost all values of ρ, Maine and New Hampshire are the most correlated contiguous states and that Missouri and Tennessee are the least correlated contiguous states. This is shown in Figure 1.1. Model 2 implies precisely the opposite situation, i.e., that Missouri and Tennessee are the most correlated and Maine and New Hampshire are the least correlated.

Property (ii) says that, regarded as functions of ρ, two covariances between two fixed

pairs of first-order neighbors, say (i,j) and (l,m), may intersect. Representation (1.4) indicates that such a behavior is due to the fact that on general graphs $L(j,j)W^r(i,j) - L(m,m)W^r(l,m)$ need not have the same sign for each r. In the case of Model 2, 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. Note that this characteristic of the covariance function of CAR(1) models, as the previous one, holds not only for first-order neighbors but also for higher-order neighbors.

The most striking aspect of property (iii) is that $\gamma_{i,j}(\rho)$ need not have the same sign over the interval $\lambda_1^{-1} < \rho < 0$. This is due to the fact that in general two vertices are connected by both walks of odd length and walks of even length. It is easily seen from (1.4) that if $d(i,j)$ is odd (resp. even), then $\gamma_{i,j}(\rho)$ is negative (resp. positive) as ρ approaches 0 from the left, but may change sign for more negative values of ρ, as $\sum_{r=2,r\ even}^{\infty} \rho^r W^r(i,j)$ may become greater (resp. smaller) than $\sum_{r=1,r\ odd}^{\infty} \rho^r W^r(i,j)$.

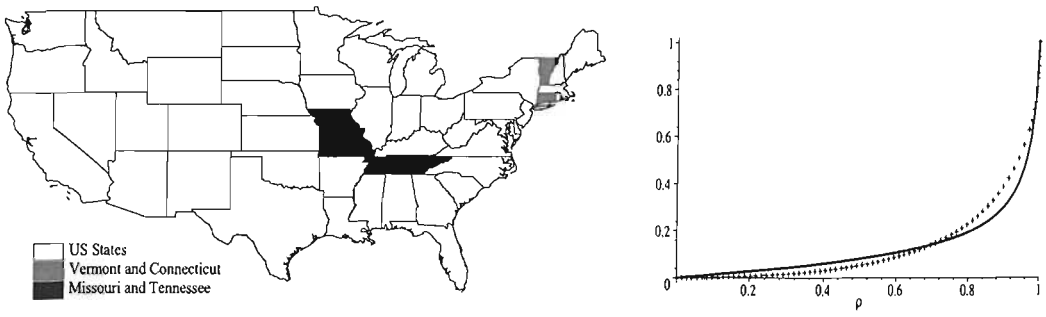
We now turn to analyze the behavior of $\gamma_{i,j}(\rho)$ as the distance $d(i,j)$ changes. Since the models are generally heteroskedastic (see Proposition 1.4.4 below), it is more appropriate to consider the correlations, $\gamma^*_{i,j}(\rho) = \gamma_{i,j}(\rho)/(\gamma_{i,i}(\rho)\gamma_{j,j}(\rho))^{1/2}$. In the limit for ρ approaching 0, we have:

Proposition 1.3.3 *As $\rho \to 0$, $\gamma^*_{i,j}(\rho)$ decreases in absolute value with $d(i,j)$, for any $i,j = 1,...,n$.*

However, contrary to what happens for autoregressions in the time domain, the result in Proposition 1.3.3 need not hold for non-vanishing ρ. To be more precise, we define:

Definition 1.3.4 *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-stable if, for any four vertices i,j,l,m such that $d(i,j) < d(l,m)$ and over the whole parameter space of the model, the absolute value of the correlation between y_i and y_j is non-smaller than the correlation between y_l and y_m. Otherwise, the model is said to be G-unstable.*

The reason why autoregressive models on graphs are not guaranteed to be G-stable can, again, be deduced from representation (1.4); on general graphs, the total weight of the r-walks between two vertices i and j need not be decreasing in $d(i,j)$, so that for sufficiently large ρ the covariance (and the correlation) between two vertices may be larger than the covariance (correlation) between two vertices that are closer together. This is particularly transparent



(a) (b)

Figure 1.2: The correlations, as a function of ρ, implied by Model 1 for a pair of first-order neighbors (thick line; Missouri and Tennessee) and for a pair of second-order neighbors (crosses; Vermont and Connecticut).

in the case on Model 2, where the total weight of the r-walks between two vertices is just the number of r-walks connecting the two vertices.

Generally, for a given neighboring structure, Model 1 has better G-stability properties than Model 2. To see why this is the case, consider four vertices i, j, l, m such that $d(i,j) < d(l,m)$. In view of representation (1.4), a necessary condition for a CAR model to be G-unstable is that there is at least an $r > d(l,m)$ and an $s \leq r$ such that $L(m,m)W^r(l,m) > L(j,j)W^s(i,j)$. This is less likely to occur in Model 1 than in Model 2, because the weights of r-walks decay quicker with r in Model 1, due to the fact that the weight of a walk in Model 1 is inversely related to the product of the degrees of the vertices visited. However, even Model 1 can be G-unstable, as shown, in the context of Example 1.3.2, by Figure 1.2.

Although it has not been acknowledged in previous literature, the potential lack of G-stability is an intrinsic property of spatial autoregressive models, especially when the models are defined on irregular lattices, and represents an important difference between time series and spatial autoregressive models. Unfortunately, it seems difficult to give easy-to-check conditions for the models on general graphs to be G-stable. From the argument in the previous paragraph, however, it emerges that if it is desired to construct stable autoregressions, then one should chose the entries of W in such a way that the implied weights of the r-walks decay quickly with r.

The lack of G-stability or any of the other peculiarities of the covariance function of CAR(1) models discussed above should not necessarily be regarded as drawbacks of the 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 as the restriction to a finite lattice of a process defined on an infinite lattice. In such a case it is certainly 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 the excellent discussion in Besag and Kooperberg, 1995, Section 2.3) and involve the loss of the original Markov property (e.g., Champagnat, Idier and Goussard, 1998). Such "edge effects" definitely represent drawbacks of spatial autoregressive models. Conversely, when the models are defined on *irregular lattices*, it is hard to see any reason why the correlations between units which are not neighbors should necessarily be smaller than the correlation between two neighbors. For instance, if two non-contiguous states share more (first-order or higher-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 connecting the two corresponding observational units (note that the number of walks of length two connecting two units is equal to the number of neighbors shared by the two units).

We continue our analysis with three propositions, and relative discussions, containing some further second-order properties which should be of interest to users of CAR(1) models. We have said above that the behavior of $\gamma^*_{i,j}(\rho)$ can be irregular for negative ρ. We now specify in which sense the behavior is regular for positive ρ.

Proposition 1.3.5 *For positive ρ and any pair of distinct units i and j, $\gamma^*_{i,j}(\rho)$ is positive and increasing in ρ.*

In particular, it is easily checked that the first derivative of $\gamma^*_{i,j}(\rho)$ evaluated at $\rho = 0$ is equal to $[W(i,j)W(j,i)]^{1/2}$ if i and j are neighbors, 0 otherwise. Proposition 1.3.5 asserts that a positive ρ implies that all pairs of random variables are positively autocorrelated. Conversely when $\rho < 0$ the correlations may be positive or negative, but, by our comment to property (iii) above, are certainly not all positive. The next proposition is concerned with the behavior of the correlations at the right extreme of the parameter space of a CAR(1) model.

Proposition 1.3.6 *When ρ approaches λ_n^{-1} from the left, $\gamma^*_{i,j}(\rho) \to 1$, for each $i, j = 1, ..., n$.*

Interestingly, the behavior of the correlations close to the other extreme of the parameter

space is different. Indeed, it can be deduced from the proof of Proposition 1.3.6 that the limit of $\gamma^*_{i,j}(\rho)$ for ρ approaching λ_1^{-1} from the right does not need to be $+1$ or -1.[2] In the particular case of Model 1, Proposition 1.3.6 is a simple application of the celebrated Matrix-Tree-Theorem for the *Laplacian matrix* $D - A$ of an unweighted graph G; e.g., Cvetković *et al.* (1980), p. 38.[3] It should also be remarked that although Proposition 1.3.6 holds for any graph, in some circumstances—and especially when the graph is very irregular and Model 2 is used—$\gamma^*_{i,j}(\rho)$ can approach 1 almost vertically. For instance, for the Model 2 of Example 1.3.2, the correlation between Maine and New Hampshire is less than 0.2 for any $\rho < \rho^*$ with ρ^* extremely close to λ_n^{-1}, and goes to one very rapidly for $\rho^* < \rho < \lambda_n^{-1}$.

The different limits of $\gamma^*_{i,j}(\rho)$ at the right and left extreme of the parameter space is only one manifestation of the already discussed difference in behavior for negative and positive values of ρ. This is in contrast with time series autoregressions, whose correlations functions are either odd or even functions of the autocorrelation parameter. We have:

Proposition 1.3.7 *For any pair of distinct observational units i and j, $\gamma^*_{i,j}(\rho)$ is: (i) an odd function if and only if G is bipartite and $d(i,j)$ is odd; (ii) an even function if and only if G is bipartite and $d(i,j)$ is even.*

Obviously, the variances $\gamma_{i,i}(\rho)$ fall within case (ii) of the proposition. The combination of Propositions 1.3.6 and 1.3.7 implies that, on a bipartite graph G, when ρ approaches λ_1^{-1} from the right, $\gamma^*_{i,j}(\rho)$, $i,j = 1,...,n$, tends to $+1$ if $d(i,j)$ is even and to -1 if $d(i,j)$ is odd.

We conclude our discussion of CAR(1) models with a remark on the interpretation of $\gamma_{i,j}(\rho)$ in terms of walks in a graph. Since the number of linearly independent non-negative powers of W is—as they all share the same eigenspaces—equal to the number s, say, of distinct eigenvalues of W, it follows that the infinite series representation (1.4) could be concentrated to a weighted sum of s of the powers of W, say $\gamma_{i,j}(\rho) = L(j,j)\sum_{r\in\{r_1,...,r_s\}} c_r(\rho)W^r(i,j)$. Thus, $\gamma_{i,j}(\rho)$ could be interpreted in terms of contributions coming only from the walks between i and j of s lengths. The reason why we have not adopted this representation is that in practice the coefficients $c_r(\rho)$ are not known in closed form, unless W is an highly structured matrix.

[2] Because λ_1 is not necessarily simple and, even when it is simple, an eigenvector associated to λ_1 may have i-th and/or j-th components equal to zero.

[3] Incidentally, the 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; see Besag *et al.* (1991) and Besag and Kooperberg (1995).

1.3.3 SAR(1) Model

The main objective of this subsection is to show that the results obtained above for CAR(1) models can be extended in a simple manner to SAR(1) models. Contrary to the case of a CAR(1) model, in a SAR(1) model the eigenvalues of the matrix W are generally complex. By the Perron-Frobenius theorem for nonnegative irreducible matrices, however, W admits a real positive eigenvalue, whose modulus is non-smaller than that of the remaining eigenvalues. With a slight abuse of notation, such an eigenvalue is denoted by λ_n.

A SAR(1) model is the family of n-variate centered Gaussian distributions with covariance matrix $\Sigma_{s,1} = \sigma^2(I - \rho W)^{-1}V(I - \rho W')^{-1}$, which is p.d. as long as ρ is different from the reciprocal of the non-zero real eigenvalues of W. We denote the covariance and the correlations in a SAR(1) model respectively by $\phi_{i,j}(\rho)$ and $\phi^*_{i,j}(\rho)$, for $i, j = 1, ..., n$. Recall that, contrary to what happens for the matrices W and L in a CAR(1) model, in a SAR(1) model there is no joint restriction between the matrices W and V (that is, W and V can be chosen independently). In most applications $V = I$, which has the virtue of making the (null) hypothesis of i.i.d. data testable in the SAR model.

Given a graph with first distance matrix A (see Section 1.3.1), the most popular specifications of the weights matrix W in applied work are as in Models 1 and 2 above, that is, $W = D^{-1}A$ or $W = A$. We now briefly discuss the interpretation of the covariance structure implied by these two specifications, when $V = I$ and, without any loss of generality, $\sigma^2 = 1$.

When $W = D^{-1}A$ and $|\rho| < 1$, $\Sigma_{s,1}$ admits the power series representation

$$\sum_{r_1,r_2=0}^{\infty} \{\rho^{r_1+r_2} \left(D^{-1}A\right)^{r_1} \left(AD^{-1}\right)^{r_2}\}.$$

It is easily seen that such an expression implies that any walk between i and j (in the graph with adjacency matrix W) contributes to the covariance $\phi_{i,j}(\rho)$, and that the contribution of an r-walk $(i_0, i_1, ..., i_r)$ to $\phi_{i_0,i_r}(\rho)$ is $\rho^r \sum_{l=0}^{r} d_{i_l} / \prod_{l=0}^{r} d_{i_l}$.[4] On the other hand, when $W = A$, $\Sigma_{s,1}$ is equal to $(I - \rho A)^{-2}$ (the square of the covariance matrix of Model 2), which, when $|\rho| < \lambda_n^{-1}$, can be represented as $\sum_{r=0}^{\infty} (r+1)\rho^r A^r$. In view of the graph theoretic interpretation of the entries of A^r, we have a straightforward comparison with Model 2, whose covariance matrix can be represented as $\sum_{r=0}^{\infty} \rho^r A^r$; the covariance structure of the two models on the same graph G is determined by weighting the same number of walks with

[4]Note that if $V = D^{-2}$, then $\Sigma_{s,1} = (D - \rho A)^{-2}$, the square of the covariance matrix of Model 1. In such a case, when $|\rho| < 1$, the contribution of an r-walk is $\rho^r s_r / \prod_{l=0}^{r} d_{i_l}$, where s_r denotes the r-th elementary symmetric polynomial in the $r+1$ variables $d_{i_0}, d_{i_1}, ..., d_{i_r}$.

different (but similar) functions of ρ, namely ρ^r in the CAR(1) model and $(r+1)\rho^r$ in the SAR(1) model. For $\rho < e^{-1}$ the function $(r+1)\rho^r$, $r \geq 0$, is decreasing with r, while for larger values of ρ it first increases and then decreases.

The arguments in the previous paragraph show that the second-order structure of the two most popular versions of the SAR(1) model is determined in a very similar way as the second-order structure of CAR(1) models. In particular, this means that all the peculiarities of the correlation structure of CAR(1) models listed above will also characterize the two popular SAR(1) models. For the sake of brevity, details of such a characterization and extensions to more complicated choices of the matrices W and V are left to the reader.

It is perhaps worth remarking that, despite the similarities emphasized above, SAR(1) models are very different from CAR(1) models in many other respects. For instance, except for very particular cases, (i) SAR(1) models on a graph G do not satisfy the classical definition of the Markov property with respect to G (see, for instance, Speed and Kiiveri, 1986); (ii) they are curved exponential families (see Lemma 1.2.3 above); (iii) they do not admit a simple statistical interpretation for their parameter ρ.

Although it is difficult to derive a statistical interpretation of the parameter ρ of a SAR(1) model, the following proposition shows that, in the region $0 < \rho < \lambda_n^{-1}$, the larger ρ is, the larger all the correlations $\phi^*_{i,j}(\rho)$ are, for any W and V.

Proposition 1.3.8 *For $0 < \rho < \lambda_n^{-1}$ and any pair of distinct units i and j, $\phi^*_{i,j}(\rho)$ is positive and increasing in ρ.*

Therefore, in the region $0 < \rho < \lambda_n^{-1}$, the parameter ρ can be thought of as a measure of the spatial correlation in y. When $\rho > \lambda_n^{-1}$, the behavior of the correlations as a function of ρ is not as in Proposition 1.3.8 and the just-mentioned interpretation of ρ does not hold (in fact, the restriction $\rho < \lambda_n^{-1}$ is usually imposed on the parameter space of SAR models). The next lemma compares the correlations of CAR(1) and SAR(1) models when the two models are constructed on the basis of the same weights matrix, when $V = L = I$, and over the whole region $0 < \rho < \lambda_n^{-1}$.

Lemma 1.3.9 *For $V = L = I$, a fixed W and $0 < \rho < \lambda_n^{-1}$, the correlation between y_i and y_j in a SAR(1) model is always greater than the corresponding correlation in a CAR(1) model.*

Finally, the two following results represent the counterparts of Propositions 1.3.6 and 1.3.7. They hold for any SAR(1) model.[5]

[5]Kelejian and Robinson (1995) contains a numerical investigation of properties of SAR models connected

Proposition 1.3.10 *When ρ approaches λ_n^{-1}, $\phi_{i,j}^*(\rho) \to 1$, for each $i, j = 1, ..., n$.*

Proposition 1.3.11 *For any pair of distinct observational units i and j, $\phi_{i,j}^*(\rho)$ is: (i) an odd function of ρ if and only if G is bipartite and $d(i,j)$ is odd; (ii) an even function of ρ if and only if G is bipartite and $d(i,j)$ is even.*

1.3.4 Multi-Parameter Models

In this section, we discuss how our analysis of CAR and SAR models can be extended from the one-parameter case ($p = 1$) to the multi-parameter case ($p > 1$). The use of multi-parameter models is particularly relevant when the data are observed over irregular spatial configurations. In that case, it is more unlikely that a family of distributions indexed by one parameter (or two, including τ^2 or σ^2) is sufficiently rich to provide a useful representation of the associations amongst the random variables in y. This is essentially due to the fact that the more irregular a spatial configuration is, the more uncertain the specification of a weights matrix is likely to be. In addition, the adoption of multi-parameter models can be regarded as an attempt to overcome the restrictive characteristics of the first-order models discussed above.

Given a graph G, there are two fundamentally different ways of increasing the dimension of the parameter space of CAR(1) and SAR(1) models, the distinction being based on whether or not the resulting multi-parameter models maintain the same conditional independence structure as the original one-parameter model.[6] The first approach, which does maintain the original conditional independence structure, consists of splitting the edge set E of the graph G in p disjoint subsets $E_1, ..., E_p$, and associating a parameter to each subset, so that C (or S) is parametrized as $\sum_{l=1}^{p} \rho_l W_l$ with $W_l(i,j)$ equal to $W(i,j)$ if $(i,j) \in E_l$, to 0 otherwise.[7] For instance, on a rectangular 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 classifying edges will vary from application to application. We are not aware of work addressing this issue, but one criterion that might prove to be useful to classify edges is their closeness to the borders of the graph, as could be measured, for

to the results in Propositions 1.3.10 and 1.3.11.

[6] For details on the conditional independence structure of a Gaussian model, see, for instance, Speed and Kiiveri (1986). In the case of a CAR model, two distinct variables y_i and y_j are conditionally independent given all the remaining random variables in y, if and only if $W_l(i,j) = 0$, for each $l = 1, ..., p$.

[7] Note that for CAR models, such a construction is hampered by the restriction that each $L^{-1}W_l$ must be symmetric.

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 the one-parameter models, additional parameters are associated to different degrees of interaction (in a sense to be specified below). In this case the models are usually called p-th order models and denoted by CAR(p) and SAR(p). Contrary to the time series case, in a spatial setting there is no obvious way to perform such an extension. In the following, we show how the use of a graph theoretic perspective clarifies the differences among three alternatives that have been proposed in the literature. For the sake of simplicity, we limit ourselves to the case $W = A$, that is, to the case when the underlying graph is unweighted.

Given a graph G, a first possibility is to take $W_l = A^l$, $l = 1, ..., p$ (e.g., Huang, 1984). The matrices A^l have some non-zero elements on their diagonals, which requires some modifications in our definition of CAR models (see Martin, 1990). The fact that the entries of the powers A^l count the numbers of l-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, and Ross and Harary, 1952, for an earlier discussion of the same issue in a non-spatial context). As a consequence, a second possible way of defining higher-order models is to set $W_l = P_l$, $l = 1, ..., p$, P_l being the matrices with (i,j)-th entry equal to the number of paths of length l between i and j. Evidently, $P_l \neq A^l$, for each $l > 1$, because of the symmetry of A (if A were a triangular matrix, for instance representing temporal unilateral interaction, we would have $P_l = A^l$, for each l). A third possibility, which in a sense is the most natural in the graph metric induced by the graph distance $d(\cdot,\cdot)$, is to take $W_l = A_l$ (see Anselin and Smirnov, 1996). Of course, the distance matrix A_l is just the Boolean version of P_l.

An important observation here is that, given a graph G, the three ways of constructing higher-order models imply the same conditional independence structure, both in the case of CAR and in the case of SAR models. More precisely, two variables y_i and y_j in a CAR(p) model are conditionally independent—given all the remaining variables—if and only if $d(i,j) > p$, whereas they are conditionally independent in a SAR(p) model if and only if $d(i,j) > 2p$.

Regardless of how the higher models are constructed, their main difference (in terms of correlation properties) from the first-order models is that every walk between two vertices is weighted differently according to its possible compositions in terms of steps of length

$1, ..., p$. As a simple example, let us consider a CAR(2) model on a path formed by three vertices. If $W_2 = A_2$, then the covariance between the two extremes of the path is equal to $\rho_1^2 + \rho_2 + 2\rho_2\rho_1^2 + \rho_2^3 + ...$ (the third term, for instance, indicates that there are two ways of "walking" from one extreme to the other by composing two 1-walks with one 2-walk), whereas if $W_2 = A^2$, then the covariance is $\rho_1^2 + \rho_2 + 6\rho_2\rho_1^2 + 2\rho_2^2 + 4\rho_2^3 + ...$ (here the third term indicates that there are six ways of going from one extreme to the other with the same composition of walks as above).

In the rest of the paper we limit ourselves to the third way of constructing higher-order spatial autoregressive models and, unless otherwise specified, the names CAR(p) and SAR(p) on a graph G will be reserved for the models where the first p distance matrices of G are used as weights matrices.

The CAR(p) and SAR(p) models on a graph G may be regarded as particular cases (obtained for $q = 1$ and $q = 2$, respectively) of the large class of models $N_n(0, \Omega_{p,q})$, with p.d. covariance matrix

$$\Omega_{p,q} = \sigma^2 \left(I - \sum_{l=1}^{p} \rho_{q,l} A_l \right)^{-q},$$

where p and q are positive integers, with $p \leq d$. When $\rho_1, ..., \rho_p$ are such that the eigenvalues of $\sum_{l=1}^{p} \rho_l A_l$ are all smaller than one in absolute value, $\Omega_{p,q}$ admits the representation[8]

$$\Omega_{p,q} = \sigma^2 \sum_{r=0}^{\infty} \binom{r+q-1}{r} \left(\sum_{l=1}^{p} \rho_l A_l \right)^r . \tag{1.5}$$

It is then clear that the correlation properties derived above for the CAR(1) and the SAR(1) models on a graph can be generalized to the models $N_n(0, \Omega_{1,q})$, $q > 2$.

The analysis of the models $N_n(0, \Omega_{p,q})$ with $p > 1$ is more complicated, but, as we are about to see, some interesting properties can be derived for particular classes of graphs.

1.4 Symmetries and Regularities

In this section we study properties of CAR(p) and SAR(p) models on graphs satisfying some symmetries (in a group theoretic sense) or some regularities (in a combinatorial sense). The results obtained here are less directly useful to practitioners than those in the previous section, but they help to clarify the connection between the covariance properties of spatial autoregressions and the properties of the underlying graphs.

[8] If further flexibility were required, the models could also be extended to have covariance matrix equal to a generalized hypergeometric function with matricial argument.

An *automorphism* of a graph G (e.g., Biggs, 1993) is a permutation g on the vertex set $V(G)$ 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. 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, any covariance matrix $\Omega_{p,q}$ (defined at the end of the previous section) is invariant under $Aut(G)$, in the sense that $P_g \Omega_{p,q} P_g' = \Omega_{p,q}$, for each $g \in Aut(G)$.[9]

For notational convenience, let $\varphi_{i,j}$ denote $\Omega_{p,q}(i,j)$, for any positive integers $p \leq d$ and q. An immediate consequence of the invariance just described is that $\varphi_{i,j} = \varphi_{l,m}$ if there exists $g \in Aut(G)$ such that $g(i) = l$ and $g(j) = m$, for $i, j, l, m = 1, ..., n$ (this property, of course, holds also for the correlations). Let us look at an example.

Example 1.4.1 In Figure 1.3 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. For this graph, $\varphi_{i,2} = \varphi_{j,5}$ when $(i,j) = (2,5), (1,1), (3,3), (4,4), (6,6)$.

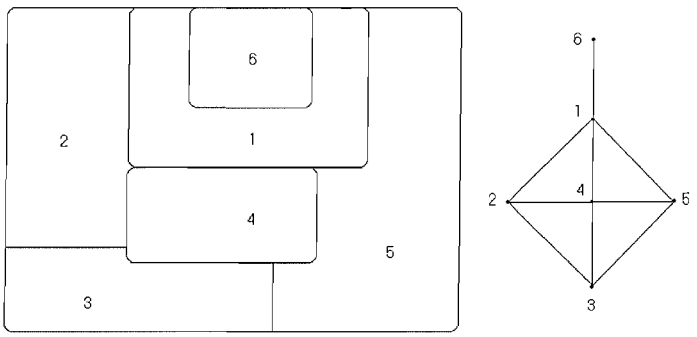


Figure 1.3: A planar configuration of observational units and a corresponding graph.

It is clear that the covariance properties of models defined on graphs having large automorphism group will be particularly simple. The following definition can be found, for instance, in Biggs (1993).

Definition 1.4.2 *A graph G is said to be: vertex-transitive if $Aut(G)$ acts transitively*

[9] In group theoretic terminology, this is equivalent to saying that any $\Omega_{p,q}$ 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 $V(G) \times V(G)$; James (1957) can be consulted for an application of these notions in statistics.

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 $\varphi_{i,i}$ do not depend on i if G, whereas if G is edge-transitive, then that the covariances $\varphi_{i,j}$ between first-order 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 1.4.

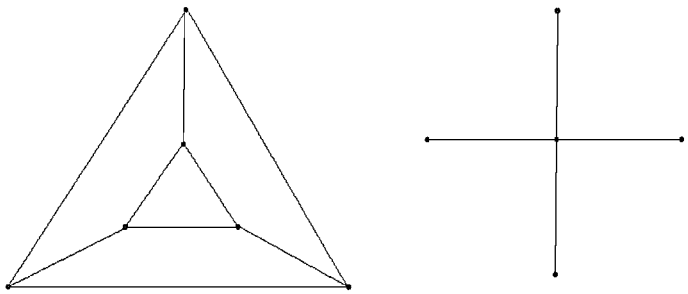


Figure 1.4: A vertex-transitive graph (left) and an edge-transitive graph (right).

It is of interest, at this point, to characterize a notion of stationarity for our models on graphs. Given the metric induced by the graph distance $d(\cdot,\cdot)$, it is natural to consider the following definition (we do not distinguish between weak and strong stationarity, as this paper is concerned only with Gaussian models).

Definition 1.4.3 *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 $d(i,j)$.*

A sufficient condition for the models $N_n(0, \Omega_{p,q})$ to be G-stationary is, evidently, distance-transitivity of G. However the condition is *not* necessary, in the same way as vertex-transitivity of G is *not* necessary for the models to be homoskedastic. It turns out that necessary and sufficient conditions for homoskedasticity and G-stationarity of the models $N_n(0, \Omega_{p,q})$ are given in terms of combinatorial regularities, rather than group theoretic symmetries, of the underlying graphs. 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 from a vertex i does not depend on i, for each $r \geq 0$ (Godsil, 1993, Sec. 5.3). For the first-order models, we have:

Proposition 1.4.4 *Any first-order model $N_n(0, \Omega_{1,q})$ on a graph G is homoskedastic if and only if G is walk-regular.*

The property does not generally extend to higher-order models (with the distance matrices as weights matrices). This is easily seen from (1.5), because walk-regularity does not necessarily imply that the powers of A_l, $l \geq 2$, have constant diagonal. Conversely, if the powers of A were used as weights matrices, then the higher-order models on a walk-regular graph would always be homoskedastic.

Consider now the class of *distance-regular* graphs. A graph is said to be distance-regular if $\{A_0, A_1, ..., A_d\}$ is an *association scheme*, or, equivalently, if $Span\{A_0, A_1, ..., A_d\}$ is a *Bose-Mesner algebra*; e.g., Bannai and Ito (1984). Another known basis of the Bose-Mesner algebra spanned by the distance matrices of a distance-regular graph is $\{A^0, A^1, ..., A^d\}$. 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 (and their generalization) is emphasized by the next result.

Theorem 1.4.5 *Any model $N_n(0, \Omega_{p,q})$ on a graph G is G-stationary if and only if G is distance-regular.*

This result, together with Proposition 1.4.4, is in agreement with the well-known fact that finite autoregressive models are homoskedastic or stationary only in very special circumstances; see, for instance, Besag and Kooperberg (1995), p. 735. A weaker condition than G-stationarity, which is of interest to the user of the models has been discussed in Section 1.3.2 for the CAR(1) model, under the name of G-stability.

In addition, the notion of distance-regular graph enables us to elucidate the relationship between the curvature of a spatial autoregression on a graph G and the regularities of G, in the following sense.

Theorem 1.4.6 *Any model $N_n(0, \Omega_{d,q})$, $q \geq 2$, on a graph G is a regular exponential family if G is distance-regular with diameter d, a curved exponential family otherwise*

1.5 Summary of Results and Extensions

In this final section, we summarize the results of the paper and discuss some issues that have not been considered in the paper.

Our two main contributions to the literature on spatial autoregressive models are as follows. Firstly, we have provided an interpretation of the covariance structure of spatial autoregressions. The interpretation follows from the fact that, when the neighborhood structure of the set of the observational units is represented by a graph, the covariance between two variables y_i and y_j in a CAR(1) model can be viewed as the *generating function* of the total weight of the walks of the same length between the two vertices i and j (see eq. (1.4)). We have argued—by examining some examples and by deriving more formal results—that such an interpretation provides a valid tool for understanding the correlation structure of CAR(1) models. Extensions to SAR(1) models are straightforward, whereas the treatment of multi-parameter CAR and SAR models is more involved, but may be based on that of the first-order models. The graph theoretic interpretation also sheds 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 when the models are defined on irregular lattices, due to the fact that the information used to specify such matrices is typically uncertain. Secondly, we have studied the properties enjoyed by the models when the underlying graphs satisfy symmetries or regularities. This has enabled us to give necessary and sufficient conditions for G-stationarity and homoskedasticity of the models. Rather than by imposing restrictions on the graph, homoskedasticity or G-stationarity could, of course, be achieved by changing the marginal distributions of some of the variables y_i (for instance by suitably choosing the matrix V in a SAR model). We have not pursued this approach in the paper, as it seems to be impractical for arbitrary configurations of the observational units.

Next, we discuss two issues that are relevant to cast the spatial models studied in present paper—which are Gaussian models whose inverse covariance matrix has a linear structure—into a more general context. The first issue regards the choice of Gaussianity as the joint density of a spatial process, the second the choice between parametrizing a covariance matrix or its inverse.

- It is natural to ask how the choice of Gaussianity affects the interaction structure of a spatial process defined on a graph. The question is best discussed in the context of

Markov random fields, of which our CAR models are an example (see Besag, 1974, or Section 6.4 of Cressie, 1993). By the Hammersley-Clifford theorem, Gaussianity implies that a Markov random field is a *pairwise interaction processes* (see the same references as above for details); roughly speaking, a random field is a pairwise interaction process if its log-density does not depend on the data through interaction terms $y_{i_1} \cdots y_{i_m}$ of order m higher than two. 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. Note that, in particular, bipartite graphs do not contain cliques of size greater than two, because they do not admit cycles of odd length (see Section 1.3.1). Hence, Gaussianity does not restrict (in the above sense) the interaction structure of Markov random fields defined on a graphs without cliques of size greater than two, such as bipartite graphs. On the other hand, it may well be possible that non-Gaussian models provide a better representation of data observed over graphs with cliques of size greater than two; see Besag and Tjelmeland (1998) and Lee *et al.* (2001). This would be an important aspect to consider in future research, since the graphs of irregular lattices do generally contain cliques of size greater than two.

- There have been debates in the spatial statistics literature about whether *autoregressive models* or *covariance models*—i.e., models parametrizing directly the covariance matrix rather than its inverse—are more suitable to analyze the spatial structure of a set of data. Our simple view on the subject is that, from a pure modelling point of view, the choice between the two classes of models should mainly be based on the fact that the former are useful when it is desired—perhaps because of the availability of some a priori information—to restrict the conditional independence structure of a set of random variables, while the latter should be preferred when it is desired to control directly the covariance function.[10] The present paper can be regarded as an investigation of the behavior of a covariance matrix when its inverse has a linear structure.

Finally, we believe that the graph theoretic treatment given in the paper would also be helpful to study the properties of inferential procedures in the context of spatial autoregressions, but this certainly goes beyond the scope of the present paper.

[10] Of course, in practice, the choice between the two specifications may be based on other factors, such as the computational convenience of inferential procedures.

Appendix. Proofs

Proof of Lemma 1.2.1 Setting $\rho_0 = -1$ and $W_0 = I$, it is readily verified that the a CAR model is a family of densities (1.3) with $\eta_l = \rho_l/(2\tau^2)$ and $s_l(y) = y'L^{-1}W_l y$, for $l = 0, 1, ..., p$. Since the matrices W 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_l such that $\Sigma_{c,p} = -(\sum_{l=0}^{p} \eta_l W_l)^{-1} L$ 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 a 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 the Ω is open, which completes the proof of the proposition.

Proof of Lemma 1.2.2 By the discussion above the statement of the lemma, a SAR model is a regular exponential family if and only if each matrix B_{l_1,l_2}, for $l_1 = 0, 1, ..., p$ and $l_2 = 1, ..., l_1$, is a linear combination of the matrices $B_{l,0}$, for $l = 0, 1, ..., p$. Such a condition implies that if $B_{l_1,l_2}(i,j) \neq 0$ for some pair of units i, j and some pair of indexes l_1, l_2, then $B_{l,0}(i,j) \neq 0$ for at least one $l = 0, 1, ..., p$. Note that $B_{l_1,l_2}(i,j) \neq 0$ if and only if there exists a unit k such that $W_{l_1}(i,k) \neq 0$ and $W_{l_2}(j,k) \neq 0$, and that $B_{l,0}(i,j) \neq 0$ implies $W_l(i,j) > 0$. It follows that a necessary condition for a SAR model to be a regular exponential family is: if for any pair of units i, j and any pair of indexes l_1, l_2 there exists a k such that $W_{l_1}(i,k) \neq 0$ and $W_{l_2}(j,k) \neq 0$, then $W_l(i,j) > 0$ for at least one $l = 0, ..., p$. But, in view of the assumption of irreducibility of $\sum_{l=0}^{p} W_l$, such a condition is equivalent to requiring that for any pair of units i, j, there is an $l = 0, ..., p$ such that $W_l(i,j) > 0$. The proof of the lemma is therefore completed.

Proof of Lemma 1.2.3 If $I - \rho W$ is not full, then the SAR(1) model with spatial weights matrix W is a curved exponential family by Lemma 1.2.2. Thus, the dimension of $\Delta_1 = span\{V^{-1}, V^{-1}W + W'V^{-1}, W'V^{-1}W\}$, i.e., the dimension of a a minimal sufficient statistic of a SAR(1) model, must be less than 3. That it is 2 is an obvious consequence of the fact that V^{-1} and $V^{-1}W + W'V^{-1}$ are linearly independent.

Proof of Lemma 1.2.5 The matrix $\Sigma_{s,p}^{-1}$ is the square of a matrix belonging to the $(p+1)$-dimensional subspace Ψ_p. Therefore, if Ψ_p is a quadratic subspace, $\Sigma_{s,p}^{-1} \in \Psi_p$, i.e. the SAR($p$) model with covariance matrix $\Sigma_{s,p}$ is a regular exponential family. This proves

the sufficiency of the condition in the lemma. Conversely, if a SAR(p) model is a regular exponential family, $\Sigma_{s,p}^{-1}$ belongs to a $(p+1)$-dimensional subspace of $\mathcal{L}(n)$, say Θ_p. It then follows immediately that the matrices $W_0, ..., W_p$ belong to Θ_p and that, since they are linearly independent by assumption, they span Θ_p. Hence $\Theta_p = \Psi_p$, which implies that Ψ_p is a quadratic subspace, Definition 1.2.4 being satisfied with $Q = \Sigma_{s,p}^{-1/2}$.

Proof of Proposition 1.3.3 Exploiting representation (1.4), which always holds in a neighborhood of $\rho = 0$, we obtain that

$$\gamma_{i,j}^*(\rho) = \frac{L(j,j)\sum_{r=d(i,j)}^{\infty}(\rho^r W^r(i,j))}{[L(i,i)L(j,j)\sum_{r=0}^{\infty}(\rho^r W^r(i,i))\sum_{r=0}^{\infty}(\rho^r W^r(j,j))]^{\frac{1}{2}}}.$$

It is then clear that the larger $d(i,j)$ is, the slower $\gamma_{i,j}^*(\rho)$ goes to zero as $\rho \to 0$.

Proof of Proposition 1.3.5 For $0 < \rho < \lambda_n^{-1}$ and any pair of units i and j, (1.4) implies that $\gamma_{i,j}(\rho)$, and hence $\gamma_{i,j}^*(\rho)$ is positive. Next, let a superposed dot denote differentiation with respect to ρ. Direct differentiation of $\gamma_{i,j}^*(\rho) = \gamma_{i,j}(\rho)/[\gamma_{i,i}(\rho)\gamma_{j,j}(\rho)]^{1/2}$ shows that $\dot{\gamma}_{i,j}^*(\rho)$ is positive in the region $0 < \rho < \lambda_n^{-1}$ if and only if $2\dot{\gamma}_{i,j}(\rho)/\gamma_{i,j}(\rho) > \dot{\gamma}_{i,i}(\rho)/\gamma_{i,i}(\rho) + \dot{\gamma}_{j,j}(\rho)/\gamma_{j,j}(\rho)$. We demonstrate that such an inequality holds by showing that $\dot{\gamma}_{i,j}(\rho)\gamma_{i,i}(\rho) > \dot{\gamma}_{i,i}(\rho)\gamma_{i,j}(\rho)$ for any $i \neq j$ and $0 < \rho < \lambda_n^{-1}$. Exploiting the expansion $L(j,j)\sum_{r=d(i,j)}^{\infty}\rho^r W^r(i,j)$ of $\gamma_{i,j}(\rho)$, the last inequality may be rewritten as

$$\sum_{r=1}^{\infty}\rho^{r-1}\left\{rW^r(i,j) + \sum_{k=0}^{r}\left\{kW^k(i,j)W^{r-k}(i,i)\right\}\right\} > \sum_{r=1}^{\infty}\rho^{r-1}\sum_{k=0}^{r}\left\{kW^k(i,i)W^{r-k}(i,j)\right\}, \tag{1.6}$$

which certainly holds if

$$\sum_{k=0}^{r}\left\{kW^k(i,j)W^{r-k}(i,i)\right\} \geq \sum_{k=0}^{r}\left\{kW^k(i,i)W^{r-k}(i,j)\right\}, \tag{1.7}$$

for each positive integer r and with strict inequality for at least one r. After some simple algebraic manipulation and letting $\bar{r}$ denote $\lfloor r/2 \rfloor - 1$, the last condition turns out to be equivalent to requiring that

$$\sum_{k=0}^{\bar{r}}\left\{(r-2k)\,W^k(i,i)W^{r-k}(i,j)\right\} \geq \sum_{k=0}^{\bar{r}}\left\{(r-2k)\,W^{r-k}(i,i)W^k(i,j)\right\}, \tag{1.8}$$

for each positive integer r (strict inequality certainly holding for $r = 1$). We now need some new notation: given z positive integers $h_1 < h_2 < ... < h_z$, we denote by $c_{h_1,...,h_z}$ the quantity $W^{r-h_z}(i,i)\prod_{t=3}^{z}\{W^{h_t-h_{t-1}}(i,i)W^{h_2-h_1}(i,j)\}$. Considering the graph with adjacency matrix W, $c_{h_1,...,h_z}$ can be interpreted as the weight of the all the walks $(i, l_1, ..., l_{r-1}, j)$ with $l_{r-h_1} = ... = l_{r-h_z} = i$ and no other repetition of i (here and in the rest of this proof, by weight of a

walk from i to j we mean the product of the weights of its steps, without including—as in the definition given in Section 1.3.1—the factor $L(j,j)$, which has been conveniently eliminated in (1.6) above). We can then rewrite the right-hand side of (1.8) as

$$\sum_{z=1}^{\bar{r}-1} \sum_{h_1,...,h_z=1}^{\bar{r}} \left\{ \left(zr - 2\sum_{t=1}^{z} h_t \right) c_{h_1,...,h_z} \right\}. \tag{1.9}$$

Observe that if a graph contains a walk $(i, l_1, ..., l_{r-1}, j)$ with $l_{r-h_1} = ... = l_{r-h_z} = i$ and $h_1 < h_2 < ... < h_z$, then it also contains a walk $(i, m_1, ..., m_{r-1}, j)$ with same weight and with $m_{h_z - h_{z-1}} = ... = m_{h_2 - h_1} = i$. It follows that the left-hand side of (1.8) must be at least as large as

$$rW^r(i,j) + \sum_{z=2}^{\bar{r}-1} \sum_{h_1,...,h_z=1}^{\bar{r}} \{[(z-1)r + 2(h_1 - h_z)] c_{h_1,...,h_z}\}, \tag{1.10}$$

where, evidently,

$$W^r(i,j) = \sum_{z=1}^{\bar{r}-1} \sum_{h_1,...,h_z=1}^{\bar{r}} c_{h_1,...,h_z}. \tag{1.11}$$

Let us now consider two cases, according to whether or not $W^r(i,j)$ is less than the right-hand side of (1.8). In the latter case (1.8) holds trivially, because the coefficients $r-2k$ are decreasing in k. Conversely, it is easily seen that if $W^r(i,j)$ is less than $\sum_{k=d(i,j)}^{\bar{r}} \{W^{r-k}(i,i) W^k(i,j)\}$, then there is at least one walk of length r from i to j visiting i at least twice in its second half, that is, at least one of the coefficients $c_{h_1,...,h_z}$, for $z > 1$ and $h_1, ..., h_z = 0, ..., \bar{r}$ is positive. Then, putting (1.9), (1.10) and (1.11) together, it emerges that (1.8) holds for each positive r even when $W^r(i,j)$ is less than the right-hand side of (1.8), which completes the proof of the second and last part of Proposition 1.3.5.

Proof of Proposition 1.3.6 Rewrite the covariance matrix $\Sigma_{c,1} = (I - \rho W)^{-1} L$ as

$$\Sigma_{c,1} = L^{1/2} (I - \rho L^{-1/2} W L^{1/2})^{-1} L^{1/2}.$$

The matrix $L^{-1/2} W L^{1/2}$ appearing in such an expression is symmetric because $L^{-1}W$ is, and has eigenvalues $\lambda_1, ..., \lambda_n$ because it is similar to W. Thus, it admits a spectral decomposition $\sum_{l=1}^{n} \lambda_l q_l q_l'$, where q_l is an eigenvector of $L^{-1/2} W L^{1/2}$ associated to λ_l. It follows at once that

$$\Sigma_{c,1} = L^{1/2} \sum_{l=1}^{n} \left(\frac{1}{1 - \rho\lambda_l} q_l q_l' \right) L^{1/2}.$$

According to the Perron-Frobenius theorem for nonnegative irreducible matrices (applied to $L^{-1/2} W L^{1/2}$), λ_n is a simple eigenvalue and there exists an entrywise positive eigenvector q_n. It follows that the (eigenprojection) matrix $q_n q_n'$ is entrywise positive, from which it is easily seen that $\gamma^*_{i,j}(\rho) \to (q_n)_i (q_n)_j / [(q_n)_i^2 (q_n)_j^2]^{1/2} = 1$ as $\rho \to \lambda_n^{-1}$, for $i, j = 1, ..., n$.

Proof of Proposition 1.3.7 Evidently, any correlation $\gamma^*_{i,j}(\rho)$ is symmetric about the origin or the vertical axis if and only if the covariance $\gamma_{i,j}(\rho)$ satisfies the same symmetry. Thus, it suffices to prove the proposition for any function $\gamma_{i,j}(\rho)$, with $i \neq j$. If G is bipartite, then $\lambda_1 = \lambda_n$ by Lemma 1.3.1, and therefore the power series representation (1.4) is valid over the whole parameter space of the CAR(1) model on G. It is well known that a bipartite graph does not contain any cycles of odd length, which entails that $W^r(i,j) = 0$ if r and $d(i,j)$ have different parity. This is because if $W^r(i,j)$ were different from zero for r and $d(i,j)$ with different parity, there would be a closed walk in G of length $r + d$, which would be odd. Then G would not be bipartite, for a closed walk of odd length always contains a cycle of odd length. As a consequence,

$$\gamma_{i,j}(\rho) = \begin{cases} L(j,j) \sum_{r=1,r\ odd}^{\infty} \rho^r W^r(i,j), & \text{if } d(i,j) \text{ is odd,} \\ L(j,j) \sum_{r=1,r\ even}^{\infty} \rho^r W^r(i,j), & \text{if } d(i,j) \text{ is even.} \end{cases}$$

The sufficiency of the conditions in (i) and (ii) is then a consequence of the symmetries of the function ρ^r. The necessity part follows trivially from the fact that if G is not bipartite, then the function $\gamma_{i,j}(\rho)$ is neither odd nor even, as $\lambda_1^{-1} \neq -\lambda_n^{-1}$ by Lemma 1.3.1.

Proof of Proposition 1.3.8 By using a power series expansion of $(I - \rho W)^{-1}$ (and of its transpose), it is easily seen that, for $0 < \rho < \lambda_n^{-1}$, $\Sigma_{s,1}$ is nonnegative. Hence, $\phi^*_{i,j}(\rho)$ is positive for $0 < \rho < \lambda_n^{-1}$ and any pair of units i and j. That $\phi^*_{i,j}(\rho)$ is increasing in ρ can be proved by obvious modification of the argument used in the proof of Proposition 1.3.5.

Proof of Lemma 1.3.9 When $V = L = I$, $\Sigma_{c,1} = (I - \rho W)^{-1}$ and $\Sigma_{s,1} = \Sigma^2_{c,1}$. Clearly, in the region $0 < \rho < \lambda_n^{-1}$, $\phi_{i,j}(\rho) > \gamma_{i,j}(\rho)$ for any $i \neq j$, because $\phi_{i,j}(\rho) = \sum_{r=0}^{\infty} \rho^r W^r(i,j)$ and $\gamma_{i,j}(\rho) = \sum_{r=0}^{\infty} (r+1)\, \rho^r W^r(i,j)$. It follows that $\phi^*_{i,j}(\rho) > \gamma^*_{i,j}(\rho)$ if and only if

$$\frac{\left(\phi_{i,j}(\rho)\right)^2}{\left(\gamma_{i,j}(\rho)\right)^2} > \frac{\phi_{i,j}(\rho)\phi_{i,j}(\rho)}{\gamma_{i,i}(\rho)\gamma_{j,j}(\rho)}. \tag{1.12}$$

We show that (1.12) holds for any $i \neq j$, by showing that $\phi_{i,j}(\rho)\gamma_{i,i}(\rho) > \phi_{i,i}(\rho)\gamma_{i,j}(\rho)$ for any $i \neq j$. The last inequality may be rewritten as

$$\sum_{r=0}^{\infty} \left\{ \rho^r \sum_{k=0}^{r} \left[(k+1) W^k(i,j) W^{r-k}(i,i) \right] \right\} > \sum_{r=0}^{\infty} \left\{ \rho^r \sum_{k=0}^{r} \left[(k+1) W^k(l,i) W^{r-k}(i,j) \right] \right\}.$$

But this is clearly equivalent to

$$\sum_{r=0}^{\infty} \left\{ \rho^r \sum_{k=0}^{r} \left[k W^k(i,j) W^{r-k}(i,i) \right] \right\} > \sum_{r=0}^{\infty} \left\{ \rho^r \sum_{k=0}^{r} \left[k W^k(l,i) W^{r-k}(i,j) \right] \right\},$$

and therefore the lemma follows by applying inequality (1.7), which has been established in the course of proving Proposition 1.3.5.

Proof of Proposition 1.3.10 Rewrite $\Sigma_{s,1}$ as $V^{\frac{1}{2}}\bar{\Sigma}V^{\frac{1}{2}}$, where $\bar{\Sigma} = (I - \rho\bar{W})^{-1}(I - \rho\bar{W}')^{-1}$, with $\bar{W} = V^{-\frac{1}{2}}WV^{\frac{1}{2}}$. Consider the spectral decomposition $\sum_{l=1}^{n}\lambda_l(\bar{\Sigma})q_l(\bar{\Sigma})q_l(\bar{\Sigma})'$ of $\bar{\Sigma}$, where $q_l(\bar{\Sigma})$ is an eigenvector of $\bar{\Sigma}$ associated to the eigenvalue $\lambda_l(\bar{\Sigma})$. By using a power series expansion of $(I - \rho\bar{W})^{-1}$ (and of its transpose), plus the fact that $\bar{W}$ is similar to W, one sees that, for $0 < \rho < \lambda_n^{-1}$, $\bar{\Sigma}$ is nonnegative and irreducible. Thus $\lambda_n(\bar{\Sigma})$ is simple by the Perron-Frobenius theorem, and is the only eigenvalue that does not go to a finite number as $\rho \to \lambda_n^{-1}$, because, as it is easily seen, $\bar{\Sigma}^{-1}$ has rank $n-1$ when $\rho = \lambda_n^{-1}$. Next, observe that, as $\rho \to \lambda_n^{-1}$, $q_n(\bar{\Sigma}) \to V^{-1/2}q_n$, where q_n is the eigenvector of W associated to λ_n (because when $\rho = \lambda_n^{-1}$, $\bar{\Sigma}^{-1}$ has an eigenvector $V^{-1/2}q_n$ corresponding to its smallest eigenvalue 0). The proposition is then proved by using a spectral decomposition of $\bar{\Sigma}$ and applying a similar argument as in the last part of the proof of Proposition 1.3.6.

Proof of Proposition 1.3.11 Expressing it as the ratio of the cofactor of $\Sigma_{s,1}^{-1}(i,j)$ to the determinant of $\Sigma_{s,1}^{-1}$, it is apparent that $\phi_{i,j}(\rho)$ is a rational function and as such is completely determined by the values it takes on any non-empty open interval included in its domain. Hence, by the remark at the beginning of the proof of Proposition 1.3.7, in order to show that a correlation $\phi_{i,j}^{*}(\rho)$, for $i \neq j$, is symmetric about the origin or about the vertical axis, it suffices to prove that the covariance $\phi_{i,j}(\rho)$ satisfies the same symmetry in the region $|\rho| < \lambda_n^{-1}$. For $|\rho| < \lambda_n^{-1}$, the covariance matrix of a SAR(1) model admits the representation

$$\Sigma_{s,1} = \sum_{r=0}^{\infty}\{\rho^r W^r\}\, V \sum_{r=0}^{\infty}\{\rho^r (W')^r\}\,,$$

and therefore its (i,j)-th entry can be written as

$$\phi_{i,j}(\rho) = \sum_{r,s=1}^{\infty}\left\{\rho^{r+s}\sum_{l=1}^{n}\left[V(l,l)\,W^r(i,l)W^s(j,l)\right]\right\}.$$

From this point on the proof is similar to that of Proposition 1.3.7. If G is bipartite and $d(i,j)$ is odd, then it is impossible to find an l such that $W^r(i,l)W^s(j,l) > 0$ for positive integers r and s having the same parity. This is so because otherwise there would be a closed walk of length $d(i,j) + r + s$ in G, which would be odd; therefore G could not bipartite, for a closed walk of odd length always contains a cycle of odd length. By the same argument, if G is bipartite and $d(i,j)$ is even, then there is no l such that $W^r(i,l)W^s(j,l) > 0$ for some positive integers r and s having different parity. For $|\rho| < \lambda_n^{-1}$, we then have that for a

SAR(1) model on a bipartite graph,

$$\phi_{i,j}(\rho) = \begin{cases} \sum_{r,s=1;\ r+s\ odd}^{\infty} \{\rho^{r+s} \sum_{l=1}^{n} [V(l,l) W^r(i,l) W^s(j,l)]\}, & \text{if } d(i,j) \text{ is odd,} \\ \sum_{r,s=1;\ r+s\ even}^{\infty} \{\rho^{r+s} \sum_{l=1}^{n} [V(l,l) W^r(i,l) W^s(j,l)]\}, & \text{if } d(i,j) \text{ is even.} \end{cases}$$

The proposition follows from the fact that the function ρ^{r+s} is odd if $r+s$ is odd and even if $r+s$ is even.

Proof of Proposition 1.4.4 Rewrite $\Sigma_{c,1}$ as $\left(\rho^{-1}I - A\right)^{\#} / (\rho \left|\rho^{-1}I - A\right|)$, where $B^{\#}$ denotes the adjoint matrix of a matrix B, and let $G\backslash i$ be the graph obtained from G by deleting the vertex i. Then it is immediate that, for any $i = 1, ..., n$,

$$\Omega_{1,q}(i,i) = \left(\frac{P\left(G\backslash i; \rho^{-1}\right)}{\rho P\left(G; \rho^{-1}\right)} \right)^{q}, \tag{1.13}$$

where $P(G) = P(G; \rho^{-1})$ denotes the characteristic polynomial $\left|\rho^{-1}I - A\right|$ (in the formal variable ρ^{-1}) of the adjacency matrix of G. Since a graph G is walk-regular if and only if $P\left(G\backslash i\right)$ does not depend on $i \in V(G)$ (Godsil, 1993, p. 81), the variances $\Omega_{1,q}(i,i)$ do not depend on i if and only if G is walk-regular.

Proof of Theorem 1.4.5 (Sufficiency) Recall, the distance matrices of a distance-regular graph span a Bose-Mesner algebra. Since, as it is well known, a Bose-Mesner algebra is closed under matrix generalized inversion, we have that, when the graph G is distance-regular and for any $1 \leq p \leq d$ and $q > 0$, $\Omega_{p,q}$ is a linear combination of the distance matrices of G. Thus $\Omega_{p,q}$ is the covariance matrix of a G-stationary model, by Definition 1.4.3. (Necessity) Let D be $Span\{A_0, ..., A_d\}$. It is easily seen from representation (1.5) that $\Omega_{p,q} \in D$ (i.e. the model is G-stationary) implies $A_1^r \in D$, for $0 \leq r \leq d$. Since these $d+1$ powers are linearly independent (because for any $0 < t \leq d$, $A_1^t(i,j) \neq 0$ implies $A_1^r(i,j) = 0$, $0 \leq r < t$), they form a basis for D. Therefore, if the distance matrices of G belong to D, then they are polynomials of maximum degree d in A_1, which is a sufficient condition for G to be distance-regular (Bannai and Ito, 1984, pp. 190-192).

Proof of Theorem 1.4.6 (Sufficiency) Straightforward, because if G is distance-regular, then $Span\{A_0, ..., A_d\}$ is an algebra of symmetric matrices and hence a quadratic subspace. Hence any model $N_n(0, \Omega_{d,q})$, $q \geq 2$, on a distance-regular graph is a regular exponential family by obvious extension of Lemma 1.2.5. (Necessity) For a model $N_n(0, \Omega_{d,q})$, $q \geq 2$, to be regular exponential, $\Omega_{d,q}^{-1}$ must belong to a $(d+1)$-dimensional subspace of $\mathcal{L}(n)$, say $\Phi_{d,q}$. If $\Omega_{d,q}^{-1} \in \Phi_{d,q}$, also $A_1^r \in \Phi_{d,q}$, $0 \leq r \leq d$, because of representation (1.5). Since the matrices A_1^r, $0 \leq r \leq d$, are linearly independent they span $\Phi_{d,q}$, and hence G is distance-regular, by the same argument as in the proof of Theorem 1.4.5.

Chapter 2

Power Properties of Invariant Tests for Spatial Autocorrelation in Linear Regression

Abstract

Many popular tests for residual spatial autocorrelation in the context of the linear regression model belong to the class of invariant tests. This paper derives a number of exact properties of the power function of such tests. In particular, we extend the work of Krämer (2005, Journal of Statistical Planning and Inference 128, 489-496) by characterizing the circumstances under which the limiting power, as the autocorrelation increases, vanishes. More generally, the analysis in the paper sheds new light on how the power of invariant tests for spatial autocorrelation is affected by the matrix of regressors and by the spatial structure. A numerical study aimed at assessing the practical relevance of the theoretical results is included.

Keywords: Cliff-Ord test; invariant tests; linear regression model; point optimal tests; power; similar tests; spatial autocorrelation.

2.1 Introduction

Testing for residual spatial autocorrelation in the context of the linear regression model (e.g., Cliff and Ord, 1981, Anselin, 1988, Cressie, 1993) is now recognized as a crucial step in much empirical work in economics, geography and regional science. The present paper is concerned with finite sample power properties of tests used for this purpose. More specifically, our main objective is to understand how power is affected by the regressors and by the spatial structure.

So far, the power properties of tests for residual spatial autocorrelation have received much less attention than those of tests for residual serial correlation, and have mainly been studied by Monte Carlo simulation (see Florax and de Graaff, 2004, and references therein). Very few attempts have been made to derive exact properties of such tests, two notable exceptions being King (1981) and Krämer (2005). The former paper has established that the most popular test for spatial autocorrelation in regression residuals, the Cliff-Ord test, is locally best invariant for an important class of alternatives. The latter paper has generalized some results previously available for tests of serial autocorrelation (see Krämer, 1985, and Zeisel, 1989); in particular, Krämer (2005) has shown analytically that there are cases in which the power of some tests for spatial autocorrelation (namely, those whose associated test statistics can be expressed as ratios of quadratic forms in the regression residuals) can vanish as the spatial autocorrelation in the data increases. In general, it is fair to state that, while there is some evidence in the literature that the properties of tests for spatial autocorrelation can be very sensitive to the regressors and to the spatial structure, little is known about which combinations of regressors and spatial structures lead to low or high power.

Of course—given the popularity of the linear regression model and the pervasiveness of the issue of spatial autocorrelation in many empirical investigations—a large number of procedures are available for the purpose of testing for residual spatial autocorrelation, and one can choose among them on the basis of the suspected form of autocorrelation or on the basis of the desired properties of the testing procedure. In this paper, we confine ourselves to a rather simple, but extremely popular, framework. We assume that the regression errors follow a (first-order) *spatial autoregressive process* (e.g., Cressie, 1993) and we focus on *invariant tests* (e.g., Lehmann and Romano, 2005). Even in this simple setup the analytical investigation of exact power properties of tests is complicated. Because of the availability of many approximating techniques for power functions, this is not necessarily a problem when interest lies in the properties of a test in the context of a given model, i.e., when both the matrix of

regressors and the spatial structure are fixed. However, when interest is, as in this paper, in how the properties of a test depend on the regressors and on the spatial structure, none of the available numerical or analytical approximations is likely to yield conclusive results. One feature of our approach is that some new properties of the power function of invariant tests are deduced directly from the density of the pertinent *maximal invariant* avoiding the need for complicated expressions for power functions or approximations to them.

Our contributions are as follows. Firstly, we extend the results of Krämer (2005) in several directions: we formulate conditions, that are in general very simple to check, for the limiting (as the autocorrelation increases) power of any invariant test to be 0, 1, or in $(0, 1)$; we prove that, for any given spatial structure and irrespective of the size of the tests, there exists an infinite number of subspaces spanned by the regressors such that the limiting power of a locally best or point optimal invariant test vanishes; we characterize such "hostile" subspaces. Secondly, we discuss some conditions that are sufficient for unbiasedness of invariant tests for spatial autocorrelation and for monotonicity of their power function. Such conditions are not necessary, but they help to understand the causes of undesirable properties of the tests.

These results call for caution in interpreting the outcome of tests for residual spatial autocorrelation, especially when the number of degrees of freedom is low and large autocorrelation is suspected. The results have also implications outside a formal hypothesis testing framework, because they imply that there are circumstances in which the practice of interpreting the Cliff-Ord statistic, or even the Moran statistic when the model does not contain regressors, as an autocorrelation coefficient (e.g., Cliff and Ord, 1981, Anselin, 1988) cannot be justified.

The remainder of the paper is organized as follows. Section 2.2 presents the theoretical framework of the paper, i.e., the testing problem and the tests considered. Section 2.3 analyzes the limiting power of invariant tests for spatial autocorrelation. This is done by first considering the general case of a model with arbitrary regressors, and then specializing the results to zero-mean models. A numerical study of the practical relevance of the results is included. Further insights into the role played by the regressors and the spatial structure in determining the power of invariant tests of autocorrelation are gained in Section 2.4, by analyzing the conditions for unbiasedness of the tests and monotonicity of their power functions. Section 2.5 concludes. All proofs are collected in the Appendix.

2.2 The Setup

2.2.1 The Testing Problem

Consider a *fixed* and *finite* set of n observational units, such as the regions of a country, and let $y = (y_1, ..., y_n)'$, where y_i denotes the random variable observed at the i-th (according to some arbitrary ordering) unit. We assume that y is modelled according to a Gaussian linear regression model, i.e.,

$$y \sim N(X\beta, \sigma^2\Sigma(\rho)), \tag{2.1}$$

where X is a non-stochastic $n \times k$ matrix of rank $k < n$, β is a $k \times 1$ vector of unknown parameters, σ^2 is an unknown positive parameter and ρ is a scalar unknown parameter belonging to some connected subset Ψ of the set of values of ρ such that $\Sigma(\rho)$ is positive definite. We assume that the function $\rho \to \Sigma(\rho)$ is differentiable, and that the parameters of the model are identified (in the sense that the parameter space of the model does not contain two distinct points indexing the same distribution) and functionally independent. In this paper, we will often refer to the case of a general $\Sigma(\rho)$, but will be mostly concerned with the specific covariance structures implied by spatial autoregressive process.

There are two distinct classes of Gaussian spatial autoregressive processes: conditional autoregressive (CAR) processes and simultaneous autoregressive (SAR) processes. They are both discussed extensively in many books and articles in the statistics and econometrics literature (e.g., Whittle, 1954, Besag, 1974, Cliff and Ord, 1981, Anselin, 1988, Cressie, 1993), to which we refer for details concerning the construction and interpretation of the models. Here, we only briefly define the covariance matrices implied by the models. As in most of the theoretical and empirical literature on spatial autoregressive processes, we confine ourselves to *first-order* (or one-parameter) processes. Such processes are specified on the basis of a fixed $n \times n$ (*spatial*) *weights matrix* W, chosen to reflect a priori information on relations among the n observations. Typically, the (i, j)-th entry of W is set to zero if i and j are not neighbors according to some metric that is deemed to be relevant for the phenomenon under analysis, and is set to some non-zero number, possibly reflecting the degree of interaction, otherwise. For instance, if the observational units are the regions of a country, one may set $W(i, j) = 1$ if i and j share a common boundary, $W(i, j) = 0$ otherwise.

Let I denote the $n \times n$ identity matrix. Provided it is symmetric and positive definite, the matrix $\Sigma(\rho)$ implied by a CAR specification is

$$\Sigma(\rho) = (I - \rho W)^{-1}L, \tag{2.2}$$

where L is a fixed $n \times n$ diagonal matrix such that $L^{-1}W$ is symmetric ($\sigma^2 L(i,i)$ represents the variance of y_i conditional on all the remaining random variables in y). We remark that, even without reference to CAR models, structure (2.2) constitutes a very natural framework in which to study tests for autocorrelation; see, e.g., Anderson (1948), Kadiyala (1970), Kariya (1980), King (1980).

On the other hand, provided that $I - \rho W$ is nonsingular, a SAR process implies

$$\Sigma(\rho) = (I - \rho W)^{-1} V (I - \rho W')^{-1}, \tag{2.3}$$

where V is a fixed $n \times n$ symmetric and positive definite matrix.

For both CAR and SAR models, we assume:

(i) $W(i,i) = 0$, for $i = 1, ..., n$;

(ii) $W(i,j) \geq 0$, for $i, j = 1, ..., n$;

(iii) W is an *irreducible* matrix (e.g., Gantmacher, 1974, Ch. 13).

Condition (i) is required by the way CAR models are constructed, and is assumed for SAR models merely for convenience. Condition (ii) is not required by the definition of the models, but is virtually always satisfied in empirical applications. Condition (iii) is a natural assumption in a spatial context; in graph theoretic terms, it amounts to requiring that the graph with adjacency matrix W (that is, the graph with the n observational units as vertices and an edge from i to j if and only $W(i,j) \neq 0$) is *strongly connected*, i.e., has a path between any two distinct vertices (e.g., Cvetković *et al.*, 1980, p. 18). Observe that (non-circular) AR(1) models are not in our class of SAR processes, because the matrix W necessary to write the covariance matrix of an AR(1) process as in equation (2.3) would be triangular and hence reducible. Also note that, as a consequence of the symmetry of $L^{-1}W$, in CAR models $W(i,j) = 0$ if and only if $W(j,i) = 0$, for $i, j = 1, ..., n$. This implies that, in CAR models, W can be assumed to be irreducible without loss of generality, because otherwise the model could be decomposed into the product of (at least) two processes.

In the context of model (2.1), we wish to test the null hypothesis $\rho = 0$ versus the one-sided alternative $\rho > 0$ (here and throughout, $\rho > 0$ stands for $\mathbb{R}^+ \cap \Psi$, i.e., we leave it implicit that ρ must belong to the parameter space of the model). The choice of a one-sided alternative rather than a two-sided one is dictated by the fact that the former is more relevant in the context of many popular specification of $\Sigma(\rho)$, including those implied by CAR and SAR processes, due to the interpretation of ρ in such processes (see below).

We assume—and this is an important point for the reading of the present paper—that $\Sigma(0) = I$. Since, as far as our testing purposes are concerned, this does not involve any loss of generality, from now on and unless otherwise specified we reserve the term "CAR model" to refer to the family of distributions

$$N(X\beta, \sigma^2(I - \rho W)^{-1}), \tag{2.4}$$

(for a fixed W) and the term "SAR model" to refer to the family of distributions

$$N(X\beta, \sigma^2 \left[(I - \rho W')(I - \rho W)\right]^{-1}), \tag{2.5}$$

(again, for a fixed W). The normalization to $\Sigma(0) = I$ emphasizes a crucial difference between CAR and SAR models, with regards to our testing problem: for CAR models there is no loss of generality in assuming that W is a symmetric matrix, whereas for SAR models we have to allow explicitly the possibility of a nonsymmetric W. In fact, we shall see that there are substantial differences between SAR models with a symmetric weights matrix, henceforth referred to as *symmetric SAR models*, and SAR models with a nonsymmetric weights matrix—henceforth referred to as *asymmetric SAR models*. The most popular nonsymmetric weights matrices in applications are those obtained by row-standardizing a preliminary matrix (e.g., Anselin, 1988). In the rest of the paper, a *row-standardized* W is one that can be written as $W = D^{-1}A$, where A is a symmetric $(0-1)$ matrix and D is the diagonal matrix with $D(i,i) = \sum_{j=1}^{n} A(i,j)$, $i = 1, ..., n$.

By the Perron-Frobenius theorem, W admits a positive eigenvalue that is (algebraically and geometrically) simple and non-smaller in modulus than any other eigenvalue. We denote such an eigenvalue by λ. For both CAR and SAR models we take the set $\mathbb{R}^+ \cap \Psi$ to be the interval $(0, \lambda^{-1})$. Such a restriction is necessary for positive definiteness of the covariance matrix of a CAR model. For a SAR model, it is not necessary, but has the advantage of guaranteeing connectedness of the parameter space and of avoiding some undesirable properties of the covariance structure implied by the model.

When $0 < \rho < \lambda^{-1}$, it is easily established (e.g., Gantmacher, 1974, p. 69) that conditions (ii) and (iii) imply that the covariance between any two variables y_i and y_j in both CAR and SAR models is strictly positive (similarly, it can be shown that when $\rho < 0$ the covariances may be positive or negative, but not all of them are positive in a left neighborhood of λ^{-1}). Thus, the hypothesis $\rho > 0$ represents positive spatial autocorrelation, a much more common phenomenon in practice than negative spatial autocorrelation.

2.2.2 The Tests

This paper is concerned with *invariant* tests (see, e.g., Lehmann and Romano, 2005). Roughly speaking, these are the tests that preserve the symmetries satisfied by the testing problem in question. More precisely, a test is said to be invariant with respect to a certain group of transformations of the sample space if it is based on a test statistic that is constant on each orbit of that group. A necessary and sufficient condition for this type of invariance is that the test statistic is a function of a *maximal invariant* under that group.

Our problem of testing $\rho = 0$ against $\rho > 0$ in model (2.1) is invariant with respect to the group of transformations $y \longrightarrow ay + Xb$, where $a \in \mathbb{R}^+$ and $b \in \mathbb{R}^k$ (e.g., Kariya, 1980, or King, 1980). A maximal invariant under this group is $v = Cy/\|Cy\|$, where C is an $(n-k) \times n$ matrix such that CC' is the identity matrix of order $n-k$ and $C'C$ is $M = I - X(X'X)^{-1}X'$, and $\|\cdot\|$ denotes the Euclidean norm. For some positive integer m, let $S_m = \{v \in \mathbb{R}^m : \|v\| = 1\}$ denote the unit m-dimensional sphere. The distribution of v depends on the single parameter ρ, and has density, with respect to the normalized Haar measure on S_{n-k},

$$pdf(v;\rho) = \left|C\Sigma(\rho)C'\right|^{-\frac{1}{2}} \left[v'\left(C\Sigma(\rho)C'\right)^{-1}v\right]^{-\frac{n-k}{2}} \tag{2.6}$$

(see Kariya, 1980, equation (3.7)). Since $pdf(v;\rho)$ does not depend on v when ρ vanishes, testing $\rho = 0$ in $N(X\beta, \sigma^2\Sigma(\rho))$ by means of an invariant test reduces to testing uniformity of v on S_{n-k}.

Besides the "principle of invariance", there are at least two other reasons why invariant tests are particularly appealing for our testing problem. Firstly, invariant tests can be implemented easily. Since an invariant test statistic must depend on y only through v, its distribution under the null (and also under the alternative) is free of nuisance parameters, and hence critical values can, in general, be obtained accurately by Monte Carlo or other numerical methods. In fact, the class of *similar* tests for $\rho = 0$ coincides with that of invariant tests (Hillier, 1987). Secondly, expression (2.6) turns out to be proportional to the *marginal likelihood* of ρ, which has often been found to provide a better basis for inference about ρ than the full likelihood of model (2.1) (particularly when k is large with respect to n); see, for instance, Diggle (1988), Tunnicliffe Wilson (1989) and Rahman and King (1997).[1]

[1] The literature on the comparison between maximum likelihood and restricted maximum likelihood, REML, is also relevant here, although REML usually refers to the marginal likelihood of all the covariance parameters, i.e., both ρ and σ^2 in our case.

Despite the elimination of the nuisance parameters achieved in (2.6), it is well known that, in general, a uniformly most powerful invariant (UMPI) test does not exist for the testing problem under consideration. In such a situation, one can resort to a test that is optimal according to some exact criterion (see, for instance, Cox and Hinkley, 1974, p. 102), or to a test that has less clear-cut optimality properties but performs well in general, such as a likelihood ratio test (which is an invariant test, as proved for instance in Cox and Hinkley, 1974, p. 173) or its restricted version based on $pdf(v;\rho)$. The present paper is particularly concerned with the tests—named point optimal invariant (POI) tests by King (1988)—that are the most powerful amongst all invariant tests against a specific alternative $\rho = \bar{\rho} > 0$, and with the locally best invariant (LBI) test, which is obtained as the limiting case for $\bar{\rho} \to 0$. In general, and certainly for our testing problem, the locally most powerful test coincides with the test maximizing the slope of the power function at $\rho = 0$ (see Lehmann and Romano, 2005, p. 339). The choice of POI and LBI tests is mainly motivated by the fact that POI tests define the upper bound (the so-called power envelope, see below) to the power attainable by any invariant test of a fixed size, and by the popularity of LBI tests, especially in the context of the spatial models we are concerned with in this paper.

The size of a critical region (henceforth c.r.) is denoted by α and, to avoid trivial cases and unless otherwise specified, is assumed to be in $(0,1)$. The critical value for a size-α test will be denoted by c_α. The POI (or best invariant) c.r. at the point $\bar{\rho}$, obtained by application of the Neyman-Pearson Lemma to the density (2.6), is defined by

$$v' \left(C\Sigma(\bar{\rho})C'\right)^{-1} v < c_\alpha. \tag{2.7}$$

Denoting by $\pi_{\bar{\rho}}(\rho)$ the power of such a c.r., the power envelope of size-α invariant tests is the function that associates the value $\pi_\rho(\rho)$ to each $\rho \geq 0$. When needed, we will emphasize the dependence of $\pi_\rho(\rho)$ on X by writing $\pi_\rho(\rho, X)$. The LBI c.r. for testing $\rho = 0$ against $\rho > 0$ is defined by

$$v'CA_0C'v < c_\alpha, \tag{2.8}$$

where $A_0 = d\Sigma^{-1}(\rho)/d\rho\big|_{\rho=0}$ (King and Hillier, 1985). When $-A_0$ is equal to some spatial weights matrix W (or to $W + W'$), as it is in the case of CAR or SAR models, the LBI test is known in the literature as Cliff-Ord test (see Cliff and Ord, 1981, and King, 1981). The Cliff-Ord test represents the generalization to regression residuals of the Moran test (Moran, 1950), and is, by far, the most popular test for spatial autocorrelation in regression models.[2]

[2]Being based on only the first derivative of $\Sigma(\rho)$, the LBI tests for $\Sigma(\rho) = I$ have generally non-trivial

Before we continue, some notation is in order. For a $q \times q$ symmetric matrix Q, we denote by $\lambda_1(Q), ..., \lambda_q(Q)$ its eigenvalues, labeled in non-decreasing order of magnitude; by $m_i(Q)$ the multiplicity of $\lambda_i(Q)$, for $i = 1, ..., q$; by $f_1(Q), ..., f_q(Q)$ a set of orthonormal (with respect to the Euclidean norm) eigenvectors of Q, with the eigenvector $f_i(Q)$ being pertinent to the eigenvalue $\lambda_i(Q)$; by $E_i(Q)$ the eigenspace associated to $\lambda_i(Q)$, for $i = 1, ..., q$. In all of the above quantities, we suppress the reference to Q when Q is a (symmetric, with $q = n$) weights matrix. Thus, for a symmetric W, $\lambda_n = \lambda$.

2.3 Limiting Power

This section contains the main results of the paper. Broadly speaking, they concern the role of the regressors in determining power properties of invariant tests for autocorrelation. In Section 2.3.1, we discuss some preliminary results in the context of the general model (2.1). Then, we focus on the limiting behavior of the power function, as the autocorrelation increases, in CAR and SAR models, with general regressors in Section 2.3.2, and without regressors in Section 2.3.3. Finally, in Section 2.3.4 we report results from numerical experiments aimed at assessing the practical relevance of the theoretical results.

2.3.1 Preliminaries

Consider the issue of how X in $N(X\beta, \sigma^2\Sigma(\rho))$ affects the the power properties of invariant tests of $\rho = 0$ versus $\rho > 0$, for some covariance structure $\Sigma(\rho)$. The following proposition sets the scene for the analysis to follow. It is concerned with comparing the envelope $\pi_\rho(\rho, X)$, for an $X \neq 0$, with the envelope when $X = 0$ (here and throughout a zero matrix is simply denoted by 0).

Proposition 2.3.1 *Consider testing $\rho = 0$ versus $\rho > 0$ in model $N(X\beta, \sigma^2\Sigma(\rho))$. For any $X \neq 0$, any $\rho > 0$, and any α,*

$$\pi_\rho(\rho, X) \leq \pi_\rho(\rho, 0). \tag{2.9}$$

In (2.9) equality is attained if and only if, for some $i = 2, ..., n-1$, $\text{col}(X) \subseteq E_i(\Sigma(\rho))$ and $\alpha = \Pr(v'\Sigma^{-1}(\rho)v < \lambda_i^{-1}(\Sigma(\rho))$.

power against a large class of alternative specifications of $\Sigma(\rho)$, which is at the same time a strength and a weakness of such tests. In any case, this does not detract from the fact that it is important to study their performance against particular alternatives, spatial autoregressive models in our case.

The conditions for equality in (2.9) are extremely restrictive, because they pose very severe constraints on X, α, and $\Sigma(\rho)$. Proposition 2.3.1 asserts that, except when these conditions are met, the presence of any $X \neq 0$ in $N(X\beta, \sigma^2\Sigma(\rho))$ has a detrimental effect (with respect to the case $X = 0$, and as long as β is unknown) on the maximum power achievable by an invariant test for testing $\rho = 0$ versus $\rho > 0$.

Two comments arise naturally from Proposition 2.3.1. The first comment is that the comparison in the proposition involves models (the one with $X = 0$ and the one with an $X \neq 0$) with different degrees of freedom. An interesting question is which matrices X of a fixed dimension $n \times k$ are favorable, and which are less favorable, to our testing purposes (from the point of view of the maximum power achievable by invariant tests). Such a question is a difficult one, because, for a given matrix $\Sigma(\rho)$, in general the answer *depends on* ρ *and* α. Some partial answers are available in the literature for regression models with AR(1) errors; see, e.g., Tillman (1975). The second comment is that, in practice, one is usually more concerned with the power of a specific test than with the power envelope. For a general $\Sigma(\rho)$, Proposition 2.3.1 does certainly not imply that the power function of a particular test when $X = 0$ is uniformly (over $\rho > 0$) non-smaller than when $X \neq 0$ (it is interesting, however, that such an implication does hold when $\Sigma(\rho)$ is that of a CAR model and the test in question is a POI or LBI test, because for a zero-mean CAR model the POI c.r. (2.7) does not depend on $\bar{\rho}$, i.e., there exists a UMPI test, and hence the power function of any POI or LBI test coincides with $\pi_\rho(\rho, 0)$).

To deal with the issues raised in the previous paragraph, we will focus on large values of ρ in CAR and SAR models. Exact power properties of invariant tests will be deduced directly from the density of the maximal invariant v. For convenience, we now list some fundamental properties of $pdf(v;\rho)$, valid for any fixed ρ. Let $\Omega_\rho = C\Sigma(\rho)C'$, and let $\tilde{E}_i(\Omega_\rho) = S_{n-k} \cap E_i(\Omega_\rho)$, $1 \leq i \leq n-k$. The density $pdf(v;\rho)$ is antipodally symmetric (that is, $pdf(v;\rho) = pdf(-v;\rho)$) and, more specifically, is constant on the regions of constant $v'\Omega_\rho^{-1}v$ (geometrically, such regions are the intersection of the surfaces of a sphere and of an ellipsoid in $\mathbb{R}^{n-k}$). It follows immediately that:

(i) $pdf(v;\rho)$ is maximized over S_{n-k} when $v'\Omega_\rho^{-1}v$ is minimized, that is, when $v \in \tilde{E}_{n-k}(\Omega_\rho)$. Note that $\tilde{E}_{n-k}(\Omega_\rho)$ consists of two antipodal points if and only if $m_{n-k}(\Omega_\rho) = 1$;

(ii) $pdf(v;\rho)$ is strictly decreasing as v moves from $\tilde{E}_i(\Omega_\rho)$ to $\tilde{E}_j(\Omega_\rho)$ along any geodesic of S_{n-k}, for any $1 \leq j < i \leq n-k$;

(iii) upon rotation to a coordinate system provided by a set of $n-k$ orthogonal eigenvectors of Ω_ρ, $pdf(v;\rho)$ is invariant to the sign of each component of the vector v.

Property (i) will be used to derive some of the results to follow. Property (ii) implies that any invariant c.r. that is not centrally symmetric (we say that an invariant c.r. $\Phi \in S_{n-k}$ is centrally symmetric if $t \in \Phi$ implies $-t \in \Phi$) is dominated uniformly over ρ (in terms of power) by a centrally symmetric c.r. of the same size. Because of this reason, from now on we assume that an invariant c.r. is centrally symmetric. This corresponds to enlarging the group of transformations with respect to which we require invariance to include also the transformation $y \rightarrow -y$. Property (iii) is not exploited directly in this paper, but is very useful when thinking geometrically about our testing problem, for it implies that the study of $pdf(v;\rho)$ can be limited to a single orthant of S_{n-k}.

The following preliminary result links the limit, as ρ approaches some positive value a (from the left, and with a an accumulation point of Ψ), of the power function of an arbitrary invariant c.r. to the limiting eigenstructure of Ω_ρ. We denote $\lim_{\rho \rightarrow a} \Omega_\rho$ by Ω, the limit being taken entrywise.

Lemma 2.3.2 *Suppose that $\Sigma(\rho)$ is positive definite for $\rho \in (0,a)$ and for $\rho \rightarrow a$. If $\lambda_{n-k}(\Omega)$ is finite, then the power of any invariant c.r. for testing $\rho = 0$ against $\rho > 0$ in model $N(X\beta, \sigma^2\Sigma(\rho))$ tends, as $\rho \rightarrow a$, to a number strictly between 0 and 1. If $\lambda_{n-k}(\Omega)$ is infinite and simple, then the power of an invariant c.r. for the same testing problem tends, as $\rho \rightarrow a$, to 1 if the c.r. contains $f_{n-k}(\Omega)$, to 0 otherwise.*

Lemma 2.3.2 holds for a very general class of matrices $\Sigma(\rho)$, including CAR and SAR models, and the (time-series) stationary AR(1) model. For the latter model, the power of the Durbin-Watson and some related tests as $\rho \rightarrow 1$ has been investigated extensively; see Krämer (1985), Zeisel (1989) and Bartels (1992). Lemma 2.3.2 shows how some results on the power of such tests can be extended to any invariant (similar) test for serial correlation.

Of course, when $a \in \Psi$ the power of any c.r. must be in $(0,1)$, and this is reflected in Lemma 2.3.2 by the fact that $\lambda_{n-k}(\Omega)$ is finite when $a \in \Psi$. The possibility that, in the setting of Lemma 2.3.2, the power of a certain c.r. vanishes as ρ goes to the boundary of Ψ should be regarded as a problem of the statistical model, rather than of a particular test. A simple geometric argument clarifies this point. Let $\nu = n - \lim_{\rho \rightarrow a}\{rank(\Sigma^{-1}(\rho))\}$, with $\Sigma(\rho)$ as in Lemma 2.3.2. Note that for $\lambda_{n-k}(\Omega) = \infty$ it is necessary that $\nu > 0$. When $\nu > 0$, the model $N(X\beta, \sigma^2\Sigma(\rho))$ tends, as $\rho \rightarrow a$, to a family of (improper) distributions defined on a

ν-dimensional subspace, say S_ν, of $\mathbb{R}^n$. This is easily seen by observing that the contours of $N(X\beta, \sigma^2\Sigma(\rho))$ are the ellipses $(y - X\beta)'\Sigma^{-1}(\rho)(y - X\beta) = k$, which also shows that, for any fixed β, S_ν is the translation of $\lim_{\rho \to a}\{E_n(\Sigma(\rho))\}$ by $X\beta$. It is then clear that the limiting power, as $\rho \to a$, of a certain test—not necessarily invariant—depends on the position of the c.r. in $\mathbb{R}^n$ relative to S_ν (if the test is invariant, the relative position does not depend on β). In particular, the power of a test vanishes whenever the intersection between S_ν and the c.r. has 1-dimensional Lebesgue measure zero. In CAR and SAR models (when $a = \lambda^{-1}$) and in stationary AR(1) models (when $a = 1$), $\nu = 1$.[3] Note that the stationarity assumption on the AR(1) model is not superfluous, in that generally $v = 0$ otherwise. On the contrary, for the CAR and SAR models considered in this paper v is always 1. This represents an important difference between time-series and spatial autoregressive models, from the point of view of testing for residual autocorrelation.

Clearly, the above geometric argument does not depend on the normality assumption, but holds for any elliptically symmetric distribution. It also holds for any c.r.; when the c.r. is invariant, the conditions in Lemma 2.3.2 can be exploited. Moreover, further progress can be made by focusing on a specific class of matrices $\Sigma(\rho)$, those implied by CAR and SAR models in the rest of this section.

2.3.2 Main Results

In this subsection we focus on the power of invariant tests in CAR and SAR models when $\rho \to \lambda^{-1}$ (from the left). Accordingly, from now on, by "limiting power" we mean the limit of the power function as $\rho \to \lambda^{-1}$. The restriction to the case $\rho \to \lambda^{-1}$ is of practical relevance, because (a) it corresponds to studying power when it is most needed, i.e., when the autocorrelation in the data, and hence the inefficiency of the OLS estimator of β, is large; (b) often, in order to fit real data, spatial autoregressive models require a large value of ρ (e.g., Besag and Kooperberg, 1995).

In order to state the key result of this section some new notation is needed. Henceforth, an invariant critical region defined as a subset of S_{n-k} is denoted by Φ_v, whereas its image on the sample space $\mathbb{R}^n$ is denoted by Φ_y. The column space of the matrix X, often referred to as the

[3] A very similar situation occurs in regression models such that $\Sigma(\rho)$, rather than $\Sigma^{-1}(\rho)$, tends to a singular matrix as ρ tends to some vale a. In this case, the distributions are defined, as $\rho \to a$, in a subspace of $\mathbb{R}^n$ of dimension $\lim_{\rho \to a}\{rank(\Sigma(\rho))\}$. Examples are a spatial moving average model (i.e., a model with covariance matrix equal to the inverse of that of a SAR model), and a fractionally integrated white noise, with ρ being the differencing parameter and $a = 1/2$ (see Kleiber and Krämer, 2005).

"regression space", is denoted by $\text{col}(X)$. The entrywise positive and normalized eigenvector of W pertaining to λ is denoted by f. Existence and uniqueness of f are guaranteed by the Perron-Frobenius theorem.

Theorem 2.3.3 *In CAR and SAR models, the limiting power of an invariant c.r. Φ_y for testing $\rho = 0$ against $\rho > 0$ is:*

- *in* $(0,1)$ *if* $f \in \text{col}(X)$;
- 1 *if* $f \in \Phi_y \backslash \text{col}(X)$;
- 0 *otherwise.*

The theorem asserts that, to some degree, the limiting power of Φ_y is determined by which of three mutually disjoint subsets of the sample space f belongs to. Of course, the result can be restated on the space S_{n-k} of the maximal invariant, in which case f must be replaced by $Cf/\|Cf\|$ and the three subsets become $\{0\}$, $\Phi_v \backslash \{0\}$ and $\overline{\Phi_v \cup \{0\}}$.

Theorem 2.3.3 is strongly related to Theorems 1 and 2 in Krämer (2005), the most important differences being: (a) the class of tests considered there (i.e., tests that can be expressed as ratios of quadratic forms in the regression residuals) and the class considered in the present paper (i.e., invariant tests) are different, although they certainly intersect; (b) our result does not require symmetry of W. We stress that Theorem 2.3.3 holds for *any* invariant (similar) c.r., regardless of the analytical form of the associated test statistic. Thus, it also holds for tests whose test statistics are analytically complicated, or, as in the case of a likelihood ratio test based on the full or the marginal likelihood, unavailable in closed form.

The practical usefulness of Theorem 2.3.3 is in providing simple conditions for the limiting power of any invariant c.r. to vanish, given any matrices X and W. Consider an invariant c.r. that rejects $\rho = 0$ for small values of some statistic $T(y)$, i.e.,

$$\Phi_y = \{y \in \mathbb{R}^n : T(y) < c_\alpha\}. \tag{2.10}$$

Theorem 2.3.3 asserts that the limiting power of such a c.r. is 0 if $T(f) < c_\alpha$, 1 if $T(f) \geq c_\alpha$, in (0,1) if $f \in \text{col}(X)$. These conditions are typically very simple to check because, in most cases, (i) f is either known (e.g., it is a vector of identical entries when W is row-standardized) or can be computed efficiently (e.g., by the power method); (ii) since Φ_y is similar, c_α can be obtained accurately by simulation methods. For instance, it is readily verified that, for

CAR or SAR models, the limiting power of a POI test is 0, 1, or strictly between 0 and 1, depending on whether

$$f'\left(\Sigma(\bar{\rho})M_{\bar{\rho}} - c_\alpha M\right) f \tag{2.11}$$

is respectively positive, negative, or zero (zero occurring if and only if $f \in \text{col}(X)$). Analogously, the limiting power of a LBI test is 0, 1, or strictly between 0 and 1, depending on whether

$$f'\left(MA_0M - c_\alpha M\right) f \tag{2.12}$$

is respectively positive, negative or zero. Note that, since they refer to test statistics that are ratios of quadratic forms in y, the conditions based on (2.11) and (2.12) reduce, in the case of a symmetric SAR model, to conditions given in Krämer (2005). The specific form of the test statistics also implies that, for POI or LBI tests, c_α can also be obtained by exploiting one of the many approximations available for the distribution of a quadratic form in a vector uniformly distributed on a hypersphere.

Further remarks concerning Theorem 2.3.3 follow.

Remark 2.3.4 The condition $f \in \text{col}(X)$, under which the limiting power of an invariant test is in $(0,1)$, is satisfied whenever W in a CAR or SAR model is row-standardized and an intercept is included in the regression, because row-standardization implies that f has identical entries. Note that here whether W refers to a model before or after normalization to $\Sigma(0) = I$ is irrelevant, because the condition $f \in \text{col}(X)$ is invariant under any invertible linear transformation of $y \sim N(X\beta, \sigma^2\Sigma(\rho))$, where $\Sigma(\rho)$ is that of a CAR or SAR model. For any weights matrix that is not row-standardized, in general $f \notin \text{col}(X)$, with the consequence that the limiting power of an invariant test is either 0 or 1.

Remark 2.3.5 An important and immediate consequence of Theorem 2.3.3 is that, in CAR and SAR models, the limit of the envelope $\pi_\rho(\rho)$ as $\rho \to \lambda^{-1}$ is 1 if $f \notin \text{col}(X)$, and is in $(\alpha, 1)$ otherwise. Hence, as $\rho \to \lambda^{-1}$ in CAR and SAR models, the null hypothesis $\rho = 0$ can be distinguished (by means of an invariant tests) from the alternative hypothesis $\rho > 0$ with zero type II error probability only if $f \notin \text{col}(X)$.

Remark 2.3.6 Consider an invariant test, constructed on the basis of some assumed weights matrix. By Theorem 2.3.3, whether its limiting power is 0, 1, or in $(0,1)$ depends on the "true" W appearing in the CAR or SAR model only through f (which, for instance, is the same for any row-standardized W), and does not depend on whether the model is a CAR or a SAR model. This property implies some robustness of invariant tests to model misspecification, when the spatial autocorrelation is large.

In the rest of this subsection we take a close look at the case in which the limiting power vanishes, and, consequently, we restrict attention to the case $f \notin \text{col}(X)$.

Suppose that, for a given CAR or a SAR model, one finds, by application of Theorem 2.3.3, that the limiting power of a certain c.r. Φ_y vanishes. Theorem 2.3.3 itself guarantees that if Φ_y is enlarged so as to include f, then its limiting power jumps to 1. From a practical point of view, a question of concern is how large Φ_y must be in order to avoid the vanishing of the limiting power. We define (allowing, for convenience and contrary to what is done elsewhere in the paper, α to take the value 1):

Definition 2.3.7 *For an invariant c.r. for testing $\rho = 0$ against $\rho > 0$ in a CAR or SAR model, α^* is the infimum of the set of values of $\alpha \in (0, 1]$ such that the limiting power does not vanish.*

When $f \notin \text{col}(X)$, α^* is a measure of the distinguishability between the null hypothesis $\rho = 0$ and the alternative $\rho \to \lambda^{-1}$. A large α^* indicates that a large size is necessary to avoid the zero limiting power problem; $\alpha^* = 1$ means that the limiting power is 0 for any α; $\alpha^* = 0$ indicates that the limiting power is 1 for any α.[4]

When an invariant c.r. is in form (2.10) (and $f \notin \text{col}(X)$), α^* is the probability that $T(y) < T(f)$ under the null hypothesis $y \sim N(X\beta, \sigma^2 I)$, or, by invariance,

$$\alpha^* = \Pr(T(y) < T(f);\ y \sim N(0, I)). \tag{2.13}$$

Thus, α^* can be computed accurately by simulation or other numerical methods. We stress that α^* depends on X (through $\text{col}(X)$, because of the invariance property of the tests), W, the choice of test and the choice between a CAR and a SAR specification. In particular, for a given error process, a given test, and a given k, α^* may depend to a very large extent on $\text{col}(X)$. Numerical examples will be given in the next subsection. In the following, we will explore the dependence of α^* on $\text{col}(X)$ by studying the circumstances in which $\alpha^* = 0$ and those in which $\alpha^* = 1$. We will first give a lemma that holds for any test based on a quadratic form in the maximal invariant, and then we will apply the lemma to POI and LBI tests. Extensions of the analysis below to more general tests (a likelihood ratio test, say) are possible, but may be more involved.

Consider an invariant c.r. of the form

$$\Phi_v(B) = \{v \in S_{n-k} : v'Bv < c_\alpha\}, \tag{2.14}$$

[4]Recall that we are here focusing on the case $f \notin \text{col}(X)$. If $f \in \text{col}(X)$, α^* is always zero, by Theorem 1, and hence uninformative. In order to study the power of invariant tests when $f \in \text{col}(X)$, one could define α^* as the infimum of the set of values such that the limiting power is greater than some positive value, but this is not pursued in the present paper.

where B is an $(n-k) \times (n-k)$ known symmetric matrix independent of α. For instance, any c.r. based on a ratio of quadratic forms in the OLS residuals can be written in this form. Typically B will depend on X and W (but could depend on a weights matrix different from the one appearing in CAR models, thus allowing for the possibility of misspecification of W). We have:

Lemma 2.3.8 *Consider, in the context of CAR or symmetric SAR models, testing $\rho = 0$ against $\rho > 0$ by means of a c.r. $\Phi_v(B)$. Provided that $f \notin \text{col}(X)$, $\alpha^* = 0$ if and only if $Cf \in E_1(B)$, and $\alpha^* = 1$ if and only if $Cf \in E_{n-k}(B)$.*

Lemma 2.3.8 implies that, in a CAR or SAR model and for a c.r. $\Phi_v(B)$, $\alpha^* \in (0,1)$ as long as Cf is not an eigenvector of B associated to the smallest or the largest eigenvalue of B. The important question remains of whether the extremes $\alpha^* = 0$ and $\alpha^* = 1$ are attainable, and, if so, in which circumstances. In particular, it is of interest to understand whether for a fixed W in a CAR or SAR model, α^* has a non-trivial (i.e., smaller than 1) upper bound as $\text{col}(X)$ ranges over the set of all subspaces of $\mathbb{R}^n$ of low (with respect to n) dimension. Obviously, given a certain model and a certain c.r., one would hope that α^* is small, since the limiting power vanishes whenever $\alpha^* > \alpha$.

To answer the above question we focus on POI tests ($B = \Omega_{\bar{\rho}}^{-1}$) and LBI tests ($B = CA_0C'$). First, we consider the case $\alpha^* = 0$. Two conditions that are easily seen to lead to $\alpha^* = 0$ (i.e., to $Cf \in E_1(B)$) for POI and LBI tests are (i) W symmetric and $X = 0$ and (ii) W symmetric and $f \perp \text{col}(X)$. More generally, the following sufficient condition can be established.

Proposition 2.3.9 *Consider, in the context of CAR or symmetric SAR models, testing $\rho = 0$ against $\rho > 0$ by means of a POI or LBI c.r. Provided that $f \notin \text{col}(X)$, $\alpha^* = 0$ if $E_{n-k}(\Omega_\rho)$ does not depend on ρ for $\rho > 0$.*

Remark 2.3.10 Although the condition in Proposition 2.3.9 is not necessary, a simple geometric argument suggests that when the condition is not met α^* is 0 only in very special circumstances. Let the center of the c.r. Φ_v based on a certain test statistic be the set of points of S_{n-k} that are in Φ_v for any α. For instance, the center of $\Phi_v(B)$ is $E_1(B) \cap S_{n-k}$. Clearly, for a certain c.r., $\alpha^* = 0$ if and only if, as $\rho \to \lambda^{-1}$, $pdf(v;\rho)$ vanishes anywhere outside the center of that c.r. Now, as long as $f \notin \text{col}(X)$, $pdf(v;\rho)$ tends, as $\rho \to \lambda^{-1}$, to be concentrated in the direction of $v = Cf$ (see Section 2.3.1). Thus, for a POI test ($B = \Omega_{\bar{\rho}}^{-1}$), $\alpha^* = 0$ if and only if $Cf \in E_1(\Omega_{\bar{\rho}}^{-1}) = E_{n-k}(\Omega_{\bar{\rho}})$. This is the case if $E_{n-k}(\Omega_\rho)$ does not depend on ρ for $\rho > 0$ (by Proposition 2.3.9), but, otherwise, poses a

strong restriction on the trajectories described on S_{n-k} by the eigenvectors in $E_{n-k}(\Omega_{\bar{\rho}})$.

We now turn to characterize the case $\alpha^* = 1$ for POI and LBI tests. Theorem 1 of Krämer (2005) contains the crucial statement that, in symmetric SAR models, "given any matrix W of weights, and independently of sample size, there is always some regressor X such that for the Cliff–Ord test the limiting power disappears" (note that here "some regressor X" means $k = 1$). Now, from Theorem 2.3.3 it is clear that whether or not a particular X (with $k \geq 1$) causes the limiting power to disappear depends on α. Thus, if interpreted as holding for any α (less than 1), the above statement would imply that for any W there exist some particularly hostile regressors that cause a zero limiting power even when the size of the Cliff-Ord c.r. (i.e., the LBI c.r.) is very large. This is clearly an extremely strong property, in a negative sense, of a c.r. Unfortunately, whether it holds or not for the Cliff-Ord test in the context of a symmetric SAR model remains to be established, because the proof of Krämer's theorem holds only when $\alpha \to 0$.[5] The next theorem settles the issue and places it in a more general context. Recall that m_1 denotes the multiplicity of λ_1, for a symmetric W. Unless W satisfies particular symmetries, generally $m_1 = 1$ (see, for instance, Biggs, 1993).

Theorem 2.3.11 *Consider, in the context of CAR or symmetric SAR models, testing $\rho = 0$ against $\rho > 0$ by means of a POI or LBI c.r. For any fixed W, there exist m_1-dimensional regression spaces such that the limiting power of the selected c.r. vanishes, irrespective of α. For instance, when $m_1 = 1$, let X be a vector proportional to $f_1 + bf$, for some $b \in \mathbb{R}$. Then, the limiting power of POI and LBI tests vanishes, irrespective of α, if $|b| \geq b^*$, where b^* is a threshold that depends on the model and on the c.r. Namely, letting*

$$b_1 = \left(\frac{\lambda - \lambda_2}{\lambda_2 - \lambda_1}\right)^{\frac{1}{2}}, \; b_2 = \frac{1 - \bar{\rho}\lambda_1}{1 - \bar{\rho}\lambda}, \; b_3 = \frac{2 - \bar{\rho}(\lambda + \lambda_2)}{2 - \bar{\rho}(\lambda_2 + \lambda_1)},$$

b^* is equal to $b_1\sqrt{b_2}$ for a POI c.r. in a CAR model, $b_1 b_2 b_3$ for a POI c.r. in a symmetric SAR model, b_1 for a LBI c.r. in both models.

Theorem 2.3.11 establishes that, for any fixed W in a CAR or symmetric SAR model, there are m_1-dimensional regression spaces such that $\alpha^* = 1$. In the presence of such regression spaces, the zero limiting power problem *cannot* be solved by increasing α. Note that if an m_1-dimensional regression space causes a zero limiting power (of a POI or LBI test in a CAR or

[5] This is because d_1 in equation (12) of Krämer (2005) is not necessarily positive for any W, unless $\alpha \to 0$. As a consequence, the regressors that Krämer constructs in his proof do not need to cause the limiting power to vanish when $d_1 < 0$.

symmetric SAR model), then also all the k-dimensional regression spaces, with $k \geq m_1$, that contain it but do not contain f will yield a zero limiting power, as an obvious consequence of the fact that the power of an invariant test does not depend on β.

We now aim to show that, in the context of Theorem 2.3.11, the vanishing of the limiting power is not an event of measure zero (in a sense to be specified). In order to do so, it is convenient to introduce some new notation. Let $G_{k,n}$ denote the set, known as a Grassmann manifold, of all k-dimensional subspaces of $\mathbb{R}^n$, and let $H_k(\alpha) \subseteq G_{k,n}$, for $0 < \alpha < 1$, be the set of k-dimensional $\text{col}(X)$ such that the limiting power of a POI or LBI c.r. of size less than α vanishes (for some CAR or symmetric SAR model). Clearly, $H_k(\alpha_1) \subseteq H_k(\alpha_2)$ for any $\alpha_1 \geq \alpha_2$.

A natural measure of the size of $H_k(\alpha)$ is the probability that $\text{col}(X) \in H_k(\alpha)$, as $\text{col}(X)$ ranges over $G_{k,n}$ according to some probability distribution (with respect to the invariant measure on $G_{k,n}$, as given in James, 1954). Such a probability, which we denote by z_α, can be interpreted as the *probability of a zero limiting power* of a size-α POI or LBI test, in a CAR or symmetric SAR model (we stress that, for each realization of X, POI and LBI tests are derived by treating X as fixed). We have:

Proposition 2.3.12 *Consider, in the context of CAR or symmetric SAR models, testing* $\rho = 0$ *against* $\rho > 0$ *by means of a POI or LBI c.r. If* $\text{col}(X)$ *has density that is almost everywhere positive on* $G_{k,n}$, $k \geq m_1$, *then* $z_\alpha > 0$, *for any* W *and regardless of how large* α *or* $n - k$ *is.*

Clearly, in some circumstances z_α can be very small (e.g., z_α is usually small when $n - k$ or α are large). The important point made by Proposition 2.3.12 is that, under the stated conditions, z_α is *never* zero. In Section 2.3.4 we will compute $z_{0.05}$ numerically for some choice of W and of the probability distribution of $\text{col}(X)$.

We now provide an interpretation of the m_1-dimensional regression spaces $\text{col}(X)$ that, according to Theorem 2.3.11, are particularly hostile for testing $\rho = 0$ versus $\rho > 0$ when ρ is large. Starting from $m_1 = 1$, Theorem 2.3.11 asserts that the set of such regression spaces is a region, defined by b^*, of the plane spanned by f_1 and f; for a POI test, it is easily seen that, as $\bar{\rho}$ increases, this set becomes smaller and more concentrated in the direction of f. Generalizing to $m_1 \geq 1$, the set of the hostile regression spaces is a certain region of the $(m_1 + 1)$-dimensional subspace of $\mathbb{R}^n$ spanned by the vectors $f_1, ..., f_{m_1}, f$ (see the proof of the theorem). Consider the Moran statistic $x'Wx/x'x$ associated to a vector

$x \in \mathbb{R}^n$ and a symmetric W (the standard version of the Moran statistic would include a normalizing factor and a correction for the sample mean of x that are not relevant here). By the Rayleigh-Ritz theorem (e.g., Horn and Johnson, 1985), f represents a vector that is most autocorrelated according to the Moran statistic, and $f_1, ..., f_{m_1}$ represent vectors that are least autocorrelated. Note that λ_n, the value of the Moran static when $x = f$, is positive by the Perron-Frobenius theorem, and λ_1, the value of the Moran static when $x = f_1, ..., f_{m_1}$, is negative for $tr[W] = \sum_{i=1}^{n} \lambda_i = 0$ by assumption. Thus, Theorem 2.3.11 asserts that in CAR and symmetric SAR models it is difficult, or even impossible, to detect large positive spatial autocorrelation in the presence of regressors that can be expressed as the sum of a strongly positively autocorrelated component and a strongly negatively autocorrelated component, with the former component being the dominant one.

We mention an extension of Theorem 2.3.11 that is directly related to the interpretation just given, and can be proved similarly to Theorem 2.3.11. If $f_j + bf \in \text{col}(X)$, with $f_j \notin E_{n-1}$ and $f_j \notin E_n$, the limiting power of a LBI test in a CAR or symmetric SAR model is 1 for any α (i.e., $\alpha^* = 0$) provided that

$$|b| \leq \left(\frac{\lambda - \lambda_{n-1}}{\lambda_{n-1} - \lambda_j} \right)^{\frac{1}{2}}.$$

Expressions of this sort can be used to infer how W affects (through its spectrum, under Gaussianity) the power properties of tests of $\rho = 0$. For instance, if W is such that $\lambda - \lambda_{n-1}$ is large, then any vector X in a large region of the plane spanned by f_j and f yields $\alpha^* = 0$.

Some further remarks concerning Theorem 2.3.11 end this section.

Remark 2.3.13 With regards to the statement from Krämer (2005) reported above, Theorem 2.3.11 (i) establishes that the statement is correct when $m_1 = 1$; (ii) provides a generalization to the case $m_1 > 1$; (iii) provides a generalization to POI tests and to CAR models.

Remark 2.3.14 The strongest implication of Theorem 2.3.11 is perhaps that regression spaces such that the limiting power of a POI test vanishes exist even when $\bar{\rho}$ is large (i.e., close to λ^{-1}) and α is large. This is surprising because, by Proposition 2.4.1 below, the power at $\bar{\rho}$ of a POI test must be larger than α. Since, if $f \notin \text{col}(X)$, $\pi_\rho(\rho) \to 1$ as $\rho \to \lambda^{-1}$ (see Remark 2.3.5), it also holds that the supremum—as $\text{col}(X)$ ranges over the set of all k-dimensional, $k \geq m_1$, subspaces of $\mathbb{R}^n$—of the maximum shortcoming (e.g., Lehmann and Romano, 2005, p. 337) of any POI or LBI c.r. is always one, for any W and any α.

Remark 2.3.15 For POI and LBI tests and for any W, $T(f)$, regarded as a function from $G_{k,n}$ to $\mathbb{R}$, is continuous. Thus, by (2.13), α^* is itself a continuous, and generally smooth, function of $\text{col}(X)$,

which implies, in particular, that the regression spaces that are sufficiently close (according to some distance on $G_{k,n}$) to regression spaces yielding a large (resp. small) α^* yield a large (resp. small) α^*.

Remark 2.3.16 We have not attempted to generalize Theorem 2.3.11 to asymmetric SAR models, for two reasons. Firstly, such models generally do not satisfy the condition $f \notin \text{col}(X)$ necessary for the zero limiting problem. This is because the nonsymmetric weights matrices generally used in SAR models are row-stochastic, implying, as already noted above, that $f \in \text{col}(X)$ as long as an intercept is included in the regression. Secondly, although the proof of Theorem 2.3.11 suggests that regression spaces (of low dimension) such that the limiting power of a POI or LBI c.r. vanishes for any α always exist also in the context of asymmetric SAR models, the exact characterization of such regression spaces appears to be more involved. It should be noted, however, that an approximated characterization can be obtained from Theorem 2.3.11, by approximating an asymmetric SAR model by a CAR model with $\Sigma^{-1}(\rho) = I - \rho(W + W')$ (i.e., omitting terms in ρ^2).

2.3.3 Zero-Mean Models

In this subsection we specialize some of the above results to zero-mean (or constant-mean, by obvious extension) CAR and SAR models. For our purposes, setting $X = 0$ in the models analyzed above has two main advantages. Firstly, it clarifies—by direct comparison with the regression case—the role played by the regressors in determining power. Secondly, it allows to focus on the effect of the specification of W on power.

Let us start from the following corollary of Theorem 2.3.3.

Corollary 2.3.17 *In zero-mean CAR and SAR models, the limiting power of an invariant c.r. Φ_y for testing $\rho = 0$ against $\rho > 0$ is 1 for any α if $f \in \Phi_y$, 0 otherwise.*

It is instructive to relate Corollary 2.3.17 to the Moran statistic $y'Wy/y'y$. In the context of CAR and SAR models, the Moran statistic is usually interpreted as an autocorrelation coefficient. In view of this interpretation, the result in Corollary 2.3.17 is precisely what one would expect when W is symmetric, since in that case $y = f$ maximizes the Moran statistic. The same cannot be said when W is nonsymmetric, because in that case the Moran statistic is not, in general, maximized by f.

In fact, the differences between models with symmetric W (CAR and symmetric SAR models) and models with nonsymmetric W (asymmetric SAR models) are not only a matter of interpretation. We provide an example in the context of possibly the simplest SAR model; the same model was used by Whittle (1954) in his seminal paper on spatial autoregressions.

Example 2.3.18 A random variable is observed at n units placed along a line and, in the context of a zero-mean SAR process, it is to be tested whether $\rho = 0$ or $\rho > 0$. Suppose that it is believed that there is only first-order interaction and that the interaction amongst first-order neighbors is stronger in one direction than in the other. Accordingly, W is chosen so that $W(i,j)$ is equal to some fixed positive scalar $w \neq 1$ if $i - j = 1$, to 1 if $j - i = 1$, and to 0 otherwise, for $i, j = 1, ..., n$. In Figure 2.1, we plot the power function of the LBI test, i.e., the Moran test, and the envelope $\pi_\rho(\rho)$ for $n = 6$, $w = 10$ and $\alpha = 0.01$. The power has been computed numerically, via the Imhof method (Imhof, 1961), and is plotted against $\rho\lambda$, which ranges between 0 and 1.

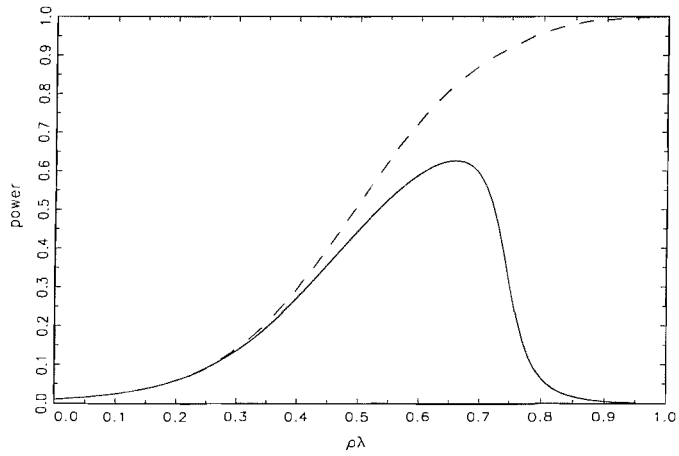


Figure 2.1: The power function of the Moran test (solid line) and the envelope $\pi_\rho(\rho)$ (dashed line) for the zero-mean asymmetric SAR model described in Example 2.3.18.

Although it is based on a model with an artificial W (for more practically relevant models, see Section 2.3.4), Figure 2.1 illustrates the theoretically important point that in a SAR model with nonsymmetric W, the limiting power of the Moran test may vanish even when the model is not contaminated by regressors. On the contrary, when W is symmetric, the power function of the Moran test always goes to 1 as $\rho \to \lambda^{-1}$ (by Lemma 2.3.8) and—as we shall see in Proposition 2.4.3 below—is monotonic. Note that this feature of the power function of the Moran test entails that there are zero-mean asymmetric SAR models in which the interpretation of the Moran statistic $y'Wy/y'y$ as an autocorrelation coefficient cannot be justified, because for such models there exist values $0 < k < \lambda^{-1}$ such that $\Pr(y'Wy/y'y > k)$ is not increasing over $0 < \rho < \lambda^{-1}$. The next result gives further insights into the problem.

Proposition 2.3.19 *In zero-mean SAR models, the limiting power of a POI or LBI c.r. for*

testing $\rho = 0$ against $\rho > 0$ is 1 for any α if and only if f is an eigenvector of W'.

The weights matrices W satisfying the condition in Proposition 2.3.19 are those such that λ is perfectly well-conditioned (e.g., Golub and Van Loan, 1996, p. 323). In practice, it turns out that the condition is very restrictive when W is nonsymmetric (whereas it is trivially satisfied when W is symmetric), and hence that in asymmetric SAR models typically $\alpha^* > 0$ even when $X = 0$. For row-standardized W's—the most popular, by far, nonsymmetric weights matrices in SAR models—the restrictiveness of the condition is emphasized by the following result.

Corollary 2.3.20 *In zero-mean asymmetric SAR models with row-stochastic W, the limiting power of a POI or LBI c.r. for testing $\rho = 0$ against $\rho > 0$ is 1 for any α if and only if W is doubly stochastic.*

Clearly, a nonsymmetric row-stochastic weights matrix W is doubly stochastic, i.e., has not only all rows but also all columns summing to 1, only in very special cases.[6] The condition in Proposition 2.3.19 remains very unlikely to be satisfied also for nonsymmetric W's that are not row-stochastic. This is essentially because, given any choice of the neighborhood structure of a set of observational units (i.e., any choice of the pairs of units deemed to be neighbors) the choice of weights yielding a well-conditioned λ is typically a very particular one,[7] and corresponds to some relevant notion of distance amongst the units only in exceptional cases.

Having argued that the condition in Proposition 2.3.19 is generally not satisfied, the interesting issue becomes to understand which (nonsymmetric) matrices W are associated to large values of α^*. Let us return to our example of a SAR model defined on a line.

Example 2.3.21 For the case of Example 2.3.18 above, the Imhof method (or some other numerical approximation to the null distribution of the Moran statistic) can be used to verify that α^* is decreasing in n and increasing in $|w - 1|$. For the particular case of Figure 1, α^* is about 0.056. Note that if one closes the line to form a circle (by setting $W(1, n) = w$

[6] Formally, this can be deduced from Birkhoff's theorem on doubly stochastic matrices, which states that any such matrix must be a convex combination of permutation matrices; e.g., Horn and Johnson, 1985. We remark that the doubly stochastic weights matrices used in SAR models by Pace and LeSage (2002) are symmetric.

[7] That such a choice exists can be seen by starting from a (nonsymmetric) matrix W, and transforming it to $S^{-1}WS$, where S is a diagonal matrix with $S(i,i) = (f_i/l_i)^{1/2}$, with l denoting the left eigenvector of W associated to λ.

and $W(n,1) = 1$), then W becomes a scalar multiple of a doubly stochastic matrix, and consequently $\alpha^* = 0$ by Proposition 2.3.19.

Numerical investigations not reported here show that, typically, for a fixed n, large values of α^* are associated to matrices W such that $W(i,j)/W(j,i)$ is large for at least one pair (i,j) (we note that this type of asymmetry yields large values of α^* even when $X \neq 0$). This suggests that the asymmetry introduced by using row-standardized weights matrices $W = D^{-1}A$ (see Section 2.2.1) does not yield very large values of α^* in zero-mean SAR models, because for such matrices $W(i,j)/W(j,i) \leq u(A)$, $i,j = 1,...,n$, where $u(A)$ denotes the ratio of the largest to the smallest row-sum of A. Note that the largest possible value of $u(A)$ over all $n \times n$ matrices A is $n-1$, obtained for the adjacency matrix of a star graph (i.e., a graph with one vertex having $n-1$ neighbors, and all other vertices having 1 neighbor). One can check that, even in this case, the value of α^* associated to the corresponding row-standardized W is very small, and decreasing in n.[8] For instance, for the Moran test, when W is the row-standardized version of the adjacency matrix of a star graph, $\alpha^* > 0.01$ only when $n < 6$. Thus, to summarize, in SAR models asymmetry of W may cause the limiting power of POI and LBI tests to disappear even when $X = 0$; for row-standardized W's, however, this typically occurs only for very small values of α or n.

2.3.4 Numerical Examples

In this subsection we report numerical results aimed at illustrating how X and W affect the exact power of tests for residual spatial autocorrelation. More specifically, the objective is to show how sensitive power can be to X when ρ is large but not necessarily in a small neighborhood of λ^{-1}, in some situations of practical interest. For simplicity, we restrict attention to the power, which we denote by $\pi_{LBI}(\rho)$, of the Cliff-Ord test in the context of a SAR model. Related numerical investigations are contained in Krämer (2005).

For some selected specifications of W, we conduct Monte Carlo experiments where X is drawn from some probability distribution, and the power is computed by the Imhof method. Because of its invariance property, the power of the Cliff-Ord test depends on X only through col(X). A natural choice for the distribution of X would then be to take $vec(X) \sim N(0, I_{nk})$,

[8] Interestingly, the effect of the asymmetry of a row-standardized weight matrix $D^{-1}A$ (or any other non-symmetric matrix that is similar to a symmetric matrix) can always be eliminated by suitably selecting V in (2.3). In fact, model (2.3) with $W = D^{-1}A$ and $V = D^{-1}$ is reduced, upon normalization to $\Sigma(0) = I$, to a SAR model with symmetric weight matrix $D^{-1/2}AD^{-1/2}$.

because this would imply that $\text{col}(X)$ is uniformly distributed on the Grassmann manifold $G_{k,n}$ (see James, 1954, for the definition of uniform distribution on $G_{k,n}$). Since, however, an intercept is in practice always included in the regression, we prefer to take $X = [\iota \mid X_1])$, with $vec(X_1) \sim N(0, I_{n(k-1)})$ (the effect on power of including an intercept will be discussed below). In the results reported below, $k = 2$, i.e., the regression includes just an intercept and an i.i.d. standard normal variate. The simulation is based on 10^6 replications of X. All computations are done in GAUSS v7. We set $\alpha = 0.05$.

We construct weights matrices from the maps of the $n = 17$ counties of Nevada and the $n = 23$ counties of Wyoming. We consider both a binary W, specified according to the queen criterion (i.e., $W[i,j] = 1$ if counties i and j share a common boundary or a common point, $W[i,j] = 0$ otherwise), and its row-standardized version. The average number of neighbors of a county is 4.35 in Nevada, 4.52 in Wyoming, whereas the sparseness of W (as measured by the percentage of zero entries) is 74.40 for Nevada and 80.34 for Wyoming. We shall see that, despite their similarities, these two spatial configurations are very different with respect to our testing purposes.

Firstly, in order to show how sensitive $\pi_{LBI}(\rho)$ is to X, in Table 2.1 we display the percentage frequency distribution of $\pi_{LBI}(\rho)$, with W as described above. We report values for $\rho = 0.9\lambda^{-1}$ and $\rho = 0.95\lambda^{-1}$, which represent points at which low power is particularly troublesome (because of the large inefficiency of the ordinary least squares estimator of β), but that are not too close to λ^{-1}. Note that, by Theorem 2.3.3, in our experiment $\lim \pi_{LBI}(\rho)$ (as $\rho \to \lambda^{-1}$) is either 0 or 1 when W is binary (as in that case $f \notin \text{col}(X)$ almost surely), whereas it is in $(0,1)$ when W is row-standardized (as in that case $f = \iota \in \text{col}(X)$). It appears from Table 2.1 that in the case of Nevada $\pi_{LBI}(\rho)$ depends to a very large extent on X, even at points that are relatively far from λ^{-1}. The dependence is less pronounced in the case of Wyoming.

Next, we consider the zero limiting power problem more closely, which requires restricting attention to binary weights matrices (so that $f \notin \text{col}(X)$ almost surely). In Table 2.2 we display $z_{0.05}$ (see Section 2.3.2), obtained as the frequency of times that (2.12) (with c_α computed by the Imhof method) is positive in our experiment. Note that $z_{0.05}$ is very large in the case of Nevada, whereas it is very small in the case of Wyoming. The table also displays the average shortcoming (i.e., $\pi_\rho(\rho) - \pi_{LBI}(\rho)$) of the Cliff-Ord test at $\rho = 0.9\lambda^{-1}$ and $\rho = 0.95\lambda^{-1}$, when $\lim \pi_{LBI}(\rho) = 0$ and when $\lim \pi_{LBI}(\rho) = 1$. It appears that the impact of the zero limiting power problem is not localized only in a very small neighborhood

Table 2.1: Percentage frequency distribution of the power $\pi_{LBI}(\rho)$ of the Cliff-Ord test, in model $y = X\beta + \varepsilon$, where ε is a SAR process and X contains an intercept and a standard normal variate. The power is computed by the Imhof method over 10^6 replications of X.

		$\pi_{LBI}(\rho)$						
	$\rho\lambda$	0.3-0.4	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1
Nevada								
binary W	0.90	0.11	0.25	28.42	71.05	0.17	·	·
	0.95	0.29	5.75	36.29	53.43	4.11	0.13	·
row-st W	0.90	·	·	0.02	0.16	41.47	58.35	·
	0.95	·	·	0.01	0.05	1.56	98.38	·
Wyoming								
binary W	0.90	·	·	·	0.02	0.69	99.29	·
	0.95	·	·	·	0.02	0.10	1.76	98.12
row-st W	0.90	·	·	·	·	·	0.50	99.50
	0.95	·	·	·	·	·	·	100

of λ^{-1}, because, on average, an X causing $\lim \pi_{LBI}(\rho) = 0$ causes shortcomings at $\rho = 0.9\lambda^{-1}$ and $\rho = 0.95\lambda^{-1}$ that are significantly larger than the corresponding shortcomings associated to an X such that $\lim \pi_{LBI}(\rho) = 1$.[9]

We remark that the probability z_α is generally very sensitive to W, n, k, the choice of a test, α, and the distribution of X. In most situations, z_α is small (but positive under the condition in Corollary 2.3.12) when $n - k$ is large (although it is possible to construct matrices W, e.g., the adjacency matrix of a star graph or a very dense matrix, such that this is not the case). This suggests that, from a practical point of view, the zero limiting power problem is mainly a small sample problem. In general, and interestingly, z_α is significantly larger when the regression includes an intercept. This is because, due to the nonnegativity of W, ι usually (and especially if the row sums of W are all of similar magnitude) yields a large value of the Moran statistic, and therefore its presence tends to put more probability mass on the regression spaces close to the hostile ones defined by Theorem 2.3.11. When W is defined on a regular grid, one can study how z_α depends on n explicitly (see Table 1 of Krämer,

[9]Note that when $m_1 = 1$, as in the examples we are considering, and $\mathrm{col}(X)$ contains the span of a vector $f_1 + bf$ with large b, the power function goes to zero (by Theorem 2.3.11), but it does so very rapidly, because the condition $f \in \mathrm{col}(X)$ is nearly satisfied and therefore the power function tends to be close to that when $f \in \mathrm{col}(X)$, which goes to a positive number as $\rho \to \lambda^{-1}$.

Table 2.2: Probability of zero limiting power ($z_{0.05}$) and average shortcoming of the Cliff-Ord test, in the case of a binary W.

		av. shortc. at $\rho\lambda = 0.90$		av. shortc. at $\rho\lambda = 0.95$	
	$z_{0.05}$	$\pi_{LBI}(\rho) \to 0$	$\pi_{LBI}(\rho) \to 1$	$\pi_{LBI}(\rho) \to 0$	$\pi_{LBI}(\rho) \to 1$
Nevada	0.77	0.20	0.16	0.32	0.24
Wyoming	$5.2 \cdot 10^{-4}$	0.15	0.03	0.26	0.02

2005). Note that z_α is related to the measure α^* by the relation $z_\alpha = \Pr(\alpha^* > \alpha)$ (where the randomness of α^* is due to that of X). In our experiment, α^* varied between $2.8 \cdot 10^{-4}$ and 0.994 for the case of Nevada, between $2.1 \cdot 10^{-7}$ and 0.430 for the case of Wyoming.

The main conclusion of our numerical study is that, in some cases of practical interest, the probability that the limiting power of the Cliff-Ord test vanishes may well be non-negligible. This obviously induces a large dependence of the power of the Cliff-Ord test on X when $\rho \to \lambda^{-1}$, but the numerical results indicate that both the power and the shortcoming may still depend to a large extent on X for values of ρ in a rather large neighborhood of λ^{-1}.

2.4 Unbiasedness and Monotonicity

In this section we discuss some conditions on model $N(X\beta, \sigma^2\Sigma(\rho))$ that are sufficient for POI and LBI tests to be unbiased (for a general $\Sigma(\rho)$) and to have power functions monotonic in ρ (for CAR or symmetric SAR models). The conditions are by no means necessary, but (i) are important to understand the structure of the testing problem under analysis; (ii) in the case of spatial autoregressive models, admit a simple interpretation.

We start from the following known, and fundamental, fact: any POI test is strictly unbiased for testing $\rho = 0$ against the specific alternative $\rho = \bar{\rho}$ for which it is constructed to be optimal. This property, for the general regression model (2.1), was derived in Theorem 1 of Kadiyala (1970) by an astute, but somewhat indirect, argument. For convenience, we restate the result in terms of the power envelope $\pi_\rho(\rho)$, and we point out (see the proof) that the result is a straightforward consequence of the Neyman-Pearson lemma.

Proposition 2.4.1 *In model $N(X\beta, \sigma^2\Sigma(\rho))$, the inequality $\pi_\rho(\rho) > \alpha$ holds for any $\rho > 0$.*

Proposition 2.4.1 is a very general result, since it holds for any X and any $\Sigma(\rho)$. However, it cannot be used to establish unbiasedness of a particular invariant test for $\rho = 0$ against

$\rho > 0$, except of course when a UMPI test exists (which is a very restrictive condition, because it requires the c.r. defined by (2.7) to be independent of $\bar{\rho}$). Next we formulate two conditions that, when taken together, lead to unbiasedness of POI and LBI tests.

By a commuting family of matrices it is meant a finite or infinite set of matrices that are pairwise commutative under standard multiplication.

Condition A *The matrices* $\Sigma(\rho)$, *for* $\rho > 0$, *form a commuting family.*

Condition A is particularly relevant in the present paper because it is satisfied by CAR and symmetric SAR models. Except for very special cases, it is not satisfied by asymmetric SAR models. A well-known characterization of a commuting family of symmetric matrices is that all its members share the same eigenvectors. This explains, in view of Proposition 2.3.9, why $\alpha^* = 0$ in zero-mean CAR and symmetric SAR models, whereas generally $\alpha^* > 0$ in zero-mean asymmetric SAR models. An important advantage of Condition A is that it allows a natural extension of many properties of the models $N(0, \sigma^2\Sigma(\rho))$ to the models $N(X\beta, \sigma^2\Sigma(\rho))$ that satisfy the next condition.

Condition B *For a fixed* $\widetilde{\rho} > 0$, $\mathrm{col}(X)$ *is spanned by* k *linearly independent eigenvectors of* $\Sigma(\widetilde{\rho})$.

An interpretation of Condition B in CAR and SAR models will be given at the end of this section. Because of the characterization mentioned above, if Condition A holds, Condition B does not depend on $\widetilde{\rho}$. Condition B, in any of its many equivalent formulations, has played a crucial role in the theoretical analysis of regression models with non-spherical errors since Anderson (1948). Although Condition B is unlikely to be met in practice, in some circumstances one may expect it to hold approximately (see the end of this section for CAR and symmetric SAR models, and Durbin, 1970, for the case of serial correlation). There is evidence in the literature that the power properties of tests for $\rho = 0$ when Condition B holds exactly are similar to those when Condition B holds approximately (e.g., Tillman, 1975, p. 971).

Letting $\mathrm{col}^{\perp}(X)$ denote the orthogonal complement of $\mathrm{col}(X)$, we have:

Proposition 2.4.2 *Assume that Conditions A and B hold. Then, in model* $N(X\beta, \sigma^2\Sigma(\rho))$, *any POI or LBI c.r. for testing* $\rho = 0$ *against* $\rho > 0$ *is unbiased. The unbiasedness is strict except when* $\mathrm{col}^{\perp}(X)$ *is a subset of an eigenspace of* $\Sigma(\rho)$, *in which case the power is* α *for any* $\rho > 0$.

In Proposition 2.4.2, as in Propositions 2.4.3 and 2.4.6 below, "any" means for any α and any $\bar{\rho}$. It is worth pointing out that, in general, Conditions A and B are not sufficient for the existence of a UMPI test for the stated testing problem, and therefore Proposition 2.4.2 is not a consequence of Proposition 2.4.1. An important counterexample in which a UMPI test exists is a CAR model satisfying Condition B (the reason why Condition B combines particularly well with a CAR specification is that the resulting model is an exponential family with number of sufficient statistics equal to the number of parameters, $k+2$). It should also be noted that Conditions A and B are not sufficient for the monotonicity in ρ of the power functions of the tests in Proposition 2.4.2, not even when $X = 0$, because, given a $\Sigma(\rho)$ satisfying Condition A, a reparametrization $\rho \to f(\rho)$ may destroy the monotonicity of the power function without causing Condition A to fail. Note that while unbiasedness is a vital property of any c.r., monotonicity of the power function in ρ is a much stronger property and may or may not be desirable depending on the specification of $\Sigma(\rho)$. In general, it is desirable whenever ρ is interpreted as an autocorrelation parameter. This is the case for CAR and SAR models. We can prove:

Proposition 2.4.3 *Assume that Condition B holds. Then, in CAR and symmetric SAR models the power function of any POI and LBI c.r. for testing $\rho = 0$ against $\rho > 0$ is non-decreasing. It is strictly increasing except when $\mathrm{col}^{\perp}(X)$ is a subset of an eigenspace of W, in which case the power is α for any $\rho > 0$.*

Proposition 2.4.3 implies that in CAR and symmetric SAR models having zero mean or, more generally, satisfying Condition B, the LBI and POI test statistics can be regarded as indexes of (residual) autocorrelation, in that they are non-decreasing (as any correlation between pairs of variables in CAR and SAR models) in ρ, over $(0, \lambda^{-1})$. Another important consequence of Proposition 2.4.3 is the monotonicity of the envelope $\pi_\rho(\rho)$, for CAR and symmetric SAR models satisfying Condition B. One would expect the same property to hold for zero-mean asymmetric SAR models, but, so far, we have found neither a proof nor a counterexample (by numerical analysis).

Remark 2.4.4 The power functions in Proposition 2.4.3 are, in fact, typically strictly increasing, because, unless $n - k$ is small or an eigenspace of W has large dimension (see Example 2.4.7 below), the chances of $\mathrm{col}^{\perp}(X)$ falling into an eigenspace of W are very low. In the special case $X = 0$, the power functions must be strictly increasing, for $\mathrm{col}^{\perp}(X) = \mathbb{R}^n$ cannot be an eigenspace of W.

Remark 2.4.5 For CAR models, Proposition 2.4.3 can alternatively be proved by showing that the

density $pdf(v;\rho)$ has a monotone likelihood ratio under Condition B, and then by using Theorem 3.4.1 of Lehmann and Romano (2005). Such an argument, however, does not extend to symmetric SAR models.

As it provides a link to the analysis in Section 2.3, the following result is also of interest.

Proposition 2.4.6 *Assume that Condition B holds. Then, in CAR and symmetric SAR models the limiting power of any POI and LBI c.r. for testing $\rho = 0$ against $\rho > 0$ is 1 if $f \notin \mathrm{col}(X)$; strictly between α and 1 if $f \in \mathrm{col}(X)$ and $\mathrm{col}^{\perp}(X)$ is not a subset of an eigenspace of W; α otherwise.*

We now provide a discussion of CAR and symmetric SAR models satisfying Condition B. From a practical perspective, the discussion is helpful to understand in which circumstances Condition B can be expected to hold approximately. We start from some examples. The most obvious case of a CAR model that satisfies Condition B is a model with mean assumed to be unknown but constant across observations and with a row-standardized W (see Section 2.2.1). On setting $L = D^{-1}$ and normalizing to $\Sigma(0) = I$, the mean of the model becomes proportional to $D^{\frac{1}{2}}\iota$, where ι is the n-dimensional vector of all ones, and the covariance matrix becomes $\Sigma(\rho) = \sigma^2(I - \rho D^{-\frac{1}{2}} A D^{-\frac{1}{2}})^{-1}$. Condition B is then satisfied because $D^{\frac{1}{2}}\iota$ is an eigenvector of $D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ (since ι is an eigenvector of $D^{-1}A$) and hence of $\Sigma(\rho)$. When other regressors are included in the model, a case in which Condition B has some chances of being met in practice is when the number of eigenspaces of W (and hence of $\Sigma(\rho)$, for CAR and symmetric SAR models) is small relative to n. This typically occurs when W satisfies a large number of symmetries, in the sense of being invariant under a large group of permutations of its index set (e.g., Biggs, 1993). The extreme case of equicorrelation serves as an illustration.

Example 2.4.7 In the context of CAR and SAR models, all regression errors are equicorrelated when W has constant off-diagonal entries (and zero diagonal entries). In that case, W is invariant with respect to the whole symmetric group on n elements and has only two eigenspaces, the one spanned by ι and the hyperplane orthogonal to it. Thus, in the case of equicorrelation, in order for Condition B to be met it suffices that every regressor in the model satisfies a single linear constraint, namely, that its entries sum to zero. Interestingly, if X contains an intercept, then $\mathrm{col}^{\perp}(X)$ is a subset of an eigenspace of W, and thus the power function of a POI or LBI c.r. is flat by Proposition 2.4.3.

The cases discussed above, albeit theoretically important, are of limited practical relevance in non-experimental contexts, since other regressors are typically used along with an intercept and the matrices W are usually not highly regular. We therefore take a more general view. Call two units i and j neighbors if $W(i,j) > 0$. Consider, for simplicity, the case when there is only one regressor, $x = (x_1, ..., x_n)'$ say, and let $\bar{x}_i$ be the weighted average $\sum_{j \neq i} W(i,j)x_j$ of the values of x observed at units that are neighbors of i (the extension to $k > 1$ is obvious). For CAR and symmetric SAR models, the eigenvectors of $\Sigma(\rho)$ are the same as those of W. Hence, in such models Condition B is met if and only if the ratio $x_i/\bar{x}_i$ does not depend on i, because, by the definition of an eigenvector of W, such a ratio must, for any i, be equal to the corresponding eigenvalue. Now, the ratio $x_i/\bar{x}_i$ may be regarded as a measure of "similarity" (as far as x is concerned) between i and its neighbors. This suggests that Condition B is approximately met (and hence the power of optimal invariant tests has desirable properties) when x is such that the degree of similarity between i and its neighbors does not change substantially with i.

2.5 Conclusion

The paper has investigated a number of properties of invariant/similar tests for autocorrelation in the context of a linear regression model with errors following a first-order conditional or simultaneous spatial autoregressive process. The main message of our analysis is that the power properties of exact tests for residual spatial autocorrelation may depend to a very large extent on the regressors, especially when the number of degrees of freedom is small and the autocorrelation is large. Intuitively, this is largely due to the fact that CAR and SAR models tend, as the autocorrelation increases, to a family of (improper) distributions on a 1-dimensional subspace of the sample space. If, in the context of a CAR or SAR model, the regressors are such that the intersection between such a subspace and a critical region has 1-dimensional Lebesgue measure zero, then the power of that critical region vanishes in the limit.

More formally, we have characterized the cases when the limiting power of invariant tests vanishes and we have shown that the minimum size α^* such that the limiting power of a POI or LBI test does not vanish may, for some spatial structures, depend on $\text{col}(X)$ to a very large extent. Furthermore, we have established that the sets of regression spaces $\text{col}(X)$ causing a zero limiting power of a size-α POI or LBI test have non-zero (invariant) measure on the set

$G_{k,n}$ of all k-dimensional subspaces of $\mathbb{R}^n$, for any α, any spatial structure and any $k > m_1$. In fact, in some circumstances, the probability content of these subsets (according to some distribution on $G_{k,n}$) may be far from negligible.

A remark concerning the distributional assumptions underlying our results is in order. As is well known (e.g., Kariya, 1980), the density of the maximal invariant (2.6) remains the same for any elliptically symmetric distributions of y, so the assumption of Gaussianity is much more than what is required to study the properties of the test considered in this paper. It should be noted, however, that while the generalization of SAR models to non-Gaussian distributions is straightforward, that is not so for CAR models; see Besag (1974).

Two possible extensions of our work are as follows. Firstly, although in this paper we have mostly focused on the power as $\rho \to \lambda^{-1}$, the techniques we have used should also prove useful to study local power, namely by studying the right derivative of the power function at $\rho = 0$. Secondly, an extension to *mixed regressive, spatial autoregressive* models (e.g., Ord, 1975, and Lee, 2002), which are not in the class of regression models (2.1) and therefore have not been considered in this paper, would be of interest.

Appendix: Proofs

Proof of Proposition 2.3.1 For some matrix X, denote by G_X the group of transformations $y \to ay + Xb$, with $a \in \mathbb{R}^+$ and $b \in \mathbb{R}^k$, and by $\Upsilon_X(\bar{\rho})$ the size-α POI c.r., defined on the sample space. By definition, $\Upsilon_X(\bar{\rho})$ is the size-α c.r. that is invariant under G_X and has maximum probability content under $N(X\beta, \sigma^2\Sigma(\bar{\rho}))$. Observe that, for any X, the probability content $\pi_{\bar{\rho}}(\bar{\rho}, X)$ of $\Upsilon_X(\bar{\rho})$ under $N(X\beta, \sigma^2\Sigma(\bar{\rho}))$ is the same as under $N(0, \Sigma(\bar{\rho}))$, by invariance under G_X. It immediately follows that, for any $X \neq 0$, any $\bar{\rho} > 0$, and any α, $\pi_{\bar{\rho}}(\bar{\rho}, X) \leq \pi_{\bar{\rho}}(\bar{\rho}, 0)$, because G_X is strictly larger than G_0 (as all transformations in G_0, i.e. $y \to ay$, are in G_X, and there are transformations in G_X, i.e., those with $b \neq 0$, that are not in G_0). Since, by the Neyman-Pearson Lemma applied to $pdf(v; \bar{\rho})$, $\Upsilon_0(\bar{\rho})$ is unique (up to a set of measure zero), a necessary and sufficient condition for $\pi_{\bar{\rho}}(\bar{\rho}, X) = \pi_{\bar{\rho}}(\bar{\rho}, 0)$, $X \neq 0$, is that $\Upsilon_X(\bar{\rho}) = \Upsilon_0(\bar{\rho})$, i.e., $y'C'[(C\Sigma(\bar{\rho})C')^{-1} - c_\alpha I]Cy < 0$ if and only if $y'[\Sigma^{-1}(\bar{\rho}) - c_\alpha I]y < 0$. Since $rank(C'RC) \leq n - k$ for any $(n-k) \times (n-k)$ matrix R, $\Upsilon_X(\bar{\rho}) = \Upsilon_0(\bar{\rho})$, $X \neq 0$, requires $rank(\Sigma^{-1}(\bar{\rho}) - c_\alpha I) \leq n - k$, and hence $c_\alpha = \lambda_i^{-1}(\Sigma(\bar{\rho}))$, $i = 2, ..., n-1$, which is equivalent to $\alpha = \Pr(y'\Sigma^{-1}(\bar{\rho})y/y'y < \lambda_i^{-1}(\Sigma(\bar{\rho}))$, $i = 2, ..., n-2$ (the cases $i = 1, n$ are excluded because α is assumed to be in $(0,1)$). It is easily seen that if $c_\alpha = \lambda_i^{-1}(\Sigma(\bar{\rho}))$,

$i = 2, ..., n-1$, $\Upsilon_0(\bar{\rho})$ is invariant under $y \to ay + Xb$, and hence is equal to $\Upsilon_X(\bar{\rho})$, if and only if $\text{col}(X) \subseteq E_i(\Sigma(\rho))$. This completes the proof of the proposition.

Proof of Lemma 2.3.2 Let $p = \lim_{\rho \to a} pdf(v;\rho) = |\Omega|^{-\frac{1}{2}} \left(v'\Omega^{-1}v\right)^{-\frac{n-k}{2}}$. If all the eigenvalues of $\Sigma(\rho)$ tend to a positive value as $\rho \to a$, then, by the Poincaré separation theorem (e.g., Rao, 1973, p. 64), all the eigenvalues of Ω are positive. It follows that the term $|\Omega|^{-\frac{1}{2}}$ of p is positive and finite if $\lambda_{n-k}(\Omega) < \infty$, and it vanishes otherwise. As for the term $(v'\Omega^{-1}v)^{-\frac{n-k}{2}}$, this is infinite if $\lambda_{n-k}(\Omega) = \infty$ and $v \in \tilde{E}_{n-k}(\Omega)$, positive and finite in any other case. Combining the results, we have that if $\lambda_{n-k}(\Omega) = \infty$, then $p = 0$ when $v \notin \tilde{E}_{n-k}(\Omega)$. Hence, by property (i) of $pdf(v;\rho)$, when $\lambda_{n-k}(\Omega)$ is infinite and simple, p must be infinite when $v \in \tilde{E}_{n-k}(\Omega)$. Also, we have that $0 < p < \infty$ if $\lambda_{n-k}(\Omega) < \infty$. The lemma now follows straightforwardly, on recalling that we are assuming that any invariant c.r. is centrally symmetric, so that it contains either both or neither of two antipodal points.

Proof of Theorem 2.3.3 Nonnegativity and irreducibility of W imply that $(I - \rho W)^{-1}$ is entrywise positive, for any $\rho > 0$ (see, e.g., Gantmacher 1974, p. 69, and recall that when we write $\rho > 0$ we implicitly assume $\rho < \lambda^{-1}$). It follows that, for both CAR and SAR models and for any $\rho > 0$, $\Sigma(\rho)$ is positive and hence, by Perron's theorem (e.g., Horn and Johnson, 1985, Theorem 8.2.11), that $\lambda_n[\Sigma(\rho)]$ is simple. Also, observe that, for both CAR and SAR models, as $\rho \to \lambda^{-1}$, $\lambda_n[\Sigma(\rho)] \to \infty$ and all of the other eigenvalues of $\Sigma(\rho)$ tend to a finite value, because, as it is easily verified, $rank[(I - \lambda^{-1}W')(I - \lambda^{-1}W)] = n - 1$. For CAR and symmetric SAR models and for any $\rho > 0$, $f_n[\Sigma(\rho)] = f$ and thus the spectral decomposition $\Sigma(\rho) = \sum_{i=1}^{n} \{\lambda_i[\Sigma(\rho)] f_i[\Sigma(\rho)] f_i'[\Sigma(\rho)]\}$ shows that the matrix $\lambda_n^{-1}[\Sigma(\rho)]\,\Omega_\rho$ tends to $Cff'C'$ as $\rho \to \lambda^{-1}$. The same limit result holds also for asymmetric SAR models, since in that case $f_n[\Sigma(\rho)] \to f$ as $\rho \to \lambda^{-1}$ (because $(I - \lambda^{-1}W')(I - \lambda^{-1}W)$ has an eigenvector f corresponding to its smallest eigenvalue 0). Now, since $rank(Cff'C') \leq rank(ff') = 1$, all eigenvalues of $Cff'C'$ are zero except possibly one, which must then be equal to $\bar{\lambda} = tr\,[Cff'C'] = f'Mf$. If $f \notin \text{col}(X)$, then $\bar{\lambda}$ is a simple positive eigenvalue of $Cff'C'$ and has an associated eigenvector equal to Cf, for

$$Cff'C'Cf = Cff'Mf = \bar{\lambda}Cf.$$

It is easily seen that for any CAR or SAR model with $f \notin \text{col}(X)$, Cf is also an eigenvector of $\lim_{\rho \to \lambda^{-1}} \Omega_\rho$, with (simple) eigenvalue equal to $\lim_{\rho \to \lambda^{-1}} \{\lambda_n[\Sigma(\rho)]\,\bar{\lambda}\} = \infty$. If $f \in \text{col}(X)$, then $Cf = 0$ and thus $\Omega_\rho = \sum_{i=1}^{n-1} \{\lambda_i^{-1}[\Sigma(\rho)] C f_i f_i' C'\}$, which tends to a matrix whose entries are all finite. Hence, when $f \in \text{col}(X)$, $\lim_{\rho \to \lambda^{-1}} \lambda_{n-k}(\Omega_\rho)$ must be finite. The theorem now follows by applying Lemma 2.3.2 with $a = \lambda^{-1}$.

Proof of Lemma 2.3.8 From (2.13), we have that, provided that $Cf \neq 0$, $\alpha^* = 0$ if and only if $Cf/\|Cf\| = \arg\min_{v \in S_{n-k}}\{v'Bv\}$, and $\alpha^* = 1$ if and only if $Cf/\|Cf\| = \arg\max_{v \in S_{n-k}}\{v'Bv\}$. The proposition follows by application of the Rayleigh-Ritz theorem (e.g., Horn and Johnson, 1985).

Proof of Proposition 2.3.9 It can be deduced from the proof of Theorem 2.3.3 that, for a CAR or SAR model with $f \notin \text{col}(X)$, $E_{n-k}(\Omega_\rho)$ tends, as $\rho \to \lambda^{-1}$, to a 1-dimensional subspace containing Cf. It follows that if $E_{n-k}(\Omega_\rho)$ does not depend on ρ for $\rho > 0$, it must be spanned by Cf for any $\rho > 0$, and hence, by Lemma 2.3.8 with $B = \Omega_{\bar{\rho}}^{-1}$, $\alpha^* = 0$ for any POI test. Since this property holds for any $\bar{\rho} > 0$, it also holds for the LBI test.

Proof of Theorem 2.3.11 We start from the case of the LBI test, which is notationally simpler than that of POI tests. By Lemma 2.3.8, for CAR and symmetric SAR models the limiting power of a LBI test vanishes for any α (less than 1) if and only if $f \notin \text{col}(X)$ and $Cf \in E_1(CWC')$. For a fixed W, consider the m_1-dimensional subspaces belonging to the span of $f_1, ..., f_{m_1}, f$, and denote by Θ the set of all such subspaces that do not contain f and are not E_1. It is easily shown that if $\text{col}(X) \in \Theta$, CWC' admits the eigenpairs (λ_i, Cf_i), $i = m_1 + 1, ..., n-1$. But then, by the symmetry of CWC' and the fact that the vectors Cf_i, $i = m_1 + 1, ..., n-1$ are pairwise orthogonal (because the f_i are), CWC' must also admit an eigenvector in the subspace spanned by $Cf_1, ..., Cf_{m_1}, Cf$. Since when $\text{col}(X) \in \Theta$ such a subspace is 1-dimensional, it follows that Cf is an eigenvector of CWC', i.e.,

$$CWMf = \tilde{\lambda}Cf \tag{2.15}$$

for some eigenvalue $\tilde{\lambda}$. Thus, a $\text{col}(X) \in \Theta$ causing the limiting power of the LBI test to disappear for any α exists if and only if $\tilde{\lambda} \leq \lambda_{m_1+1}$. Observe that as $\text{col}(X) \in \Theta$ approaches a subspace orthogonal to E_1, $Mf/\|Mf\|$ tends to a vector in E_1, which implies that $\tilde{\lambda} \to \lambda_1$ (note that, by the definition of Θ, no $\text{col}(X) \in \Theta$ is orthogonal to E_1). Thus, by the continuity of the eigenvalues of a matrix (CWC' here) in the entries of the matrix itself plus the fact that $\lambda_1 < \lambda_{m_1+1}$, a $\text{col}(X) \in \Theta$ such that $\tilde{\lambda} \leq \lambda_{m_1+1}$ always exists. The extension to POI tests, for any $\bar{\rho} > 0$, is straightforward and is obtained by replacing W with $\Sigma(\bar{\rho})$ and λ_i by $(1 - \rho\lambda_i)^{-r}$, $i = 1, ..., n$, in the arguments used above, for any CAR ($r = 1$) or symmetric SAR model ($r = 2$).

The second part of the theorem considers explicitly the case $m_1 = 1$. Take X to be a vector proportional to $f_1 + bf$, so that $\text{col}(X) \in \Theta$ as long as $b \neq 0$. For the LBI test, we just need to establish which values of b yield $\tilde{\lambda} \leq \lambda_2$ (existence of an infinite number

of such values of b follows from the first part of the theorem). Recalling that the vectors f_1 and f are normalized, it is easily seen that $Mf = (f - bf_1)/(1 + b^2)$. Plugging such an expression in (2.15), and using the fact that $Cf_1 = -bCf$ (since $CX = 0$), we obtain $\tilde{\lambda} = (\lambda + b^2\lambda_1)/(1 + b^2)$. Hence, $\tilde{\lambda} \leq \lambda_2$ requires $|b| \geq [(\lambda - \lambda_2)/(\lambda_2 - \lambda_1)]^{1/2}$, proving the part of the theorem relative to the LBI test when $m_1 = 1$ (note that the non-uniqueness of f_1 does not affect this result). By obvious extension, the limiting power of a POI test disappears for any α if $|b| \geq \{(\lambda_n[\Gamma(\bar{\rho})] - \lambda_2[\Gamma(\bar{\rho})])/(\lambda_2[\Gamma(\bar{\rho})] - \lambda_1[\Gamma(\bar{\rho})])\}^{1/2}$. The proof of the theorem is then completed on substituting $\lambda_i[\Gamma(\bar{\rho})] = (1 - \bar{\rho}\lambda_i)^{-p}$ in the last inequality, with $p = 1$ for a CAR model and $p = 2$ for a symmetric SAR model.

Proof of Proposition 2.3.12 By Theorem 2.3.11, for POI or LBI tests in CAR or symmetric SAR models with any W, there exist $\text{col}(X) \in G_{k,n}$ such that $\alpha^* = 1$. Let $T(y)$ represents the test statistic associated to a POI or LBI test. Then, by equation (2.13), the subspaces $\text{col}(X)$ yielding $\alpha^* = 1$ are those that maximize $T(f)$, regarded as a function from $G_{k,n}$ to $\mathbb{R}$. Next, observe that $T(f)$ is continuous at its points of maximum, which implies that, for any α, it is possible to find a neighborhood (defined according to some distance on $G_{k,n}$) of the points of maximum such that any $\text{col}(X)$ in this neighborhood causes the limiting power of size-α tests to disappear. This implies immediately that $H_k(\alpha)$ has non-zero invariant measure on $G_{k,n}$ (see James, 1954), for any $0 < \alpha < 1$ and for $k = m_1$, and for any POI or LBI c.r. in any CAR or symmetric SAR model. Since the power of an invariant test does not depend on β, the proposition also holds for $k > m_1$.

Proof of Corollary 2.3.17 The result follows immediately by taking $C = I$ in Theorem 2.3.3.

Proof of Proposition 2.3.19 Observe that if f is an eigenvector of W', it must be associated to λ. To see this, call ϕ the eigenvalue of W' associated to f. Transposing both left and right hand sides of the equation $W'f = \phi f$ and post-multiplying them by f yield $f'Wf = \phi$. But then $\phi = \lambda$, because it must also hold that $f'Wf = \lambda$. Let $\Gamma(\rho) = [(I - \rho W')(I - \rho W)]^{-1}$. By Lemma 2.3.8 with $B = \Gamma^{-1}(\bar{\rho})$, in order to prove the statement of the proposition regarding POI tests, we need to show that $W'f = \lambda f$ is necessary and sufficient for $f \in E_n(\Gamma(\bar{\rho}))$, for any $\bar{\rho} > 0$. Clearly, if this holds for any $\bar{\rho} > 0$, it also holds for $\bar{\rho} \to 0$, establishing the part of the proposition regarding the LBI test. The necessity is straightforward, because if $\Gamma(\bar{\rho})f = \lambda_n(\Gamma(\bar{\rho}))f$, then $\Gamma^{-1}(\bar{\rho})f = \lambda_n^{-1}(\Gamma(\bar{\rho}))f$. From the latter equation we have $(1 - \bar{\rho}\lambda)(I - \rho W')f = \lambda_n^{-1}(\Gamma(\bar{\rho}))f$, which requires f to be an eigenvector of $I - \rho W'$ and hence of W' (associated to λ by the above argument). As for the sufficiency,

note that if $W'f = \lambda f$, then f is clearly an eigenvector of $\Gamma(\bar{\rho})$, for any $\bar{\rho} > 0$. By Perron's theorem (e.g., Horn and Johnson, 1985, Theorem 8.2.11), a vector in $E_n(\Gamma(\bar{\rho}))$ is entrywise nonnegative (or nonpositive), for any $\bar{\rho} > 0$. But f is entrywise positive (by the Perron-Frobenius theorem applied to W), and hence it must be in $E_n(P_{\bar{\rho}}P'_{\bar{\rho}})$, for any $\bar{\rho} > 0$, because if it were not, then, by the symmetry of $P_{\bar{\rho}}P'_{\bar{\rho}}$, it should be orthogonal to an entrywise nonnegative vector, which is impossible. This completes the proof of the proposition.

Proof of Corollary 2.3.20 If W is a row-stochastic matrix, then f has identical entries, and therefore the condition in Proposition 2.3.19 is satisfied if and only if the columns of W, as its rows, sum to 1.

Proof of Proposition 2.4.1 The assertion follows by applying Corollary 3.2.1 of Lehmann and Romano (2005) to the density (2.6), plus the assumption of identification of model (2.1).

Proof of Proposition 2.4.2 For a POI test, we need to prove that $\pi_{\bar{\rho}}(\rho) \geq \alpha$ for any positive ρ and $\bar{\rho}$ and any size α. If unbiasedness holds for any $\bar{\rho} > 0$, then it also holds for the LBI test. Letting

$$M_{\bar{\rho}} = I - X[X'\Sigma^{-1}(\bar{\rho})X]^{-1}X'\Sigma^{-1}(\bar{\rho}),$$

the matrix $C'\left(C\Sigma(\bar{\rho})C'\right)^{-1}C$ can be rewritten as $\Sigma^{-1}(\bar{\rho})M_{\bar{\rho}}$ (e.g., Lemma 2 of King, 1980). Thus, for $0 \leq \rho < \lambda^{-1}$,

$$\pi_{\bar{\rho}}(\rho) = \Pr\left(\frac{y'\Sigma^{-1}(\bar{\rho})M_{\bar{\rho}}y}{y'My} < c_\alpha;\ y \sim N(0, \Sigma(\rho))\right). \tag{2.16}$$

Under Conditions A and B, $M_{\bar{\rho}} = M$ and, as is easily seen by exploiting the fact that Condition B is equivalent to the existence of an invertible matrix A such that $\Sigma(\widetilde{\rho})X = XA$, the matrices $\Sigma^{-1}(\bar{\rho})$ and M commute for any $\bar{\rho} > 0$. Hence,

$$\pi_{\bar{\rho}}(\rho) = \Pr\left(\frac{z'\Sigma(\rho)\Sigma^{-1}(\bar{\rho})Mz}{z'\Sigma(\rho)Mz} < c_\alpha\right),$$

where $z \sim N(0, I)$. Moreover, under Conditions A and B, the matrix M has an eigenvalue 0 with eigenspace spanned by the k eigenvectors of $\Sigma(\rho)$ that are in $\mathrm{col}(X)$, and an eigenvalue 1 with eigenspace spanned by the remaining eigenvectors of $\Sigma(\rho)$. Let H be the set of indexes i of the $n-k$ eigenvalues $\lambda_i[\Sigma(\rho)]$ associated to a set of linearly independent eigenvectors of $\Sigma(\rho)$ that are not in $\mathrm{col}(X)$. Note that, when Condition A holds, H does not depend on ρ. Under Conditions A and B, the power of a POI c.r. can then be expressed as

$$\pi_{\bar{\rho}}(\rho) = \Pr\left(\frac{\sum_{i\in H}\lambda_i[\Sigma(\rho)]\lambda_i^{-1}[\Sigma(\bar{\rho})]z_i^2}{\sum_{i\in H}\lambda_i[\Sigma(\rho)]z_i^2} < c_\alpha\right), \tag{2.17}$$

and its size as

$$\alpha = \Pr\left(\frac{z'\Sigma^{-1}(\bar{\rho})Mz}{z'Mz} < c_\alpha\right) = \Pr\left(\frac{\sum_{i\in H}\lambda_i^{-1}[\Sigma(\bar{\rho})]z_i^2}{\sum_{i\in H}z_i^2} < c_\alpha\right). \tag{2.18}$$

Observe now that the sequences $\lambda_i[\Sigma(\rho)]$, $i \in H$, and $\lambda_i^{-1}[\Sigma(\bar{\rho})]$, $i \in H$, are oppositely ordered in the sense of Hardy *et al.*, 1952, p. 43. Then, the application of Tchebychef's inequality (Hardy *et al.*, 1952, Theorem 43) to the weighted arithmetic means (with weights $z_i^2 / \sum_{i \in H} z_i^2$) of the $\lambda_i[\Sigma(\rho)]$, $i \in H$, and of the $\lambda_i^{-1}[\Sigma(\bar{\rho})]$, $i \in H$, yields that

$$\sum_{i \in H} \lambda_i[\Sigma(\rho)] z_i^2 \sum_{i \in H} \lambda_i^{-1}[\Sigma(\bar{\rho})] z_i^2 \geq \sum_{i \in H} z_i^2 \sum_{i \in H} \lambda_i[\Sigma(\rho)] \lambda_i^{-1}[\Sigma(\bar{\rho})] z_i^2,$$

for any vector $z \in \mathbb{R}^n$, with equality holding only if all the $\lambda_i[\Sigma(\rho)]$ or all the $\lambda_i^{-1}[\Sigma(\bar{\rho})]$, $i \in H$, are the same. Rearranging the terms of the above inequality, one finds that the statistic appearing in expression (2.17) is stochastically larger (e.g., Lehmann and Romano, 2005, p. 70) than that appearing in expression (2.18), and hence that $\pi_{\bar{\rho}}(\rho) \geq \alpha$, for any $\bar{\rho} > 0$, any $\rho > 0$ and any size α. If there are at least two indexes $i, j \in H$ such that $\lambda_i[\Sigma(\rho)] \neq \lambda_j[\Sigma(\rho)]$, i.e., if $\text{col}^{\perp}(X)$ is not a subset of an eigenspace of $\Sigma(\rho)$, then the last inequality is strict (as we are assuming $\alpha \neq 0, 1$). The proof of the proposition is completed.

Proof of Proposition 2.4.3 For a CAR or a symmetric SAR model, $\lambda_i[\Sigma(\rho)] = (1 - \rho\lambda_i)^{-r}$, for $i = 1, ..., n$, and with $r = 1$ for a CAR model, $r = 2$ for a symmetric SAR model. Inserting such expressions in equation (2.17) from the proof of Proposition 2.4.2, we obtain that the power function of a POI c.r. is non-decreasing in ρ if the statistic

$$t_{\bar{\rho}}(\rho) = \left(\sum_{i \in H} \left(\frac{1}{1 - \rho\lambda_i} \right)^r z_i^2 \right)^{-1} \sum_{i \in H} \left(\frac{1 - \bar{\rho}\lambda_i}{1 - \rho\lambda_i} \right)^r z_i^2$$

is non-increasing in ρ for any vector $z \in \mathbb{R}^n$. Direct differentiation of $t_{\bar{\rho}}(\rho)$ with respect to ρ and some simple manipulation show that such a condition is satisfied if

$$\sum_{i,j \in H} a_{i,j} z_i^2 z_j^2 \leq 0, \tag{2.19}$$

with the coefficients $a_{i,j}$ defined by

$$a_{i,j} = \lambda_j \frac{(1 - \bar{\rho}\lambda_j)^r - (1 - \bar{\rho}\lambda_i)^r}{(1 - \rho\lambda_i)^r (1 - \rho\lambda_j)^{r+1}}.$$

It is immediately verified that, for each $i, j \in H$ such that $i \neq j$, $a_{i,j} + a_{j,i} \leq 0$, with strict inequality if $\lambda_i \neq \lambda_j$. Thus, given that $a_{i,i} = 0$ for any $i \in H$, (2.19) holds, the inequality being strict if there exist at least one pair of distinct eigenvalues λ_i, λ_j with $i, j \in H$, i.e., if $\text{col}^{\perp}(X)$ is not a subset of an eigenspace of W. The statement in the proposition relative to the POI c.r.s is therefore proved, and the one relative to LBI follows immediately.

Proof of Proposition 2.4.6 Under Condition B, if $f_i \notin \text{col}(X)$, for $i = 1, ..., n$, then $f_i \in \text{col}^{\perp}(X)$. It follows that, if $f_i \notin \text{col}(X)$, for $i = 1, ..., n$ and when $\Sigma(\rho)$ is that of

a CAR or symmetric SAR model, $\Omega_\rho C f_i = C\Sigma(\rho) M f_i = C\Sigma(\rho) f_i = \lambda_i(\Sigma(\rho)) C f_i$, i.e., $\{Cf_i, f_i \notin \mathrm{col}(X), j = 1, ..., n\}$ is a set of $n-k$ orthogonal eigenvectors of Ω_ρ. Thus, in particular, $E_{n-k}(\Omega_\rho)$ does not depend on ρ. The proposition now follows by Theorem 2.3.3, and Propositions 2.3.9 and 2.4.3.

Chapter 3

Spatial Design Matrices and Associated Quadratic Forms: Structure and Properties

Abstract

The paper provides significant simplifications and extensions of results obtained by Gorsich, Genton, and Strang (2002, Eigenstructures of spatial design matrices, Journal of Multivariate Analysis 80, 138-165) on the structure of spatial design matrices. These are the matrices implicitly defined by quadratic forms that arise naturally in modelling intrinsically stationary and isotropic spatial processes. We give concise structural formulae for these matrices, and simple generating functions for them. The generating functions provide formulae for the cumulants of the quadratic forms of interest when the process is Gaussian, second-order stationary and isotropic. We use these to study the statistical properties of the associated quadratic forms, in particular those of the classical variogram estimator, under several assumptions about the actual variogram.

Keywords: cumulants; intrinsically stationary processes; Kronecker products; quadratic forms; spatial design matrices; variogram.

3.1 Introduction

In modelling spatial data—in general in d dimensions—observed at sites labelled by points in some subset of $\mathbb{R}^d$, it is often assumed that the process is intrinsically stationary and isotropic (see below and Cressie, 1993). Such models are then—intuitively at least—generalizations of familiar stationary time series models defined on the line (the case $d = 1$), and we shall see that there is quite a formal structure that reflects this relationship (Theorem 3.2.3 below).

In this paper, as in the recent paper by Gorsich, Genton, and Strang (2002) (hereafter abbreviated to GGS), we assume that the observational sites are located on a uniform grid in $\mathbb{R}^d$, with n sites on each of d axes. Sites may then be labelled by elements of the set $\Gamma = \Gamma(n, d)$ of sequences $\alpha = (\alpha(1), ..., \alpha(d))$ of non-negative integers satisfying $0 \leq \alpha(i) \leq (n-1)$ for $i = 1, ..., d$, and, to avoid ambiguity, we order the sequences in Γ *lexicographically.* Extensions to the case of a rectangular grid are straightforward, but for simplicity we confine our results to the hypercubic grid.

Denoting the observed process by $\{Z(\alpha); \alpha \in \Gamma\}$, intrinsic stationarity entails the assumptions that $E(Z(\alpha))$ is constant, and that, for $\alpha \neq \beta$, $\gamma(\alpha, \beta) = Var(Z(\alpha) - Z(\beta))$ depends on (α, β) only through $(\alpha - \beta)$, and the isotropy assumption that $\gamma(\alpha, \beta)$ depends on (α, β) only through $h = \|\alpha - \beta\|^2$, the squared Euclidean distance between the sites α and β. In that case the function $2\gamma(h)$ defined by

$$2\gamma(h) = Var(Z(\alpha) - Z(\beta)) \tag{3.1}$$

is called the *variogram* of the process $Z(\alpha)$. Note that, here and throughout, we use h to denote the *squared* Euclidean distance $\|\alpha - \beta\|^2 = \sum_{i=1}^{d} (\alpha(i) - \beta(i))^2$ between sites, rather than (as is more common) $\|\alpha - \beta\|$ itself. This is notationally more convenient later. Henceforth we take h to be strictly positive unless otherwise indicated.

The natural estimator for $2\gamma(h)$ is based on the function

$$q_h = \sum_{N(h)} (z(\alpha) - z(\beta))^2, \tag{3.2}$$

where $z(\alpha)$ denotes the observed value of $Z(\alpha)$, and $N(h)$ is the set of (unordered) pairs (α, β) satisfying $\|\alpha - \beta\|^2 = h$. Note that both $\gamma(0) = 0$ and $q_0 = 0$. Statistics of this form are also of interest more generally in the context of modelling spatial processes.

For $h > 0$ the expression on the right in (3.2) may be written as a quadratic form

$$q_h = z' L_h z = z'(D_h - A_h)z, \tag{3.3}$$

where $z = (z(\alpha); \alpha \in \Gamma)$ denotes the N-dimensional vector of observations, L_h and A_h are symmetric, and D_h is a diagonal matrix. Here and throughout $N = n^d = |\Gamma|$, the cardinality of Γ, denotes the total sample size. The matrix of this quadratic form, L_h, is the $N \times N$ *spatial design matrix* at distance $\sqrt{h}$, and D_h and $-A_h$ are, respectively, the diagonal and off-diagonal parts of L_h. By expanding the right side of (3.2) it is easy to see that A_h has a one in positions labelled by pairs (α, β) satisfying $\|\alpha - \beta\|^2 = h$, and zeros elsewhere, and that the diagonal element in row α of D_h is the number of sequences $\beta \in \Gamma$ satisfying $\|\alpha - \beta\|^2 = h$, i.e., the sum of the elements in row α of A_h. The matrices $L_h = L_h(n, d)$ in (3.3) are, in GGS, denoted by $A^{(d)}(n^d, h)$, with $h = \|\alpha - \beta\|$. The matrix A_h may be interpreted as the adjacency matrix of a graph $G(\Gamma, h)$ with vertex set Γ and edges the pairs $(\alpha, \beta) \in \Gamma \times \Gamma$ for which $\|\alpha - \beta\|^2 = h$. In that context L_h is known as the Laplacian matrix of the graph $G(\Gamma, h)$ (see, Mohar, 1997, for instance). Statistics of the type (3.2) have been studied extensively for the case $d = 1$, beginning with von Neumann *et al.* (1941).

As already mentioned, an important application of the quadratic forms q_h is to the estimation of the variogram in geostatistics. Let $N_h = |N(h)|$ denote the cardinality of the set $N(h)$. The statistic $2\hat{\gamma}(h) = q_h/N_h$, is an unbiased estimator of $2\gamma(h)$, and is often referred to as the classical variogram estimator (see Section 3.3.2 below, and GGS and the references therein). However, for other purposes it is also of interest to consider the statistics

$$q_h^* = 2 \sum_{N(h)} z(\alpha) z(\beta) = z' A_h z, \tag{3.4}$$

based on just the off-diagonal part of L_h. To give just a few examples: (i) the statistic q_h^*, normalized by $z'z$, is used to test for spatial autocorrelation at distance $\sqrt{h}$ (see Moran, 1950); (ii) if the covariance matrix of the process belongs to the linear span of (some of) the matrices A_h, that is, if the spatial process is not only intrinsically stationary and isotropic, but also second-order stationary, the statistic $q_h^*/(2N_h)$ is (when the process has zero mean) an unbiased estimator of the covariance function at distance $\sqrt{h}$ (see Section 3.3.2); (iii) if the process is assumed to be Gaussian with precision matrix (inverse covariance matrix) that is a linear combination of matrices I_N and $\{A_h,\ h \in H_p\}$, where H_p contains p distinct values of h and I_N denotes the $N \times N$ identity matrix, then a p-th order conditional autoregression is obtained (Besag, 1974). The matrices A_h, $h \in H_p$, play the role of spatial weights matrices, and the quadratic forms $(z'z, q_h^*,\ h \in H_p)$, are minimal sufficient statistics for the parameters of the model, and thus form the basis for inference on those parameters.

The problem of interest here is to give structural formulae for the matrices A_h, and

thereby for D_h and L_h. Thus, we continue the work of GGS, whose aim was to analyze the eigenstructure of the matrices L_h, with a view to deducing the properties of statistics like q_h and q_h^*, or more specifically of the variogram estimator $2\hat{\gamma}(h)$. It is well-known that under Gaussian assumptions (and also more generally) the properties of q_h and q_h^* depend upon L_h and A_h, respectively, only through their eigenvalues. Our purpose in the present paper will be to simplify and extend the results given in GGS.

In Section 3.2 we first provide a complete structural representation of the matrices A_h and L_h, and then give generating functions that make their computation straightforward with a standard symbolic computation package. In principle this completely solves the eigenvalue problem, but in practice, since N is usually quite large, direct computation of the eigenvalues would be unreliable. And, as we shall see, except in special cases, both A_h and L_h are sums of non-commuting matrices. Since, in this case, it is generally not possible to express the eigenvalues of the sum in terms of those of the summands, general explicit formulae for the eigenvalues are unlikely to be accessible.

Fortunately, our generating function results do permit the computation of the cumulants of the statistics of interest very simply and directly. In Section 3.3 we use these expressions to study the properties of the statistics q_h and q_h^* under the assumption that the process $\{Z(\alpha), \alpha \in \Gamma\}$ is Gaussian, second-order stationary, and isotropic. In particular, in Section 3.3.3 we show that the earlier results can be applied to the study of the properties of the classical variogram estimator $2\hat{\gamma}(h)$ under a variety of assumptions on the actual variogram $2\gamma(h)$.

3.2 The Matrices A_h, D_h and L_h

In this section we give the main structural results for the matrices A_h, D_h and L_h. The elements of these matrices, indexed by pairs $(\alpha, \beta) \in \Gamma \times \Gamma$, are completely determined by n, d and h. The results express these matrices in $d > 1$ dimensions in terms of sums of Kronecker products of the corresponding matrices in dimension $d = 1$. We begin with the key result—a very simple structural formula for the matrices A_h.

3.2.1 Off-Diagonal Part

The matrices A_h are defined by

$$(A_h)_{\alpha,\beta} = \begin{cases} 1 & \text{if } \| \alpha - \beta \|^2 = h; \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

Evidently, setting $A_0 = I_N$, $\Sigma_{h \geq 0} A_h = J_N$, where J_q is the $q \times q$ matrix with all elements one. In dimension $d = 1$ we denote the $n \times n$ matrices A_{r^2} by F_r, $r = 0, 1, ..., n-1$. That is:

$$(F_r)_{i,j} = \begin{cases} 1 & \text{if } |i - j| = r; \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

Since $\Sigma_{r=0}^{n-1} F_r = J_n$, we have that

$$J_N = \bigotimes_1^d J_n = \bigotimes_1^d \left(\sum_{r=0}^{n-1} F_r \right) = \sum_{\alpha \in \Gamma} (F_{\alpha(1)} \otimes F_{\alpha(2)} \otimes ... \otimes F_{\alpha(d)}) \tag{3.7}$$

by the multilinearity of the Kronecker (or direct) product '$\otimes$'. Note that the elements of

$$F_\alpha^{\otimes} = F_{\alpha(1)} \otimes F_{\alpha(2)} \otimes ... \otimes F_{\alpha(d)} \tag{3.8}$$

are zeros and ones, so *exactly one term* $F_\alpha^{\otimes}$ on the right in (3.7) has a one in any given position (β, δ). In view of (3.7), the following result, proved in the Appendix, is not surprising:

Proposition 3.2.1 *Let* $\Gamma_h = \{\alpha \in \Gamma : \|\alpha\|^2 = h\}$. *Then:*

$$A_h = \sum_{\alpha \in \Gamma_h} F_\alpha^{\otimes}. \tag{3.9}$$

For example, if $h = 1$, Γ_1 consists of d sequences containing a single one and $d - 1$ zeros, so that

$$A_1 = \sum_{i=1}^{d} (I_n \otimes ... \otimes F_1 \otimes ... \otimes I_n),$$

with F_1 in the i-th position in the i-th term (see the discussion of equation (9) in GGS). Likewise, for $h = 2$, Γ_2 consists of the $\binom{d}{2}$ sequences that contain 2 ones and $d - 2$ zeros, so in the corresponding expression for A_2 each term in the sum contains F_1 twice. Notice that, in both of these low-order cases, all the sequences that appear in Γ_h are permutations of a single sequence.

An alternative proof of Proposition 3.2.1 based on known graph-theoretic results is worth recording, because it shows immediately how to generalize the result to cover index sets more complex than the uniform grid Γ, e.g., the rectangular grid mentioned in the Introduction. We refer the reader to Cvetković *et al.* (1980) for more on the graph-theoretic details.

Given graphs $G_i(V_i, E_i)$, $i = 1, ..., d$, with vertex sets V_i and edge sets E_i, the *direct product* of the G_i, $G_1 \times ... \times G_d$ is the graph $G_d^{\times}$, say, defined as follows. The vertex set of $G_d^{\times}$ is the Cartesian product $V_d^{\times} = V_1 \times ... \times V_d$ of the V_i, and if $x_i, y_i \in V_i$ for $i = 1, ..., d$, $(x_1, ..., x_d)$ and $(y_1, ..., y_d)$ are adjacent in $G_d^{\times}$ if and only if $(x_i, y_i) \in E_i$ for $i = 1, ..., d$. In our case, the matrices F_r, $r = 0, ..., n-1$, are the adjacency matrices of the (so-called distance) graphs G_r with common vertex sets $V_r = V = \{0, ..., n-1\}$, and with edge sets defined by: for $i, j \in \{0, ..., n-1\}$, $(i, j) \in E_r$ only when $|i - j| = r$. Then, $V_d^{\times} = \Gamma$, and for each $\alpha \in \Gamma$ we may define a product $G_d^{\times}(\alpha)$ of the graphs $G_{\alpha(i)}$ as above. It is known that $G_d^{\times}(\alpha)$ has adjacency matrix $F_{\alpha}^{\otimes}$ (Cvetković *et al.*, 1980, Theorem 2.21). Thus, for any subset U of Γ, the union of the graphs $G_d^{\times}(\alpha)$ has adjacency matrix $A_U = \sum_{\alpha \in U} F_{\alpha}^{\otimes}$. Proposition 3.2.1 gives the case $U = \Gamma_h$.

Call two sequences (β, δ) $h-neighbors$ if the sequence α defined by $\alpha(i) = |\beta(i) - \delta(i)|$, for $i = 1, ..., d$, is in Γ_h. This definition of neighbors—based on the Euclidean distance between points—is natural in some contexts, but in others a neighborhood structure based, say, on the L_1-norm (the length of the shortest walk connecting β to δ) may be more appropriate. The observation in the previous paragraph makes it straightforward to extend the results to follow to this case (and to neighborhood structures defined by other L_p-norms), but we omit the details.

3.2.2 Diagonal Part

The matrices D_h in (3.3) are diagonal matrices with diagonal elements $D_h(\alpha)$ equal to the number of $h - neighbors$ of α. In dimension $d = 1$ define, for each $r = 0, ..., n-1$, the diagonal matrix M_r with $i - th$ diagonal element the $i - th$ row sum of F_r, and then define, for $\alpha \in \Gamma$,

$$M_{\alpha}^{\otimes} = M_{\alpha(1)} \otimes M_{\alpha(2)} \otimes ... \otimes M_{\alpha(d)}. \tag{3.10}$$

It is straightforward to prove:

Proposition 3.2.2 $D_h = \sum_{\alpha \in \Gamma_h} M_{\alpha}^{\otimes}$.

Notice that $tr[D_h]$ is the total number of non-zero elements in A_h, so that $tr[D_h] = 2N_h$. We have now established:

Theorem 3.2.3 *The spatial design matrix at distance* $\sqrt{h}$ *is given by:*

$$L_h = \sum_{\alpha \in \Gamma_h} (M_{\alpha}^{\otimes} - F_{\alpha}^{\otimes}), \tag{3.11}$$

where $M_\alpha^\otimes$ and $F_\alpha^\otimes$ are as defined in (3.10) and (3.8).

The above expressions for the matrices A_h, D_h, and L_h involve summing over the set Γ_h. We next examine this set more closely, and give formulae for these matrices that do not involve Γ_h.

3.2.3 Generating Functions

Since h must be a sum of squares of d of the integers $(0, 1, ..., n-1)$, not all values of $h \leq d(n-1)^2$ are feasible. This is so even when $d \geq 4$, notwithstanding Lagrange's four-square theorem (Hardy and Wright, 1979, Sec. 20.5), because no term in the decomposition of h can exceed $(n-1)^2$. Thus, Γ_h in Proposition 3.2.1 can be empty, and in that case we define A_h, D_h and L_h to be zero matrices.

The values of h that yield non-vanishing matrices L_h can be read off from the expansion of the polynomial

$$(1 + t + t^4 + ... + t^{r^2} + ... + t^{(n-1)^2})^d = \sum_{h=0}^{d(n-1)^2} m_h t^h, \tag{3.12}$$

in which the coefficient m_h is evidently the number of ways in which h can be expressed as a sum of squares of d of the integers $(0, 1, ..., n-1)$, i.e., $m_h = |\Gamma_h|$ is the number of $h-$neighbors of the origin. Except for the restriction $h \leq d(n-1)^2$, the m_h evidently depend on d but not directly on n. Letting $f_n(t) = \sum_{r=0}^{n-1} t^{r^2}$, and using Wilf's (1994) notation, we may write

$$m_h = [t^h](f_n(t))^d, \tag{3.13}$$

where $[t^h]$ means "the coefficient of t^h in the expansion of the following function in powers of t". Note that $[t^h]$ is identical to the operator $(h!)^{-1}(\partial/\partial t)^h|_{t=0}$, and, as an operator, is therefore linear. A cumbersome formula for the m_h can be deduced from (3.13), but using a modern symbolic computing package it is a simple matter to compute m_h from (3.13) without having to rely on such formulae.

Similarly, letting $b_n(t) = \sum_{r=0}^{n-1} t^{r^2} x_r$, where the x_i are labels for the integers $0, 1, ..., n-1$, obeying the usual rules of multiplication, we see that, from the formal expansion of $(b_n(t))^d$,

$$[t^h](b_n(t))^d = \sum_{\alpha \in \Gamma_h} \left\{ \prod_{i=1}^{d} x_{\alpha(i)} \right\}. \tag{3.14}$$

Thus, the sequence α belongs to Γ_h only if the product $\Pi_{i=1}^d x_{\alpha(i)}$ appears on the right in (3.14).

The key to obtaining a simple representation for the matrices A_h, D_h, and hence L_h, is to notice that the scalar generating function $(b_n(t))^d$ can be generalized in such a way that, when expanded, the coefficient of t^h is precisely A_h. To see this, define the matrix

$$B_n(t) = \sum_{r=0}^{n-1} t^{r^2} F_r, \tag{3.15}$$

an $n \times n$ Toeplitz matrix with (i,j) element $t^{(i-j)^2}$. By direct expansion of the d-th Kronecker power $B_n^{\otimes}(t) = \bigotimes_1^d B_n(t)$, it is clear that A_h is the coefficient of t^h in the expansion of $B_n^{\otimes}(t)$ in powers of t. That is,

$$A_h = [t^h] B_n^{\otimes}(t). \tag{3.16}$$

Similarly, letting

$$C_n(t) = \sum_{r=0}^{n-1} t^{r^2} M_r \tag{3.17}$$

and $C_n^{\otimes}(t) = \bigotimes_1^d C_n(t)$, we see that

$$D_h = [t^h] C_n^{\otimes}(t). \tag{3.18}$$

We therefore have the simple generating-function representation for L_h given in:

Theorem 3.2.4 *The spatial design matrix at distance $\sqrt{h}$ is given by:*

$$L_h = [t^h](C_n^{\otimes}(t) - B_n^{\otimes}(t)). \tag{3.19}$$

These results evidently do not require knowledge of Γ_h: it is built in to the generating function. On the other hand, the matrices appearing in these representations of A_h, D_h and L_h are $N \times N$, and likely to be high-dimensional, so it might seem that these results would be of little practical value. On the contrary, we will see in the next section that they provide both analytically and computationally convenient information about the statistics q_h and q_h^* discussed in the Introduction, and hence about the properties of the variogram estimator $2\hat{\gamma}(h)$. Before doing so we note some further implications of these results.

It is clear that, if $\alpha \in \Gamma_h$, so is every permutation of the elements of α. Thus, Γ_h must be a union of one or more orbits in Γ under the action of the symmetric group S_d (the group of permutations of d objects). A set of orbit representatives is provided by the set $\Omega = \Omega(d,n)$ of *non-decreasing* sequences $\omega = (\omega(1), ..., \omega(d)) \in \Gamma$, with $\omega(1) \leq \omega(2) \leq ... \leq \omega(d)$. Let $\Omega_h = \{\omega \in \Omega : \|\omega\|^2 = h\}$, and, for $j = 0, ..., n-1$, $\omega \in \Omega$, let $k_\omega(j)$ denote the multiplicity of j in ω, so that $\sum_{j=0}^{n-1} k_\omega(j) = d$, and write $\nu(\omega) = \prod_{j=0}^{n-1} k_\omega(j)!$, with, as usual, $0! \equiv 1$.

With this notation it is easy to see that $m_h = d! \sum_{\omega \in \Omega_h} (\nu(\omega))^{-1}$, and since $\Gamma_h = \{\sigma\omega : \omega \in \Omega_h, \sigma \in S_d\}$, where $\sigma\omega$ denotes the permutation σ of ω, we have that

$$A_h = \sum_{\omega \in \Omega_h} \frac{1}{\nu(\omega)} F^*_\omega,$$

where $F^*_\omega = \sum_{\sigma \in S_d} F^{\otimes}_{\sigma\omega}$ is a symmetric function of the matrices $F_{\omega(1)}, ..., F_{\omega(d)}$. By an obvious extension of this argument to the off-diagonal part, and setting $M^*_\omega = \sum_{\sigma \in S_d} M^{\otimes}_{\sigma\omega}$, we can state:

Theorem 3.2.5 *The spatial design matrix at distance $\sqrt{h}$ is given by*

$$L_h = \sum_{\omega \in \Omega_h} \frac{1}{\nu(\omega)} (M^*_\omega - F^*_\omega). \tag{3.20}$$

For many values of h equation (3.14) reveals that Γ_h consists of a single orbit, which is to say that Ω_h has a single element, say ω_h. In that case $m_h = d!/\nu(\omega_h)$, and Theorem 3.2.5 gives the very simple result that $L_h = (\nu(\omega_h))^{-1}(M^*_{\omega_h} - F^*_{\omega_h})$. In the example following Proposition 3.2.1, for instance, $h = 1$, $\omega_1 = (0, .., 0, 1)$ and $\nu(\omega_1) = (d-1)!$.

Using these results we may also obtain the following generalization and simplification of Lemma 6.1 and Theorem 6.1 in GGS, which give upper bounds on the largest eigenvalues of L_h and A_h (for sets Ω_h with low cardinality), and hence upper bounds for the normalized statistics $z'L_hz/z'z$ and $z'A_hz/z'z$.

Lemma 3.2.6 *Let λ_h and μ_h denote the largest eigenvalues of A_h and L_h, respectively, and let $u_h = d! \sum_{\omega \in \Omega_h} \frac{2^{d-k_\omega(0)}}{\nu(\omega)}$. Then $\lambda_h \leq u_h$ and $\mu_h \leq 2u_h$.*

The proof of Lemma 3.2.6 is in the Appendix. If Ω_h contains only the single sequence ω_h, which contains only one non-zero term (so Γ_h contains only what GGS call "non-diagonal directions"), the matrices in the sum $\sum_{\alpha \in \Gamma_h} F^{\otimes}_\alpha$ *are* pairwise commutative, so the eigenvalues of A_h are simple functions of those of the single matrix F_r ($r = \sqrt{h}$) involved. Under the same condition, $L_h = ((d-1)!)^{-1} L^*_{\omega_h}$, with $L^*_{\omega_h} = \sum_{\sigma \in S_d} L^{\otimes}_{\sigma\omega_h}$, which is also a sum of pairwise commutative matrices. Thus, as GGS note in Lemma 5.1, in the case of non-diagonal directions the eigenvalues of L_h are simple functions of those of the matrix $(M_{\sqrt{h}} - F_{\sqrt{h}})$.

The necessary and sufficient conditions required to ensure pairwise commutativity of the summands in Theorem 3.2.5 are that Ω_h contains only the single sequence ω_h, and ω_h contains no more than one (possibly repeated) non-zero integer. Note that ω_h may correspond to what GGS would call "diagonal directions", and that these conditions are always satisfied for $h = 1, 2, 3$ (for any $d \geq h$), but otherwise clearly hold only for special values of h.

3.3 Applications

In this section we use the results established above to study the properties of the statistics $q_h^* = z'A_h z$ and $q_h = z'L_h z$. We consider first the case in which $z \sim N(0, I_N)$, but in Section 3.3.2 show how our earlier results can be used to deal with the more general case $z \sim N(0, \Sigma)$, assuming the process is second-order stationary and isotropic.

3.3.1 Properties of the Quadratic Forms q_h^* and q_h

Under the assumption $z \sim N(0, I)$, the distributions of the quadratic form $q = z'Az$, and its normalized form $\bar{q} = z'Az/z'z$, can certainly be obtained (see James, 1964, for the former, and Hillier, 2001 for the latter), but both are sufficiently complicated as to inhibit their use for practical study of, and/or tabulation of, the distribution. On the other hand, it is well known that the cumulants of $q = z'Az$ under the assumption $z \sim N(0, \Sigma)$ are given by:

$$\kappa_p = 2^{p-1}(p-1)!tr[(A\Sigma)^p],\ p = 1, 2, ... \tag{3.21}$$

(see Kendall and Stuart, 1969, Chapter 3 for the definition of cumulants, and Chapter 15 for the result given in equation (3.21)). The results in Section 3.2 allow these cumulants to be computed quite straightforwardly when $\Sigma = I_N$ and the matrix A in (3.21) is either A_h or L_h. These results are given next. First, for comparison, we summarize the properties of the analogue of q_h^* for the case $d = 1$.

In the case $d = 1$ the properties of the statistics $Q_r^* = y'F_r y,\ r = 1, ..., n-1$, with $y \sim N(0, I_n)$, have been extensively studied. The following Lemma summarizes some elementary properties of the statistics Q_r^*, all of which are either given in, or are easily deduced from, the comprehensive results in Anderson (1971):

Lemma 3.3.1 *For $r = 1, ..., n-1$, let $Q_r^* = y'F_r y$, and assume that $y \sim N(0, I_n)$. Then,*

$$E(Q_r^*) = tr[F_r] = 0;$$
$$var(Q_r^*) = 2tr[F_r^2] = 2tr[M_r] = 4(n-r).$$

Moreover, all odd cumulants of Q_r^* vanish, so the density of Q_r^* is symmetric about zero, and for $r_1 \neq r_2$, $Q_{r_1}^*$ and $Q_{r_2}^*$ are uncorrelated.

Properties of the q_h^*

With $A = A_h$ and $\Sigma = I_N$ in (3.21) we obtain the cumulants, $\kappa_{p,h}^*$, of q_h^*. Much of Lemma 3.3.1 generalizes easily to this case:

Lemma 3.3.2 *For $h \geq 1$, any $d \geq 1$, and $z \sim N(0, I_N)$,*

$$E(q_h^*) = tr[A_h] = 0,$$
$$var(q_h^*) = 2tr[A_h^2] = 2tr[D_h],$$

and, for h_1, $h_2 \geq 1$, $h_1 \neq h_2$, $q_{h_1}^*$ and $q_{h_2}^*$ are uncorrelated.

Lemma 3.3.2 is proved in the Appendix. Now, with the help of the generating function $C_n^{\otimes}(t)$ for D_h, it is straightforward to obtain a generating function for the variances $var(q_h^*)$, since

$$\begin{aligned} var(q_h^*) &= 2tr[D_h] = 2tr\left\{[t^h]C_n^{\otimes}(t)\right\} \text{ (using (3.18))} \\ &= 2[t^h]tr\left\{C_n^{\otimes}(t)\right\} \\ &= 2[t^h](tr(C_n(t))^d. \end{aligned} \tag{3.22}$$

The last step here follows from a standard property of the trace operator for Kronecker products, and the penultimate step from the fact that the operator $[t^h]$ commutes with the trace operator. Noting that $tr[M_0] = n$, and $tr[M_r] = 2(n-r)$, $r = 1, ..., n-1$, it follows from the definition of $C_n(t)$ that

$$tr(C_n(t)) = (n + 2(n-1)t + ... + 2(n-r)t^{r^2} + ... + 2t^{(n-1)^2}). \tag{3.23}$$

Since $2N_h = tr[D_h]$, these formulae provide simple and efficient methods for computing the values N_h: setting $g_n(t) = tr(C_n(t))$ we have

$$2N_h = [t^h](g_n(t))^d. \tag{3.24}$$

In general, for $d > 1$, the density of q_h^* is not symmetric about zero. The analogue of the symmetry result for the case $d = 1$ in Lemma 3.3.1 is the weaker result given in:

Lemma 3.3.3 *If ph is odd $tr[A_h^p] = 0$ (independently of d). Hence, for h odd, the distribution of q_h^* (and also its normalized form $\bar{q}_h^* = q_h^*/z'z$) is symmetric about zero.*

The next result is also of interest and, as Lemma 3.3.3, is proved in the Appendix.

Lemma 3.3.4 *For $d = 2$ and every $h \geq 1$, the distribution of q_h^* (and also its normalized form $\bar{q}_h^* = q_h^*/z'z$) is symmetric about zero.*

Lemma 3.3.4 generalizes Lemma 5 of Hillier and Martellosio (2006), which asserts that, for $d = 2$ and every $h \geq 1$, $tr[A_h^3]$, and hence $\kappa_{3,h}^*$, is zero. Note that Lemma 3.3.4 cannot be

extended to $d > 2$: for instance, when $d = 3$ and $h = 2$, $tr[A_h^3]$, and hence $\kappa_{3,h}^*$, is positive, since the sequences $(0,0,0)$, $(1,1,0)$, $(0,1,1)$ are the vertices of an equilateral triangle of side $h = 2$ (an odd-cycle of step 2 in the terminology of the proof of Lemma 3.3.4). Summarizing the results of Lemmas 3.3.3 and 3.3.4, the somewhat unexpected result has been found that the distribution of q_h^* is generally not symmetric about zero when h is even and $d > 2$, although it is symmetric for any $h \geq 1$ when $d = 1$ or 2 .

Properties of the q_h

We now deal with the case $A = L_h$ and $\Sigma = I_N$ in (3.21). Since $L_h l_N = 0$ (where l_N is an $N \times 1$ vector of ones), the results to follow continue to hold under the assumption that $z \sim N(\mu l_N, I_N)$, i.e., that the $Z(\alpha)$ have an unknown constant mean μ. We have, in either case, for the cumulants of q_h, $\kappa_{p,h} = 2^{p-1}\Gamma(p)tr[L_h^p]$, $p = 1, 2, ...$ Thus:

Lemma 3.3.5 *When $z \sim N(\mu l_N, I_N)$,*

$$E(q_h) = tr[L_h] = tr[D_h] = 2N_h, \tag{3.25}$$

and

$$var(q_h) = 2tr[L_h^2] = 2\left(tr[D_h^2] + tr[D_h]\right). \tag{3.26}$$

The result for the variance uses the facts that $tr[D_h A_h] = 0$ and $tr[A_h^2] = tr[D_h]$. The computation of $tr[D_h]$ has been discussed above, and we can compute the term $tr[D_h^2]$ from the formula:

$$tr[D_h^2] = tr\left[[t^h][s^h]C_n^{\otimes}(t)C_n^{\otimes}(s)\right] = [(ts)^h](tr\,[C_n(t)C_n(s)])^d.$$

Thus:

$$var(q_h) = 2\left\{[(ts)^h](tr\,[C_n(t)C_n(s)])^d + 2N_h\right\}. \tag{3.27}$$

From the definition of $C_n(t)$, $tr[C_n(t)C_n(s)] = \sum_{r_1,r_2=0}^{n-1} t^{r_1^2} s^{r_2^2} tr[M_{r_1}M_{r_2}]$, and it is easy to check that $tr[M_0^2] = n$ and, for $1 \leq r_1 \leq r_2 \leq n-1$,

$$tr[M_{r_1}M_{r_2}] = \begin{cases} 4(n-r_2) - 2r_1 & \text{if } r_1 + r_2 \leq n; \\ 2(n-r_2) & \text{otherwise.} \end{cases} \tag{3.28}$$

Thus, we again have a simple generating function for the variances of the statistics q_h, and hence for the variance of the variogram estimator in the "null" case ($\Sigma = I_N$) (see Section 3.3.3 below).

Higher-order cumulants and product cumulants (e.g., covariances) for both the q_h^* and the q_h can be obtained by obvious extensions of these methods. For instance,

$$tr[A_h^p] = [(t_1...t_p)^h](tr[B_n(t_1)...B_n(t_p)])^d, \tag{3.29}$$

and

$$cov(q_{h_1}, q_{h_2}) = 2tr[L_{h_1}L_{h_2}] = 2tr[D_{h_1}D_{h_2}] = 2[t^{h_1}][s^{h_2}](tr[C_n(t)C_n(s)])^d. \tag{3.30}$$

The generating functions in these expressions may, of course, simplify (as above), and this reduces the computational problem considerably. We leave other such extensions to the reader.

3.3.2 Second-Order Stationary Isotropic Processes

Under the assumption that the process is second-order stationary and isotropic—which is stronger than the intrinsic stationarity assumption mentioned in the introduction (see Cressie, 1993)—we have, as an obvious consequence of equation (3.16):

Proposition 3.3.6 *If the process* $\{Z(\alpha); \alpha \in \Gamma\}$ *is second-order stationary and isotropic, its covariance matrix* Σ *has the representation*

$$\Sigma = \sum_{h \in H} c(h) A_h, \tag{3.31}$$

where H *is a some set of values of* h *containing zero (recall that* $A_0 = I_N$*), and the coefficients* $\{c(h); h \in H\}$ *must be such that* Σ *is positive definite. Thus, from (3.16),* $\Sigma = [S_H(t)]B_n^{\otimes}(t)$*, where*

$$[S_H(t)] = \sum_{h \in H} c(h)[t^h]. \tag{3.32}$$

The operator $[S_H(t)]$ constructs a linear combination, with parameters $c(h)$, of the coefficients of the powers t^h, $h \in H$, that occur in the expansion of the function to which it is applied. Like the $[t^h]$ themselves, $[S_H(t)]$ is clearly linear. If we now assume that $z \sim N(0, \Sigma)$, with Σ as in (3.31), and take $h > 0$, we easily see that:

$$E(q_h^*) = tr[A_h\Sigma] = \sum_{k \in H} c(k) tr[A_h A_k] = \begin{cases} c(h)tr[D_h] & \text{if } h \in H; \\ 0 & \text{otherwise.} \end{cases} \tag{3.33}$$

And (since $tr[A_h] = tr[D_k A_h] = 0$),

$$E(q_h) = tr[L_h\Sigma] = \sigma^2 tr[D_h] - \sum_{k\in H\setminus\{0\}} c(k)tr[A_h A_k]$$
$$= \begin{cases} \{\sigma^2 - c(h)\}tr[D_h] & \text{if } h \in H; \\ \sigma^2 tr[D_h] & \text{otherwise,} \end{cases} \tag{3.34}$$

where we have put $c(0) = \sigma^2$. Since, under these assumptions, $\gamma(h) = \sigma^2 - c(h)$, this shows that $2\hat{\gamma}(h) = q_h/N_h$ is an unbiased estimator of the true variogram $2\gamma(h)$, for all $h > 0$, as is well-known (Cressie, 1993). Obviously, to compute the unbiased estimator $2\hat{\gamma}(h)$ one needs to know the correct scale factor N_h, and this has hitherto been unavailable for the isotropic case in general; equation (3.24) gives a simple general procedure for computing it, generalizing the special case given in Lemma 7.1 in GGS.

The variances and covariances of the statistics q_h^* and q_h for several values of h are often needed in applications. For instance, the entire covariance matrix of a vector of statistics q_h at a set of values of h is required for variogram fitting by generalized least squares (see Genton, 1988, and Cressie, 1993, Sec. 2.6.2), and this has previously been unavailable for the isotropic case. The covariances cannot easily be written down in closed form, but when Σ has the form (3.31) are easily represented in generating function form using the operators $[S_H(t)]$ defined in (3.32). Thus we easily obtain:

Lemma 3.3.7 *Suppose $z \sim N(0, \Sigma)$, with Σ of the form (3.31). Then, for any $h_1 \geq h_2$:*

$$cov(q_{h_1}^*, q_{h_2}^*) = 2tr[A_{h_1}\Sigma A_{h_2}\Sigma] = 2[s_1^{h_1}][s_2^{h_2}][S_H(t_1)][S_H(t_2)]v_n^d(s_1, s_2, t_1, t_2), \tag{3.35}$$

and

$$cov(q_{h_1}, q_{h_2}) = 2tr[L_{h_1}\Sigma L_{h_2}\Sigma] = 2[s_1^{h_1}][s_2^{h_2}][S_H(t_1)][S_H(t_2)]V_n^d(s_1, s_2, t_1, t_2), \tag{3.36}$$

where

$$v_n^d(s_1, s_2, t_1, t_2) = (tr[B_n(s_1)B_n(t_1)B_n(s_2)B_n(t_2)])^d \tag{3.37}$$

and

$$V_n^d(s_1, s_2, t_1, t_2) = v_n^d(s_1, s_2, t_1, t_2) + (tr[C_n(s_1)B_n(t_1)C_n(s_2)B_n(t_2)])^d$$
$$-2(tr[C_n(s_2)B_n(t_2)B_n(s_1)B_n(t_1)])^d. \tag{3.38}$$

Note that $cov(q_{h_1}^*, q_{h_2}^*) = 0$ when $h_1 \neq h_2$ and $h_1, h_2 \notin H$, and that the elements of the matrix defining $v_n^d(s_1, s_2, t_1, t_2)$ are positive. Thus, if the $c(h)$ in (3.31) are positive and non-decreasing in $|H|$, an increase in $|H|$ must increase $cov(q_{h_1}^*, q_{h_2}^*)$. Extensions to higher-order

cumulants are obvious, but, as in the case $\Sigma = I_N$, will entail a larger computational burden. Finally, we note that the approach used here can also be extended to the case where the precision matrix Σ^{-1}, rather than Σ itself, is a linear combination of the A_h.

3.3.3 Properties of the Classical Variogram Estimator

The above results for q_h provide the tools for studying the properties of the classical variogram estimator for a second-order stationary and isotropic process under virtually any specification for the $c(h)$. We do not intend to study the detailed properties of the variogram estimator here, but will show that the above results can be used to study the properties of $2\hat{\gamma}(h)$ under a variety of specifications for the variogram $2\gamma(h)$ (for the intrinsically stationary, but non-isotropic case, see Cressie, 1985).

We first consider the variance of $2\hat{\gamma}(h) = q_h/N_h$ as a function of h and d, assuming $\Sigma = I_N$. In Figure 3.1 we plot $var(2\hat{\gamma}(h)) = var(q_h)/N_h^2$, computed using equations (3.24) and (3.27), for $d = 1, 2, 3, 4$, and $h = 1, ..., 16$, with N held fixed at $N = 2^{12}$, so that, for $d = 1, 2, 3, 4$ we have $n = 2^{12}, 2^6, 2^4, 2^3$ respectively.

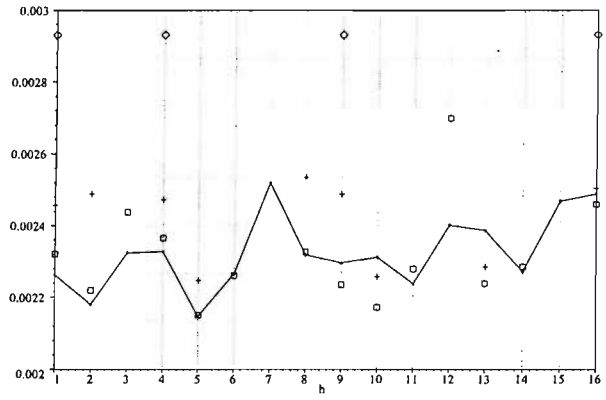
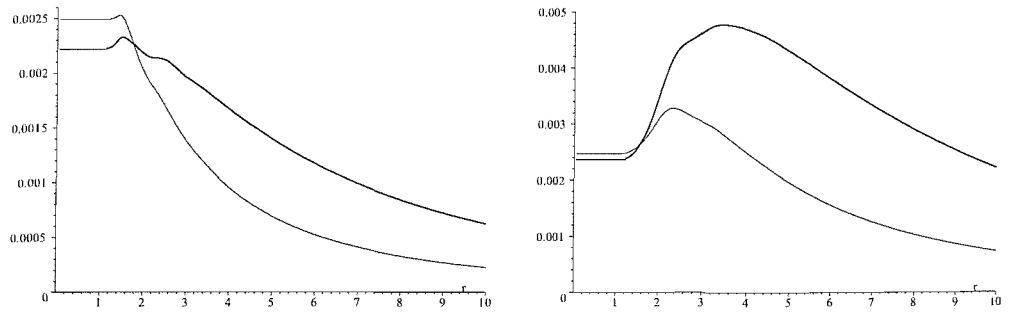


Figure 3.1: The variance of the classical estimator $2\hat{\gamma}(h)$ as a function of h and d : $d = 1$ (*diamond*), 2 (*cross*), 3 (*square*), 4 (*line*); $N = 2^{12}$; $\Sigma = I_N$.

Figure 3.1 shows that: (a) for each fixed dimension $d > 1$, the variance is quite volatile as h varies; and (b) the variance is not monotonic in d for fixed h (see for instance the value $h = 9$). Thus, in contrast to Figure 4 in GGS (where the variance could only be computed for "non-diagonal" directions), our results show that when "diagonal" directions are taken into account—as it is natural to do under the assumption of isotropy—$var(2\hat{\gamma}(h))$ is no longer monotonic either in d or in h. The volatility and non-monotonicity of the variances is attributable to variation in N_h, m_h, and the structure of Ω_h as h varies. The explanation is



(a) (b)

Figure 3.2: The variance of the classical estimator $2\hat{\gamma}(h)$ when the variogram is spherical. In (a) $h = 2$, in (b) $h = 4$. The variance is plotted for many values of the range r from 0 to 10, $N = 2^{12}$, $d = 2$ (thin line), $d = 3$ (thick line).

purely number theoretic: the number of decompositions of a particular h as a sum of squares is not related in any simple way to the values n and d.

The variance of the classical variogram estimator when Σ is of the form (3.31) can be computed using (3.36) with $h_1 = h_2$. Using this formula, one can study the behavior of $var\,(2\hat{\gamma}\,(h))$ under various specifications for the true variogram $2\gamma(h)$, i.e., of the $c(h)$ in (3.31). In Figure 3.2 we plot the variances for the case of a spherical variogram with sill 1, nugget 0 and range r, so that the $c(h)$ in (3.31) are given by

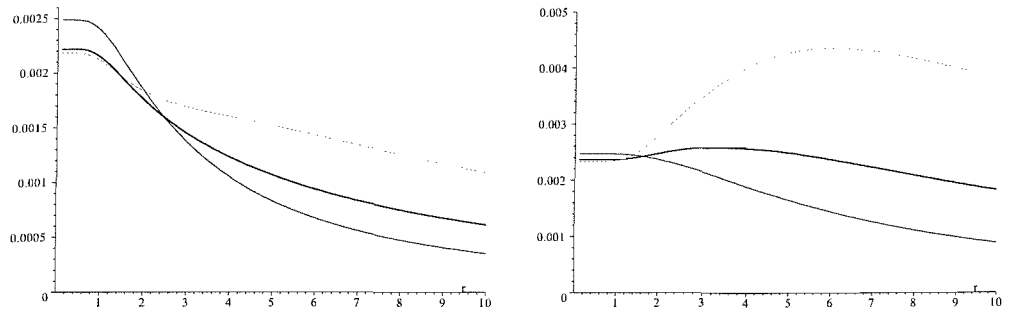
$$c(h) = c(h,r) = \begin{cases} 1 - (3\sqrt{h}/r + (\sqrt{h}/r)^3)/2 & \text{if } 0 \le h \le r^2; \\ 0 & \text{if } h > r^2. \end{cases} \tag{3.39}$$

The value of N is kept fixed, as above, at $N = 2^{12}$. We plot the variances for $d = 2$ and $d = 3$ as a function of the range r (the variogram is not valid for $d > 3$). In Figure 3.2(a) we display the results for $h = 2$ (note that this is a diagonal direction in the sense of GGS—for any d), and in Figure 3.2(b) for $h = 4$. The corresponding figure for $h = 1$ is equivalent to Figure 7 in GGS, which was produced by simulation for $N = 2^8$ (note that GGS appear to have omitted a factor 2).

In Figure 3.3 we repeat this exercise for the case of an exponential variogram with sill 1, nugget 0 and (practical) range r, so that the $c(h)$ in (3.31) are given by:

$$c(h) = c(h,r) = exp\{-3\sqrt{h}/r\},\ h \ge 0. \tag{3.40}$$

In this case, all feasible values of h will appear in equation (3.31), presenting a much larger computational task for the evaluation of $var(2\hat{\gamma}\,(h))$. Nevertheless, by exploiting the structure of the generating function (3.38) to streamline the computation, the variances can be



(a) (b)

Figure 3.3: The variance of the classical estimator $2\hat{\gamma}(h)$ when the variogram is exponential. In (a) $h = 2$, in (b) $h = 4$. The variance is plotted for many values of the (practical) range r from 0 to 10, $N = 2^{12}$, $d = 2$ (thin line), $d = 3$ (thick line), $d = 4$ (dashed line).

computed efficiently. In Figure 3.3(a) we plot the variances as a function of r for $h = 2$, and in Figure 3.3(b) those for $h = 4$, in both cases for $d = 2, 3$, and 4 (the exponential variogram is valid for all d).

With a fixed number, N, of *i.i.d.* observations, we expect the variance to decrease, at least for small h ($h \leq N^{\frac{1}{d}}$) as d increases, because the number of pairs of points available to estimate $2\gamma(h)$ (for fixed h) cannot decrease as d increases, and usually increases. But, as dependence in the data increases, or h increases, one anticipates that this effect might be overturned. Both Figures 3.3(a) and 3.3(b) show that these expectations are correct: the variances are not monotonic in r, sometimes increasing with r initially, then decreasing. And the non-monotonicity is more pronounced for larger h, and for the case of a spherical variogram. Note that the lack of smoothness for low values of r evident in Figure 3.2 arises because the spherical variogram itself is not smooth. For sufficiently large values of r—the values most likely to be used in applications—the variance for fixed h is increasing in d for both variograms—as suggested by GGS.

Of course, the usefulness of Lemma 3.3.7 is in providing a means to compute $var(2\hat{\gamma}(h))$ (and covariances) exactly in applications. For the exponential this is not a trivial computation, because as we note above, $c(h) \neq 0$ for all feasible values of h, so that $[S_H(t)]$ in (3.32) contains all feasible values. In practice, however, perfectly satisfactory accuracy can be achieved by truncating the $c(h, r)$ at some point.

3.4 Discussion

We have provided simple formulae and generating functions for the spatial design matrices implicitly defined by quadratic forms that arise in the analysis of isotropic spatial models on uniform grids, extending and simplifying the results in Genton (1988), and Gorsich, Genton, and Strang (2002). Such models are a natural generalization of familiar time series models—the one-dimensional case—and the structural results we have derived reflect this relation. These results show that in general these matrices are sums of non-commuting matrices—Kronecker products of their counterparts for the one-dimensional case—and hence that their eigenvalues are unlikely to be expressible in terms of those of the summands.

Fortunately, to study the properties of the associated quadratic forms the eigenvalues themselves are not needed: the generating functions for the matrices themselves induce generating functions for their cumulants. We provide detailed results on the means, variances and covariances of these statistics. As an important application of these results, we give simple formulae for the normalizing constant needed to produce an unbiased estimator of the variogram, and, assuming second-order stationarity, the covariance matrix needed to implement generalized least squares procedure for variogram estimation (see Cressie, 1993, Ch. 6). Finally, we briefly study some properties of the classical variogram estimator for the cases of some popular choices of the actual variogram.

For the purposes of hypothesis testing the normalized statistics $\bar{q}_h^* = z'A_h z/z'z$ and $\bar{q}_h = z'L_h z/z'z$ are of greater interest. But since exact distribution theory for such statistics is difficult, various techniques for approximating the distributions based on just the low-order cumulants have been developed (see, for instance, Durbin and Watson, 1951, Ali, 1987, or Henshaw, 1996). Although we do not implement them here, the results in Section 3.3 make such techniques quite straightforward. It is easily seen that, under the assumption that $z \sim N(0, \sigma^2 I_N)$—usually the null hypothesis—the ratios $\bar{q}_h^*$ and $\bar{q}_h$ are independent of their denominator, so that the moments of the ratios are ratios of the moments. Hence the cumulant results for q_h^* and q_h given in Section 3.3 can also be used to study or approximate the properties of $\bar{q}_h^*$ and $\bar{q}_h$ under this assumption.

It is, of course, both analytically and computationally convenient if the eigenvalues, or good approximations to them, of L_h and A_h are known. One possible device for developing approximations in the case $d = 1$ is to replace the F_r by their circular counterparts (see Anderson, 1971), and our results allow that approach to be adapted to higher dimensional

cases straightforwardly. Our work on that subject is reported in the next chapter of this thesis.

Appendix. Proofs

Proof of Proposition 3.2.1 For each pair $(\beta, \delta) \in \Gamma \times \Gamma$, define $\alpha \in \Gamma$ by $\alpha(i) = |\beta(i) - \delta(i)|$, $i = 1, ..., d$. From the definition of A_h, $(A_h)_{\beta,\delta} = 1$ if and only if $\|\alpha\|^2 = h$, or $\alpha \in \Gamma_h$. On the other hand, the (β, δ) element of $(F_{\alpha(1)} \otimes F_{\alpha(2)} \otimes ... \otimes F_{\alpha(d)})$ is one if and only if

$$|\beta(i) - \delta(i)| = \alpha(i), \text{ for } i = 1, ..., d.$$

Summing the $F_\alpha^{\otimes}$ over Γ_h must therefore yield A_h by the remark following (3.8).

Proof of Lemma 3.2.6 Let $g_h = \max_{\alpha\in\Gamma} D_h(\alpha)$ denote the maximum number of $h-$neighbors for any point in the grid Γ. The number m_h is the number of $h-$neighbors of the origin, so that $g_h \geq m_h$. Under the condition that no sequence $\alpha \in \Gamma_h$ contains an element $\alpha(i) > n/2$, we have $g_h = u_h$. To see this, suppose first that Ω_h contains just the single sequence ω_h. If $k_{\omega_h}(0) = 0$, $g_h = 2^d m_h$ because, under the stated condition, $\max_{\alpha\in\Gamma} D_h(\alpha)$ occurs at a sequence α for which the $h-$neighbors in all 2^d directions enter $D_h(\alpha)$, and m_h counts just the $h-$neighbors β in the direction for which the vector $\beta - \alpha$ has only positive components. If $k_{\omega_h}(0) > 0$, only $2^{d-k_{\omega_h}(0)}$ distinct directions are needed. Repeating the argument for each $\omega \in \Omega_h$ proves the claim $g_h = u_h$. Finally, when the condition that no $\alpha(i)$ exceeds $n/2$ is dropped, it is clear that $g_h \leq u_h$. The assertions $\lambda_h \leq u_h$, $\mu_h \leq 2u_h$ follow by Gershgorin's theorem (see Marcus and Minc, 1969).

Proof of Lemma 3.3.2 The first two cumulants are straightforward. By a standard result, $cov(q_{h_1}^*, q_{h_2}^*) = 2tr[A_{h_1} A_{h_2}]$; to show that this is zero, consider a diagonal element of $A_{h_1} A_{h_2}$:

$$(A_{h_1} A_{h_2})_{\alpha,\alpha} = \sum_{\beta\in\Gamma} (A_{h_1})_{\alpha,\beta} (A_{h_2})_{\beta,\alpha} \; \alpha \in \Gamma.$$

The product $(A_{h_1})_{\alpha,\beta}(A_{h_2})_{\beta,\alpha}$ vanishes unless both $\|\alpha - \beta\|^2 = h_1$ and $\|\beta - \alpha\|^2 = h_2$, which is impossible. Hence, for each $\alpha \in \Gamma$, every term in the sum on the right here vanishes.

Proof of Lemma 3.3.3 Consider a diagonal element of A_h^p:

$$(A_h^p)_{\alpha,\alpha} = \sum_{\beta_1,\beta_2,...,\beta_{p-1}\in\Gamma} (A_h)_{\alpha,\beta_1} (A_h)_{\beta_1,\beta_2} ... (A_h)_{\beta_{p-2},\beta_{p-1}} (A_h)_{\beta_{p-1},\alpha}, \; \alpha \in \Gamma. \tag{3.41}$$

This is non-zero only if

$$\|\alpha - \beta_1\|^2 = \|\beta_1 - \beta_2\|^2 = ... = \|\beta_{p-1} - \alpha\|^2 = h. \tag{3.42}$$

Expanding each term $\left\|\beta_i - \beta_{i+1}\right\|^2$ as $\|\beta_i\|^2 + \left\|\beta_{i+1}\right\|^2 - 2\langle\beta_i, \beta_{i+1}\rangle$ and adding the p terms gives (with $\beta_0 = \beta_p = \alpha$) :

$$2\left(\|\alpha\|^2 + \sum_{i=1}^{p-1}\|\beta_i\|^2 - \sum_{i=0}^{p-1}\langle\beta_i, \beta_{i+1}\rangle\right) = ph.$$

The left side is certainly an even integer, so when ph is odd we obtain a contradiction. Thus, when ph is odd, every term in the expression above for $(A_h^p)_{\alpha,\alpha}$ vanishes, for all $\alpha \in \Gamma$, implying $tr[A_h^p] = 0$.

Proof of Lemma 3.3.4 It is required to prove that, for $d = 2$ and every $h \geq 1$, $tr[A_h^p] = 0$ for all odd p. By equations (3.41) and (3.42) in the previous proof, $tr[A_h^p]$ is non-zero only if there is a sequence $(\alpha, \beta_1, ..., \beta_{p-1}, \alpha)$ of elements of $\Gamma(n, 2)$ such that two consecutive elements are at squared Euclidean distance h; we call such a sequence a "cycle", more precisely an odd-cycle when p is odd, and we refer to h as the "step" of the cycle. We then need to prove that no odd-cycles exist on a 2-dimensional grid. When h is odd, this is guaranteed by Lemma 3.3.3. Conversely, suppose that one or more odd-cycles of even step exist and let h^* denote the minimum step of such cycles. Observe that any element of a cycle of even step must belong to the restriction $\bar{\Gamma}(n, 2)$ of $\Gamma(n, 2)$, with an element α of $\Gamma(n, 2)$ belonging to $\bar{\Gamma}(n, 2)$ if and only if $\alpha(1) + \alpha(2)$ is even. Thus an odd-cycle of step h^* on $\Gamma(n, 2)$ is also an odd-cycle on $\bar{\Gamma}(n, 2)$. But, $\bar{\Gamma}(n, 2)$ is itself, after suitable shrinking, a uniform grid, leading to the contradiction that h^* cannot be the minimum even step of an odd-cycle on a 2-dimensional uniform grid. This completes the proof of the lemma.

Chapter 4

Circular Approximation to the Design Matrices of Isotropic Spatial Processes

Abstract

The low-order cumulants of the quadratic forms associated to the so-called spatial design matrices are needed for several inferential purposes in the context of isotropic processes defined on uniform grids. The computation required to obtain such cumulants may be intensive when the grids are large, essentially because the eigenvalues of the design matrices are generally not known in closed form. To alleviate the problem, this paper develops a circular approximation to the spatial design matrices, and discusses its use in approximating the aforementioned cumulants, in the case of independent data and of second-order stationary and isotropic processes.

Keywords: circulant matrices; isotropic spatial processes; Kronecker product; quadratic forms; second-order stationary processes; spatial design matrices.

4.1 Introduction

Let $\Gamma = \Gamma(n_1, ..., n_d)$ be a d-dimensional uniform grid with n_i sites on the i-th axis, i.e., the set of the $N = \Pi_{i=1}^{d} n_i$ sequences $\alpha = (\alpha(1), ..., \alpha(d))$ of non-negative integers satisfying $0 \leq \alpha(i) \leq n_i - 1$. For convenience, we order the sequences in Γ lexicographically. The so-called *spatial design matrices* are matrices indexed by Γ that arise naturally in the analysis of isotropic spatial processes; see Genton (1998), Gorsich, Genton and Strang (2002) and Hillier and Martellosio (2006). Here we use the notation of the latter paper, hereafter abbreviated to HM. Call two elements $\alpha, \beta \in \Gamma$ *h−neighbors* if the squared Euclidean distance between α and β is h, i.e., $\|\alpha - \beta\|^2 = \sum_{i=1}^{d} (\alpha(i) - \beta(i))^2 = h$. A spatial design matrix L_h at a squared distance $h = 1, 2, ...$ has (α, β) off-diagonal entry equal to -1 if α and β are h-neighbors and equal to 0 otherwise, and (α, α) entry equal to the number of h-neighbors of β. For convenience, every matrix L_h is decomposed in (minus) its off-diagonal part A_h and its diagonal part D_h, so that $L_h = D_h - A_h$. Note that for each $\alpha \in \Gamma$, the (α, α) entry of D_h is equal to the α-th row sum of A_h. When $h = 0$, it is convenient to define $A_0 = I_N$.

Let $F_r^{(n)}$, for $r = 0, ..., n-1$, denote the $n \times n$ matrices with (i, j) entry equal to 1 if $|i - j| = r$ and 0 otherwise. By straightforward extension of Proposition 1 in HM, A_h admits the simple representation

$$A_h = \sum_{\alpha \in \Gamma_h} F_\alpha^{\otimes}, \tag{4.1}$$

where $\Gamma_h = \{\alpha \in \Gamma : \|\alpha\|^2 = h\}$ and $F_\alpha^{\otimes}$ denotes the Kronecker product of the matrices $F_{\alpha(1)}^{(n_1)}, ..., F_{\alpha(d)}^{(n_d)}$. Let $z = (Z(\alpha); \alpha \in \Gamma)'$ be an N-dimensional vector of random variables observed on the grid and let $N(h)$ denote the set of unordered pairs of sequences that are h-neighbors in Γ. Important statistics associated to L_h are

$$q_h = \sum_{N(h)} (Z(\alpha) - Z(\beta))^2 = z' L_h z,$$

and

$$q_h^* = 2 \sum_{N(h)} Z(\alpha) Z(\beta) = z' A_h z.$$

These quadratic forms play a central role in various inferential procedures in the context of spatial processes defined on Γ. Part of the reason for this is that q_h / N_h, where N_h denotes the cardinality of $N(h)$, is an unbiased estimator of the variogram at distance $\sqrt{h}$ of an intrinsically stationary and isotropic process and $q_h^*/(2N_h)$ is (when $E(z) = 0$) an unbiased estimator of the covariance at distance $\sqrt{h}$ of a second-order stationary and isotropic process.

Moreover, the statistics $\bar{q}_h = q_h/(z'z)$ and $\bar{q}_h^* = q_h^*/(z'z)$ (often upon some normalization that is here irrelevant) are commonly employed for testing purposes. Indeed, when $d = 1$, $\bar{q}_h$ is the von Neumann statistic at time lag $\sqrt{h}$ (von Neumann *et al.*, 1941) and $\bar{q}_h^*$ is the serial correlation coefficient at time lag $\sqrt{h}$; when $d > 1$, $\bar{q}_h$ is a Geary statistic (Geary, 1954) and $\bar{q}_h^*$ is a Moran statistic (Moran, 1950). Note that when z is a vector of regression residuals, the von Neumann and the Moran statistics take the name of, respectively, Durbin-Watson and Cliff-Ord statistics (Durbin and Watson, 1951, Cliff and Ord, 1972). More applications are discussed in the introductions of Gorsich, Genton and Strang (2002) and HM.

In this paper, we assume that the random vector z has a joint Gaussian distribution. Several authors, especially in geostatistics, have been concerned with the cumulants of the quadratic forms associated to the spatial design matrices under Gaussianity. Such cumulants have direct applications (for instance, the covariance matrix of q_h is required for generalized least squares fitting of the variogram) or are useful to derive approximations to the densities of the quadratic forms (for instance, Durbin and Watson, 1951, Ali, 1987, and Henshaw, 1966). Cressie (1985) considers the case when z is a Gaussian intrinsically stationary process and analyses variances and covariances (for different values of h) of the non-isotropic counterparts of q_h, which are equivalent to q_h only in one dimension. Genton (1988) deals with the isotropic case in a general dimension d, but only for the case of independent observations and when only non-diagonal directions are considered (that is, when A_h is defined by $A_h(\alpha, \beta) = 1$ if and only if $\|\alpha - \beta\|^2 = 1$ and $\alpha - \beta$ contains $d-1$ zeros). Gorsich, Genton and Strang (2002) provide generalizations and, in addition, they study by simulation the variance of q_h/N_h under second-order stationarity and isotropy. HM derive a complete structural representation of the matrices A_h and L_h and propose generating functions for the computation of the cumulants of the associated quadratic forms.

Part of the difficulty in analyzing the properties of spatial design matrices and associated quadratic forms is that, for any $n > 2$, the matrices $F_r^{(n)}$, $r = 0, ..., n-1$, are not pairwise commutative. This makes it difficult, or perhaps even impossible, to express the eigenvalues of sums of matrices like (4.1) in terms of the eigenvalues of its summands. A consequence pointed out by HM, Section 3.1, is that the computation of the cumulants of q_h and q_h^*, albeit possible, becomes demanding when N is large. This is true not only under general assumptions—such as second-order stationarity and isotropy—on the underlying spatial process, but also (for the cumulants of order higher than two) for i.i.d. data.

In the present contribution, we overcome these problems by approximating A_h with a

matrix $\widetilde{A}_h$ having a "more convenient" structure. Because it is constructed on the basis of circulant matrices, $\widetilde{A}_h$ is named *circular* spatial design matrix. As we shall show, such a matrix is useful to approximate the density or the cumulants of q_h^*, especially when the n_i are large and of similar magnitude—which is precisely one of the cases in which an approximation is most needed. Given $\widetilde{A}_h$, one can construct approximations to the full spatial design matrix L_h, the most natural one being $\widetilde{L}_h = \widetilde{D}_h - \widetilde{A}_h$, where $\widetilde{D}_h$ is the diagonal matrix containing the row sums of $\widetilde{A}_h$. For the sake of brevity, in this paper we focus on approximations to A_h, and we only indicate in a few remarks how extensions to approximations of L_h can be carried out.

The plan of the rest of the paper is as follows. Section 4.2 introduces the preliminary notions that are necessary for the subsequent analysis. In Section 4.3 we construct the matrix $\widetilde{A}_h$. Section 4.4 discusses the quality of the approximation of q_h^* by $\widetilde{q}_h^* = z'\widetilde{A}_h z$, when data are independent. Section 4.5 extends the assessment of the performance of the approximation to second-order stationary and isotropic processes. Section 4.6 concludes. All the proofs are relegated to the Appendix.

4.2 Preliminaries

We first examine the structure of the set Γ_h. Let $\sigma\alpha$ denote the action of a permutation $\sigma \in S_d$ on $\alpha \in \Gamma$, S_d being the symmetric group of degree d. Note that for a particular h, a particular $\sigma \in S_d$ and a particular $\alpha \in \Gamma_h$, the sequence $\sigma\alpha$ may or may not be in Γ_h; more precisely, for any $\sigma \in S_d$, we have that $\sigma\alpha \in \Gamma_h$ if and only if $\alpha \in \Gamma_h$ and $\sigma\alpha \in \Gamma$.

Let $n_{\min} = \min\{n_i, i = 1, ..., d\}$. The following condition allows a simple description of the set Γ_h.

Condition A *No sequence in* Γ_h *contains an element greater than or equal to* $n_{\min}$.

It is easily seen that under Condition A, $\sigma\alpha \in \Gamma$ for any $\sigma \in S_d$ and any $\alpha \in \Gamma_h$, and, consequently, Γ_h is the union of one or more orbits in Γ under the action of S_d. A set of orbit representatives is provided by the set of *non-decreasing* sequences $\omega \in \Gamma\,(n_{\min}, ..., n_{\min})$, with $\omega(1) \leq \omega(2) \leq ... \leq \omega(d)$. For simplicity, we denote such a set by Ω, without explicit reference to the dependence on d and $n_{\min}$. The set $\Gamma_h \cap \Omega$, which we denote by Ω_h, plays a crucial role in explaining the structure of the spatial design matrices. In the simple case when Ω_h is a singleton and Condition A holds, Γ_h is an orbit of the action of S_d on Γ, but, in general, the assumption of isotropy implies that Ω_h has more than one element, which is to say that

Γ_h is the union of two or more orbits. For example, when $d = 2$, $h = 25$ and $n_1, n_2 > 5$, we have $\Omega_h = \{(0,5),(3,4)\}$. In general, that is, whether Condition A holds or not, Γ_h is the union of one or more subsets of orbits in Γ under the action of S_d. For example, when $d = 2$, $h = 25$ and $n_1 = 5, n_2 > 5$, we have $\Gamma_h = \{(0,5),(3,4),(4,3)\}$. Note that Condition A is always satisfied on a hypercubic grid.

We now define the matrices which will be used later as the building blocks of our approximation to A_h. Following Anderson (1971), let C be the circulant matrix

$$C = \begin{bmatrix} 0 & I_{n-1} \\ 1 & 0 \end{bmatrix}.$$

The matrix C is orthogonal and, for a non-negative integer r,

$$C^r = \begin{bmatrix} 0 & I_{n-a_r} \\ I_{a_r} & 0 \end{bmatrix}, \tag{4.2}$$

where a_r is the residue of r to modulus n. Its trace is

$$tr\left[C^r\right] = \begin{cases} n, & \text{if } a_r = 0; \\ 0, & \text{otherwise.} \end{cases} \tag{4.3}$$

For $r = 0, 1, ..., n-1$, define the matrices

$$\widetilde{F}_r^{(n)} = \begin{cases} \frac{1}{2}\left(C^r + C^{-r}\right), & \text{if } r = 0 \text{ or } r = \frac{n}{2}; \\ C^r + C^{-r}, & \text{otherwise.} \end{cases}$$

The matrices $\widetilde{F}_r^{(n)}$ have entries

$$(\widetilde{F}_r^{(n)})_{i,j} = \begin{cases} 1, & \text{if } |i-j| = r \text{ or } |i-j| = n-r; \\ 0, & \text{otherwise,} \end{cases}$$

and their row sums are constant, by construction, equal to 1 if $r = 0$ or $r = n/2$, 2 otherwise. Note that $\widetilde{F}_r^{(n)} = \widetilde{F}_{n-r}^{(n)}$ and therefore only $\lfloor n/2 \rfloor + 1$ of the matrices $\widetilde{F}_r^{(n)}$ are distinct, e.g. the matrices $\widetilde{F}_r^{(n)}$, $r = 0, ..., \lfloor n/2 \rfloor$ ($\lfloor \cdot \rfloor$ denotes the floor function). Moreover, $\widetilde{F}_r^{(n)} = F_r^{(n)}$, for $r = 0$ or $n/2$.

Any matrix $\widetilde{F}_r^{(n)}$ is invariant under the action of the *dihedral group* D_n of order $2n$, in the sense that $P_g \widetilde{F}_r^{(n)} P_g' = \widetilde{F}_r^{(n)}$, for each $g \in D_n$, P_g being the permutation matrix representation of g. Given such a strong invariance, it is natural to expect many multiplicities in the spectrum of each $\widetilde{F}_r^{(n)}$. The eigenvalues of $\widetilde{F}_r^{(n)}$ will be denoted by $\lambda_{r,0}^{(n)}, \lambda_{r,1}^{(n)}, ..., \lambda_{r,n-1}^{(n)}$ and, if n is odd, they are $2\cos(2\pi ri/n)$, $i = 1, ..., (n-1)/2$, each with multiplicity two and 2; if n is even

and $r \neq 0, n/2$, they are $2\cos(2\pi ri/n)$, $i = 1, ..., (n-2)/2$, each with multiplicity two, 2 and $2(-1)^r$ (Anderson, 1971, Theorem 6.5.3). Finally, if $r = n/2$, they are 1 and -1, each with multiplicity $n/2$, and if $r = 0$ the eigenvalue 1 has multiplicity n.

Remark 4.2.1 The matrices $\widetilde{F}_r^{(n)}$, $r = 0, ..., \lfloor n/2 \rfloor$, are the *distance matrices* of the graph K_n with vertex set $\{0, 1, ..., n-1\}$ and edges the pairs $(i, i+1)$, $i = 0, 1, ..., n-2$, plus the pair $(0, n-1)$ (see, for instance, Biggs, 1993). As observed in Section 2.1 of HM, graph theoretic interpretations are instructive because they suggest how to extend the results for uniform grids to other index sets. Here it is worth recalling that the matrices L_h and A_h are, respectively, the Laplacian matrix and the adjacency matrix of a graph $G(\Gamma, h)$ with vertex set Γ and edges the pairs $(\alpha, \beta) \in \Gamma \times \Gamma$ such that $\|\alpha - \beta\|^2 = h$.

Remark 4.2.2 Because the vertices of K_n can be conveniently arranged on a circle, the matrices $\widetilde{F}_r^{(n)}$, $0 \leq r \leq n/2$, define the so-called *circular serial correlation coefficients* (see Anderson, 1971, Section 6.5), which are known to have simpler statistical properties than the corresponding statistics based on the matrices $F_r^{(n)}$. In fact, the matrices $\widetilde{F}_r^{(n)}$, $r = 0, ..., \lfloor n/2 \rfloor$, span an algebra, known as *Bose-Mesner* algebra (e.g., Bannai and Ito, 1984), which admits a basis of symmetric and pairwise orthogonal idempotents, and therefore is commutative, closed under multiplication and generalized inversion.[1] It is worth noting that closure under generalized matrix inversion implies that a second-order stationary model on a circle, viz. $z \sim N(0, \Sigma)$ with $\Sigma \in \sum_{r=0}^{\lfloor n/2 \rfloor} c(r)\, \widetilde{F}_r^{(n)}$, is a full exponential family (e.g. Kass and Vos, 1997) when the covariances $c(r)$, or some diffeomorphism of them, are regarded as unknown parameters. This property has well-known inferential advantages, but, unfortunately, is lost when moving to second-order stationary models on higher-dimensional grids (cf. Section 4.5).

4.3 Circular Spatial Design Matrices

For a given $h = 0, 1, ..., \sum_{i=1}^{d} (n_i - 1)^2$, let us now consider the matrix

$$\widetilde{A}_h = \sum_{\alpha \in \Gamma_h} \widetilde{F}_\alpha^{\otimes}, \tag{4.4}$$

where

$$\widetilde{F}_\alpha^{\otimes} = \bigotimes_{i=1}^{d} \widetilde{F}_{\alpha(i)}^{(n_i)}.$$

Evidently, in dimension one we have $\widetilde{A}_h = \widetilde{F}_{\sqrt{h}}^{(n_1)}$.

We can regard $\widetilde{A}_h$ either as an approximation to the "true" matrix A_h, or as an alternative spatial design matrix in its own right. Given $\widetilde{A}_h$, it is natural to consider also the full spatial

[1] A simple way to verify this is to observe that K_n is *distance-regular* and to exploit the well-known result that the distance matrices of a distance-regular graph span a Bose-Mesner algebra; e.g., Biggs (1993).

design matrix $\widetilde{L}_h = \widetilde{D}_h - \widetilde{A}_h$, where $\widetilde{D}_h$ is the diagonal matrix containing the row sums of $\widetilde{A}_h$, and in fact, in some circumstances, this provides a useful approximation to L_h. We call the spatial design matrices $\widetilde{L}_h$ and $\widetilde{A}_h$ *circular* because they are obtained by replacing the matrices $F^{(n_i)}_{\alpha(i)}$ in L_h and A_h with their circular counterparts $\widetilde{F}^{(n_i)}_{\alpha(i)}$. In this paper we focus on approximations to A_h.

In order to describe the structure of $\widetilde{A}_h$, some new notation is in order. Let Δ be the collection of all (proper and improper) subsets of $\{1, ..., d\}$. For $D \in \Delta$, define the sequence $\epsilon_D = (\epsilon_D(1), ..., \epsilon_D(d))$ by

$$\epsilon_D(i) = \begin{cases} 0, & \text{for } i \in D; \\ n_i, & \text{otherwise.} \end{cases}$$

Geometrically, the 2^d sequences ϵ_D are the corners of the grid $\Gamma(n_1 + 1, ..., n_d + 1)$.

For a pair $(\gamma, \delta) \in \Gamma \times \Gamma$, we define the sequence $\alpha_{\gamma\delta} \in \Gamma$ by $\alpha_{\gamma\delta}(i) = |\gamma(i) - \delta(i)|$, $i = 1, ..., d$. Then

Proposition 4.3.1 *For each* $(\gamma, \delta) \in \Gamma \times \Gamma$,

$$(\widetilde{A}_h)_{\gamma,\delta} = \begin{cases} 1, & \text{if } \exists\ D \in \Delta \text{ such that } \|\alpha_{\gamma\delta} - \epsilon_D\|^2 = h; \\ 0, & \text{otherwise.} \end{cases} \tag{4.5}$$

Thus, in general, the entry $(\widetilde{A}_h)_{\gamma,\delta}$ is 1 not only for the pairs (γ, δ) such that $\alpha \in \Gamma_h$ (for such pairs $(\widetilde{A}_h)_{\gamma,\delta} = 1$ because $\{0, ..., 0\} \in \Delta$), but also for some other pairs (γ, δ) such that $\alpha \notin \Gamma_h$. More precisely, $\widetilde{A}_h$ and A_h are the same matrix only in the special circumstance mentioned in the following corollary.

Corollary 4.3.2 $\widetilde{A}_h = A_h$ *if and only if, for each* $\alpha \in \Gamma_h$ *and each* $i = 1, ..., d$, $\alpha(i)$ *is* 0 *or* $n_i/2$.

In more general circumstances, it can be deduced from Proposition 4.3.1 that, when h is small relative to the magnitude of the sides of the grid, the large part of the entries of A_h and $\widetilde{A}_h$ are the same. Formally, we have:

Corollary 4.3.3 *For a pair* $(\gamma, \delta) \in \Gamma \times \Gamma$, $(\widetilde{A}_h - A_h)_{\gamma,\delta} = 0$ *if* $\alpha_{\gamma\delta}(i) < n_i - \sqrt{h}$, *for every* $i = 1, ..., d$.

Trivially, like the $\widetilde{F}_r$ (see Remark 4.2.2 above), the matrices $\widetilde{F}^{\otimes}_{\alpha}$ are pairwise commutative:

$$\widetilde{F}^{\otimes}_{\alpha}\widetilde{F}^{\otimes}_{\beta} = \bigotimes_{i=1}^{d}\left(\widetilde{F}^{(n_i)}_{\alpha(i)}\widetilde{F}^{(n_i)}_{\beta(i)}\right) = \bigotimes_{i=1}^{d}\left(\widetilde{F}^{(n_i)}_{\beta(i)}\widetilde{F}^{(n_i)}_{\alpha(i)}\right) = \widetilde{F}^{\otimes}_{\beta}\widetilde{F}^{\otimes}_{\alpha}, \tag{4.6}$$

and hence they admit a set of common eigenvectors. It is then straightforward to prove:

Lemma 4.3.4 *For U a subset of Γ, the eigenvalues of $\widetilde{A}_U = \Sigma_{\alpha\in U}\widetilde{F}_\alpha^{\otimes}$ are given by:*

$$\lambda_U^{\otimes}(\beta) = \sum_{\alpha\in U}\left(\Pi_{i=1}^d \lambda_{\alpha(i),\beta(i)}^{(n_i)}\right),\ \beta\in\Gamma. \tag{4.7}$$

In particular, for any h (and any d and $n_1,...,n_d$), the eigenvalues of $\widetilde{A}_h$ are $\lambda_{\Gamma_h}^{\otimes}(\beta)$, $\beta\in\Gamma$. To obtain such eigenvalues, the sequences in Γ_h are needed. For any grid, and for any h, these can be derived by expanding the generating function

$$\prod_{i=1}^d\left\{\sum_{r=0}^{n_i-1}(x_r t^{r^2})\right\}. \tag{4.8}$$

Namely, $\alpha\in\Gamma_h$ if and only if $\Pi_{i=1}^d x_{\alpha(i)}$ appears as a coefficient of t^h in the expansion

$$\sum_{h=0}^{\bar{h}}\left\{t^h\sum_{\alpha\in\Gamma_h}\left(\prod_{i=1}^d x_{\alpha(i)}\right)\right\}$$

of (4.8), where $\bar{h} = \Sigma_{i=1}^d (n_i-1)^2$. Note that, because of Corollary 4.3.3, the limit (for $n_1,...,n_d\to\infty$) values of $\lambda_{\Gamma_h}^{\otimes}(\beta)$, $\beta\in\Gamma$, are the asymptotic eigenvalues of A_h; for related work, in the non-isotropic case, see Martin (1986).

The graph $\widetilde{G}(\Gamma,h)$ on Γ having adjacency matrix $\widetilde{A}_h$ is derived from $G(\Gamma,h)$ simply by joining the "opposite sides" of the uniform grid Γ, i.e., by wrapping a uniform grid onto a torus of same dimension.[2] It is easily seen that the graphs $\widetilde{G}(\Gamma,h)$, contrary to the graphs $G(\Gamma,h)$, are *regular* (i.e. each vertex has the same number of neighbors), for instance by showing that their adjacency matrices $\widetilde{A}_h$ have constant row sums:

$$\begin{aligned}\sum_{\alpha\in\Gamma_h}\widetilde{F}_\alpha^{\otimes}\bigotimes_{i=1}^d l_{n_i} &= \sum_{\alpha\in\Gamma_h}\bigotimes_{i=1}^d \widetilde{F}_{\alpha(i)}^{(n_i)} l_{n_i}\\ &= \sum_{\alpha\in\Gamma_h}\prod_{i=1}^d\left(2-\delta_{\alpha(i),0}-\delta_{\alpha(i),\frac{n_i}{2}}\right)\bigotimes_{i=1}^d l_{n_i} = \widetilde{m}_h l_N,\ \ say,\end{aligned} \tag{4.9}$$

where l_v denotes a $v\times 1$ vector of ones, and $\delta_{r,s}$ is 1 if $r=s$, 0 otherwise (the Kronecker delta). When $n_1=...=n_d=n$, we have from the last equality that the constant row sum $\widetilde{m}_h$ is

$$\widetilde{m}_h = \sum_{i=0}^d 2^{d-i}q_i, \tag{4.10}$$

q_i being the number of sequences in Γ_h with exactly i elements equal to 0 or $n/2$. Now, $\sum_{i=0}^d q_i$ is equal to the cardinality of Γ_h, denoted by m_h in HM, and therefore, if no $\alpha\in\Gamma_h$

[2] Of course, the idea of replacing a lattice by its toroidal counterpart to approximate properties of stochastic processes is very natural in many contexts; see, for instace, Anderson (1971) for the case $d=1$, Besag and Moran (1975) for the case $d=2$.

contains either 0 or $n/2$, $\widetilde{m}_h = 2^d m_h$. It is also worth mentioning that $\widetilde{m}_h$ is the maximum eigenvalue of $\widetilde{A}_h$ (and thus the maximum value of the statistic $\widetilde{q}_h^*/z'z$, usually employed for testing purposes), by Gershgorin's theorem (e.g., Marcus and Minc, 1969) plus the fact that $\widetilde{m}_h$ is an eigenvalue of $\widetilde{A}_h$ (with eigenvector l_N).

The computation of $\widetilde{m}_h$ from equations (4.9) or (4.10) requires previous computation of the elements of Γ_h. This can be avoided by computing $\widetilde{m}_h$ from the expansion of an appropriate generating function. In fact, it is easily seen that

$$\widetilde{m}_h = [t^h] \prod_{i=1}^{d} \left\{ 1 + \sum_{r=1}^{n_i - 1} \left(2 - \delta_{r, \frac{n_i}{2}} \right) t^{r^2} \right\}, \tag{4.11}$$

where $[t^h]$ is the operator extracting the coefficient of t^h from the expansion in powers of t of the function which follows (see Wilf, 1994).

The sum of all elements in $\widetilde{A}_h$ is $N\widetilde{m}_h$, which we denote by $2\widetilde{N}_h$, so that $\widetilde{N}_h$ is the number of (unordered) h-neighbors on the toroidal grid or, in graph theoretic terminology, the number of edges in $\widetilde{G}(\Gamma, h)$. The corresponding quantity for non-toroidal grids has been defined in the introduction under the name of N_h. Twice this quantity can be obtained from

$$2N_h = [t^h] \prod_{i=1}^{d} \left\{ n_i + \sum_{r=1}^{n_i - 1} 2(n_i - r) t^{r^2} \right\}; \tag{4.12}$$

see HM, eq. (26).

4.4 Accuracy of the Circular Approximation for i.i.d. Data

In this section we assess the accuracy of the circular approximation $\widetilde{A}_h$, when it is used in the quadratic form $\widetilde{q}_h^* = z'\widetilde{A}_h z$ to approximate $q_h^* = z'A_h z$, and when $z \sim N(0, I_N)$ (or $z \sim N(0, \sigma^2 I_N)$). The assumption $z \sim N(0, I_N)$ arises, for instance, as the null hypothesis in tests of autocorrelation. Let $\kappa_p(q)$ denote the p-th cumulant of a statistic q. As measures of accuracy, we consider the ratios of cumulants

$$\eta_{p,h} = k_p(\widetilde{q}_h^*)/k_p(q_h^*), \ p = 2, 3...$$

Since q_h^* and the sum of squares $z'z$ are independent when $z \sim N(0, \sigma^2 I_N)$, $\eta_{p,h}$ are also the ratios of the cumulants of the quadratic forms normalized by $z'z$ (usually employed as test statistics). As for the estimator $q_h^*/(2N_h)$—which we denote by ξ_h—of the autocovariance function of a second-order stationary and isotropic process, we define

$$\phi_{p,h} = k_p(\xi_h)/k_p(\widetilde{\xi}_h), \ p = 2, 3...,$$

where, with obvious notation, $\widetilde{\xi}_h = \widetilde{q}_h^*/(2\widetilde{N}_h)$. Observe that $\phi_{p,h} = (\eta_{2,h})^p/\eta_{p,h}$, by using the easy-to-check fact that the variance of q_h^* for independent Gaussian data is $4\sigma^2 N_h$ (cf. HM, Lemma 3). When $k_p(\widetilde{q}_h^*) = k_p(q_h^*) = 0$, we define $\eta_{p,h} = \phi_{p,h} = 1$.

In Section 4.4.1 we derive some preliminary properties of the quadratic forms $\widetilde{q}_h^*$. Then in Sections 4.4.2 and 4.4.2 we study the some properties of the ratios $\eta_{p,h}$ and $\phi_{p,h}$.

4.4.1 The Quadratic Forms Associated to $\widetilde{A}_h$

The distributions of the quadratic forms $\widetilde{q}_h^*$ and of $\widetilde{q}_h$ are complicated objects, even when $z \sim N(0, I_N)$. Nevertheless, the cumulants of $\widetilde{q}_h^*$ and $\widetilde{q}_h$ are simple functions—power sums—of the known eigenvalues of $\widetilde{A}_h$ and $\widetilde{L}_h$, respectively (the eigenvalues of $\widetilde{A}_h$ have been given immediately after Lemma 4.3.4; those of $\widetilde{L}_h$ are obtained simply by adding $\widetilde{m}_h$). To be precise, recall that the p-th cumulant of a quadratic form $y'My$ in a vector $y \sim N(\mu, I_N)$ is

$$2^{p-1}(p-1)!\left(tr\left[M^p\right] + p\mu' M^p \mu\right), \; p = 1, 2, ... \tag{4.13}$$

(see, for instance, Kendall and Stuart, 1969, Chapter 15). Thus, the cumulants of $\widetilde{q}_h$ are the same under $z \sim N(0, I_N)$ and $z \sim N(\theta l_N, I_N)$ (i.e., the variables in z have constant mean θ), since l_N is an eigenvector of $\widetilde{L}_h$ associated to the eigenvalue 0. As for $\widetilde{q}_h^*$, the p-th cumulant under $z \sim N(\theta l_N, I_N)$ is equal to the p-th cumulant under $z \sim N(0, I_N)$ plus $\widetilde{m}_h^p$, because l_N is an eigenvector of $\widetilde{A}_h$ associated to the eigenvalue $\widetilde{m}_h$.

We now briefly derive some basic properties of the quadratic forms $\widetilde{q}_h^*$. It is convenient to start from the one-dimensional case. Letting $\widetilde{Q}_r^* = z'\widetilde{F}_r^{(n)}z$ denote $\widetilde{q}_{r^2}^*$ when $d = 1$, we have:

Lemma 4.4.1 *For* $r_1, r_2 = 1, 2, ..., \lfloor n/2 \rfloor - 1$,

(i) $E(\widetilde{Q}_{r_1}^*) = 0$, $var(\widetilde{Q}_{r_1}^*) = 4n$;

(ii) $\widetilde{Q}_{r_1}^*$ *and* $\widetilde{Q}_{r_2}^*$, $r_1 \neq r_2$, *are uncorrelated;*

(iii) the density of $\widetilde{Q}_{r_1}^*$ *is symmetric about zero if and only if either* n *is even and* r_1 *is odd or* n *is an even multiple of* r_1.

For the remaining cases $r_1, r_2 = \lfloor n/2 \rfloor, ..., n-1$, note that $\widetilde{Q}_{r_1}^* = \widetilde{Q}_{n-r_1}^*$ and (for even n) $\widetilde{Q}_{n/2}^* = Q_{n/2}^* = z'F_{n/2}^{(n)}z$ (see HM, Lemma 2). To generalize Lemma 4.4.1 to $d \geq 1$, consider the following condition.

Condition B *No* $\alpha \in \Gamma_h$ *contains an element* $\alpha(i)$ *greater than or equal to* $n_i/2$.

Condition B is typically satisfied in applications, because one is usually concerned with distance h such that N_h is not too small (with respect to N_1, say); this is the case, for instance, in variogram estimation (see Journel and Huijbregts, 1978). The generalization of parts (i) and (ii) of Lemma 4.4.1 is:

Lemma 4.4.2 *For* $h_1, h_2 = 1, 2, ..., \sum_{i=1}^{d} (n_i - 1)^2$,

(i) $E(\widetilde{q}_{h_1}^*) = 0$, $var(\widetilde{q}_{h_1}^*) = 4\widetilde{N}_h$;

(ii) $\widetilde{q}_{h_1}^*$ *and* $\widetilde{q}_{h_2}^*$, $h_1 \neq h_2$, *are uncorrelated if Condition B holds for* $h = h_1, h_2$.

When Condition B does not hold for $h = h_1$ or h_2, the covariance between $\widetilde{q}_{h_1}^*$ and $\widetilde{q}_{h_2}^*$ can be computed from the eigenvalues of $\widetilde{A}_{h_1}$ and $\widetilde{A}_{h_2}$, given by Lemma 4.3.4. This is because, as a result of (4.6), the $\widetilde{A}_h$ commute:

$$\widetilde{A}_{h_1}\widetilde{A}_{h_2} = \sum_{\alpha\in\Gamma_{h_1}}\sum_{\beta\in\Gamma_{h_2}} \left(\widetilde{F}_\alpha^{\otimes}\widetilde{F}_\beta^{\otimes}\right) = \sum_{\beta\in\Gamma_{h_2}}\sum_{\alpha\in\Gamma_{h_1}} \left(\widetilde{F}_\beta^{\otimes}\widetilde{F}_\alpha^{\otimes}\right) = \widetilde{A}_{h_2}\widetilde{A}_{h_1},$$

and therefore $cov(\widetilde{q}_{h_1}^*, \widetilde{q}_{h_2}^*) = 2\sum_{\beta\in\Gamma}\{\lambda_{\Gamma_{h_1}}^{\otimes}(\beta)\,\lambda_{\Gamma_{h_2}}^{\otimes}(\beta)\}$.

Let us now consider the symmetry of the density of $\widetilde{q}_h^*$ for a general dimension d. Denote by $\kappa_p(q)$ the p-th cumulant of a statistic q, and recall that the density of q is symmetric about the origin if and only if $\kappa_p(q) = 0$ for all odd p. The symmetry of the density of q_h^* is discussed in Section 3.1 of HM. Obviously, we would like to approximate q_h^* with a statistic whose density maintains the symmetries of the density of q_h^*, or, at least, such that $k_p(\widetilde{q}_h^*) = 0$ is necessary and sufficient for $k_p(q_h^*) = 0$ for small odd p. In order to establish whether this is the case, let us consider, for a given p, the following extension of Condition B.

Condition C *No* $\alpha \in \Gamma_h$ *contains an element* $\alpha(i)$ *greater than or equal to* n_i/p.

We then have:

Lemma 4.4.3 *If* $k_p(\widetilde{q}_h^*) = 0$ *then* $k_p(q_h^*) = 0$. *The converse does not hold in general, but it does hold if* $p = 1$, p *is even, or under Condition C.*

With regards to the question of whether $k_p(\widetilde{q}_h^*) = 0$ is necessary and sufficient for $k_p(q_h^*) = 0$ for small odd p, the lemma asserts that this is the case for p at least up to $n_{\min}/h$.

We remark that whether or not the density of $\widetilde{q}_h^*$ is symmetric can be established by looking at the eigenvalues of $\widetilde{A}_h$. Indeed, because of the condition on the odd cumulants, the density of $\widetilde{q}_h^*$ is symmetric about zero if and only if the spectrum of $\widetilde{A}_h$ is symmetric about zero (in the sense that for each $\beta \in \Gamma$, there is $\gamma \in \Gamma$ such that $\lambda_{\Gamma_h}^{\otimes}(\gamma) = -\lambda_{\Gamma_h}^{\otimes}(\beta)$).

4.4.2 Variances

In many applications attention is confined to the second-order structure of the quadratic forms q_h; see, for instance, Genton (1999). We now study the ratio of variances $\eta_{2,h} = var\left(\widetilde{q}_h^*\right)/var\left(q_h^*\right)$. By Lemma 4.4.2, $\eta_{2,h} = \phi_{2,h} = \widetilde{N}_h/N_h$. Throughout this subsection we assume that the distance h is feasible on a given grid, i.e., that $N_h \neq 0$. It is also convenient to assume that Condition B holds.

Clearly, $\eta_{2,h}$ is completely determined by the values of $h, d, n_1, ..., n_d$. The dependence of $\eta_{2,h}$ on h (for given d and $n_1, ..., n_d$) or on d (for given h and having decided how $n_1, ..., n_d$ transform when d changes) is non-monotonic and non-smooth. This is essentially because of the number theoretic nature of the problem under analysis, which arises from the fact that Γ_h is the set of decompositions of the integer h as a sum of squares $\sum_{i=1}^{d} \alpha(i)^2$, with $0 \leq \alpha(i) \leq n_i - 1$ (a restricted version of a "Waring's problem"; Hardy and Wright, 1979, Chapter XX).

Nevertheless, it is sufficient to examine a few numerical examples to convince oneself that, in general, $\eta_{2,h}$ increases with h and does not decrease with d (when N is held fixed or when $n_1 = ... = n_d = n$ and n is held fixed). In some of the particular cases discussed below it is easy to check that $\eta_{2,h}$ is indeed increasing with h and non-decreasing with d (see Corollaries 4.4.6 and 4.4.7).

The dependence of $\eta_{2,h}$ on the n_i's is simpler and can be revealed by deriving a suitable representation of $\eta_{2,h}$. This is done in the next proposition, which holds for any grid Γ and any distance h. Three corollaries concerned with three particular cases follow from the proposition.

Proposition 4.4.4 *Let $k_\alpha(0)$ be the number of zeros in a sequence $\alpha \in \Gamma_h$. Then,*

$$\eta_{2,h} = \frac{N \sum_{\alpha \in \Gamma_h} 2^{-k_\alpha(0)}}{\sum_{\alpha \in \Gamma_h} \left\{2^{-k_\alpha(0)} \Pi_{i=1}^{d} (n_i - \alpha(i))\right\}}. \tag{4.14}$$

Recall that the sequences $\alpha \in \Gamma_h$ needed to evaluate (4.14) can be easily derived from (4.8). Formula (4.14) simplifies considerably if the sides of the grid are of equal length and S_d acts transitively on Γ_h. In these circumstances the set Ω_h (defined in Section 4.2), has a single element, say ω_h.

Corollary 4.4.5 *If $n_1 = ... = n_d = n$ and $\Omega_h = \{\omega_h\}$, then*

$$\eta_{2,h} = \frac{n^{d - k_{\omega_h}(0)}}{\prod\limits_{\omega_h(j) \neq 0} \left(n - \omega_h(j)\right)}.$$

In particular, for $h = 1, 2, 3$—the only distances such that the action of S_d on Γ_h is transitive in any dimension d—the ratio $\eta_{2,h}$ does not depend on d, as shown by the following corollary (note that it is necessary that $d \geq h$ for h to be feasible and $n > 2$ for Condition B to hold).

Corollary 4.4.6 *When* $n_1 = ... = n_d = n$ *and* $h = 1, 2$ *or* 3, $\eta_{2,h} = [n/(n-1)]^h$.

Another case when the ratio of variances $\eta_{2,h}$ takes a simple form is when only "non-diagonal directions" are considered (see Gorsich, Genton and Strang, 2002). This is the case when Γ_h has—or is restricted to have—only sequences lying on the main axes of Γ.

Corollary 4.4.7 *If only non-diagonal directions are considered, then*

$$\eta_{2,h} = \frac{d}{d - \sqrt{h} \sum_{i=1}^{d} \frac{1}{n_i}}. \tag{4.15}$$

Note that on a hypercubic grid, (4.15) reduces to the simple formula $\eta_{2,h} = n/(n - \sqrt{h})$.

Next, we investigate how $\eta_{2,h}$ behaves as the grid gets larger. It is an immediate consequence of Corollary 4.3.3 that $\eta_{2,h} \to 1$ from above as $n_1, ..., n_d \to \infty$. Moreover, direct differentiation of (4.14) with respect to n_i shows that $\eta_{2,h}$ is strictly decreasing in each n_i, $i = 1, ..., d$ (see also Lemma 4.4.10 below). Thus, similarly to what happens in the time series case ($d = 1$), the circular approximation works better when at least one n_i is large. It is then natural to ask whether, given the sample size N, the approximation is more accurate when the grid Γ is embedded in a hyper-cube or when it is embedded in a hyper-rectangle. To answer this question, it is convenient to neglect the restriction that the n_i's must be integers.

Proposition 4.4.8 *Given* N, $\underset{(n_1,...,n_d) \in (\mathbb{R}^+)^d}{\arg\min} \{\eta_{2,h}\} = (N^{1/d}, ..., N^{1/d})$, *for any* h *and any* d.

Thus, the ideal situation for the application of the circular approximation is when the sides of the grid are of the same, or similar, length.

Using the formulae given in this section, the computation of $\eta_{2,h}$ is straightforward for any h and on any grid $\Gamma(n_1, ..., n_d)$. To give an indication of the performance of the circular approximation, some values of $\eta_{2,h}$ are reported in Table 4.1.

4.4.3 Higher-Order Cumulants

In this section, we extend the analysis of the accuracy of the circular approximation to the higher-order cumulants.

Note that for some combinations of values of h and p on some grid Γ, $\kappa_p(q_h^*)$ may vanish. For instance, by Lemma 3.3.3 in Chapter 3, $\kappa_p(q_h^*) = 0$ on any grid when ph is odd, or, by

$h \backslash n$	20	40	60	80	100	1000
1	1.0526	1.0256	1.0169	1.0127	1.0386	1.0010
2	1.1080	1.0519	1.0342	1.0255	1.0203	1.0020
5	1.1696	1.0796	1.0520	1.0386	1.0307	1.0030

Table 4.1: The values of $\eta_{2,h}$, for $h = 1, 2, 5$, on a 2-dimensional square grid of side $n = 20, 40, ..., 100$.

Lemma 3.3.4 in Chapter 3, $\kappa_p(q_h^*) = 0$ on any 2-dimensional grid when p is odd. Because of this reason in this subsection we assume that, for a given p, Γ is large enough as for Condition C to hold. In that case, by Lemma 4.4.3, when $\kappa_p(q_h^*) = 0$, $\kappa_p(\widetilde{q}_h^*) = 0$ too, and $\eta_{p,h} = 1$ by definition.

The next lemma shows that $\kappa_p(\widetilde{q}_h^*)$—which can be obtained directly from the known eigenvalues of $\widetilde{A}_h$—provides an upper bound for $\kappa_p(q_h^*)$.

Lemma 4.4.9 *For $p = 1, 2, ...$ such that $\kappa_p(q_h^*) \neq 0$, $\eta_{p,h} > 1$.*

We can also generalize the statement made above that $\eta_{2,h}$ is strictly decreasing in each n_i, $i = 1, ..., d$.

Lemma 4.4.10 *For $p = 1, 2, ...$ such that $\kappa_p(q_h^*) \neq 0$, the ratio $\eta_{p,h}$ is strictly decreasing in each n_i, $i = 1, ..., d$, and strictly increasing in p.*

As far as the normalized versions ξ_h and $\widetilde{\xi}_h$ of the quadratic forms q_h^* and $\widetilde{q}_h^*$ are concerned, we have the following conjecture, which extends to any $p > 2$ what we know for $p = 2$ (as $\phi_{2,h} = \widetilde{N}_h / N_h \geq 1$). No numerical counterexamples or a proof have been found.

Conjecture 4.4.11 *For $p = 1, 2, ...$ such that $\kappa_p(q_h^*) \neq 0$, $\phi_{p,h} \geq 1$.*

The conjecture does no need to hold when Condition C is violated. In that case $\kappa_p(\widetilde{q}_h^*)$ may be much greater than $\kappa_p(q_h^*)$ (due to the fact that there may be p-sided polygons of side h which are embedded in the toroidal version of Γ but not in Γ itself), and the normalization by $2\widetilde{N}_h$ and $2N_h$ may not be sufficient to revert the order of the inequality $\kappa_p(\widetilde{q}_h^*) > \kappa_p(q_h^*)$ established by Lemma 4.4.9.

Lemma 4.4.12 *For $p = 1, 2, ...$ such that $\kappa_p(q_h^*) \neq 0$, the ratio $\phi_{p,h}$ is strictly increasing in p.*

In view of Lemma 4.4.10, one would expect $\phi_{p,h}$ to be decreasing in each n_i, $i = 1, ..., d$. This is indeed always the case in the numerical studies we have performed, but a proof as simple as that for $\eta_{p,h}$ does not seem possible.

In Table 4.2 some values of $\eta_{4,1}$ are reported (note that $\eta_{3,1}$ is not defined, as $k_3(q_1^*) = 0$ on any uniform grid); $k_p(\widetilde{q_h^*})$ is computed by exploiting the known eigenvalues of $\widetilde{A}_h$, and $k_p(q_h^*)$ by differentiating the generating functions given in HM. Table 4.3 gives the corresponding values of $\phi_{p,h}$.

$d \backslash n$	20	40	60	80	100
2	1.3673	1.1670	1.1084	1.0801	1.0634
3	1.3646	1.1657	1.1076	1.0795	1.0629
4	1.3631	1.1653	1.1072	1.0793	1.0627

Table 4.2: Some values of $\eta_{4,1}$ on a d-dimensional hypercubic uniform grid of side n.

$d \backslash n$	20	40	60	80	100
2	1.1138	1.0548	1.0361	1.0269	1.0215
3	1.1116	1.0536	1.0353	1.0263	1.0210
4	1.1104	1.0532	1.0350	1.0261	1.0208

Table 4.3: Some values of $\phi_{4,1}$ on a d-dimensional hypercubic uniform grid of side n.

4.5 Second-Order Stationary Isotropic Processes

In this section, we analyze the performance of the circular approximation for deriving the properties of $\xi_h = q_h^*/(2N_h)$, when the underlying spatial process is second-order stationary and isotropic processes. A stochastic process $\{Z(\alpha), \alpha \in \Lambda \subseteq \mathbb{R}^d\}$ is said to be second-order stationary and isotropic if $E(Z(\alpha))$ is constant and $cov(Z(\alpha), Z(\beta))$ depends on the sites α and β only through $\|\alpha - \beta\|$. Thus, when $\Lambda = \Gamma$, the covariance matrix of z can be represented as

$$\Sigma = \sum_{h \in H} c(h) A_h,$$

where H is a set of squared distances including $h = 0$, and $c(h)$ is the covariance between pairs of variables observed at sites at squared distance h. We assume that the $c(h)$'s, $h \in H$, are

such that Σ is positive-definite, and that the joint density of the spatial process is Gaussian (but extensions to more general elliptically symmetric distributions are certainly possible, in much the same way as, for instance, in Genton, 1999). For the sake of simplicity, we also assume $E(Z(\alpha)) = 0$.

In Section 3.2 of HM a procedure is given to compute the cumulants of q_h^* and q_h for a second-order stationary and isotropic processes. Although such a procedure is in principle applicable on any uniform grid to derive cumulants of any order, in practice the computation may become very intensive when the grid is large and/or the cardinality of H is large. This is true even for low-order cumulants, and for product cumulants such as the covariances between two quadratic forms q_h^* or q_h for two different values h. Because of this computational problem, and given also that in applications one generally needs to consider quadratic forms for several values of h, an approximation is desirable.

In the following, we limit ourselves to the mean and the variance of $\xi_h = q_h^*/(2N_h)$, but extensions to higher-order cumulants or to q_h are straightforward. Observe that $2N_h$ can be readily evaluated from the generating function (4.12) and recall that the p-th cumulant of a quadratic form $z'Mz$ in $z \sim N(0, \Sigma)$ is given by $\kappa_p = 2^{p-1}(p-1)!tr[(M\Sigma)^p]$, $p = 1, 2, ...$ (e.g., Kendall and Stuart, 1969, Chapter 15). We have

$$E(\xi_h) = \frac{1}{2N_h} tr[A_h\Sigma] = \frac{1}{2N_h} \sum_{k \in H} c(k) tr[A_h A_k] = \begin{cases} c(h), & \text{if } h \in H; \\ 0, & \text{otherwise,} \end{cases}$$

because $tr[A_h A_k] = 2N_h$ if $h = k$ and vanishes otherwise. From the results in Section 3.2 of HM, and particularly their Lemma 7, one obtains

$$var(\xi_h) = \frac{1}{2N_h^2} tr[(A_h\Sigma)^2] = \frac{1}{2N_h^2} [(s_1 s_2)^h][S_H(t_1)][S_H(t_2)] v_n^d(s_1, s_2, t_1, t_2), \tag{4.16}$$

where $[S_H(t)]$ is a linear operator constructed as

$$[S_H(t)] = \sum_{h \in H} c(h)[t^h],$$

and

$$v(s_1, s_2, t_1, t_2) = \prod_{i=1}^{d} tr[B_{n_i}(s_1) B_{n_i}(t_1) B_{n_i}(s_2) B_{n_i}(t_2)],$$

with

$$B_n(t) = \sum_{r=0}^{n-1} t^{r^2} F_r.$$

Expression (4.16) makes it clear that the memory usage and the computation required to obtain $var(\xi_h)$ may be large, because of the required multiplication of big matrices (when

at least one n_i is large) and of the large number of accesses to the generating function $v(s_1, s_2, t_1, t_2)$ (when the cardinality of H is large). In order to approximate $var(\xi_h)$, we define a "circular process" $z \sim N(0, \widetilde{\Sigma})$, with

$$\widetilde{\Sigma} = \sum_{h \in H} \widetilde{c}(h) \widetilde{A}_h.$$

At this stage, we do not insist on a particular relationship between $\widetilde{c}(h)$ and $c(h)$, but we will shortly see that, given a simplifying assumption, it is convenient to set $\widetilde{c}(h) = c(h)$.

A natural (method of moments) estimator of $\widetilde{c}(h)$ is $\widetilde{\xi}_h = \widetilde{q}_h^* / (2\widetilde{N}_h)$. Its expected value when $z \sim N(0, \widetilde{\Sigma})$ is, for a fixed h,

$$E(\widetilde{\xi}_h) = \frac{1}{2\widetilde{N}_h} tr[\widetilde{A}_h \widetilde{\Sigma}] = \frac{1}{2\widetilde{N}_h} \sum_{k \in H} \widetilde{c}(h) tr[\widetilde{A}_h \widetilde{A}_k]. \tag{4.17}$$

Note that the term $\widetilde{c}(h) tr[\widetilde{A}_h^2] / (2\widetilde{N}_h) = \widetilde{c}(h)$, for $h \in H$, always gives a non-zero contribution to $E(\widetilde{q}_h^*)$; additional terms are present only if H contains squared distances k (different from h) such that $tr[\widetilde{A}_h \widetilde{A}_k]$ does not vanish.

Let us now assume that Condition B holds for any $h \in H$. Note that such an assumption limits the range of dependence in the process z, in that it entails that two variables $Z(\alpha)$ and $Z(\beta)$ are correlated only if the i-th component of $\alpha - \beta$ is less than $n_i/2$, for each $i = 1, ..., d$. This, of course, is likely to be reasonable when all the n_i's are large, but may be restrictive otherwise.

The main analytical advantage of assuming that Condition B holds for any $h \in H$ is that it entails that the matrices $\widetilde{A}_k$, $k \in H$, are—as their non-circular counterparts—disjoint. As a consequence,

$$E(\widetilde{\xi}_h) = \begin{cases} \widetilde{c}(h), & \text{if } h \in H; \\ 0, & \text{otherwise.} \end{cases}$$

In this case, under the stated assumption, it is appropriate to take $\widetilde{c}(h) = c(h)$, for $h \in H$, so that, for any h, $\widetilde{\xi}_h$ and ξ_h have the same expectation. An approximation to $var(\xi_h)$ is therefore given by

$$var(\widetilde{\xi}_h) = \frac{1}{2\widetilde{N}_h^2} tr[(\widetilde{A}_h \widetilde{\Sigma})^2] = \frac{1}{2\widetilde{N}_h^2} tr \left[\left\{ \sum_{k \in H} (c(k) \widetilde{A}_h \widetilde{A}_k) \right\}^2 \right]$$

which, in view of the pairwise commutativity of the matrices $\widetilde{A}_h$, can be computed with very

little computational effort from

$$var(\widetilde{\xi}_h) = \frac{1}{2\widetilde{N}_h^2} \sum_{k_1,k_2 \in H} \left\{ c(k_1)c(k_2) \sum_{\beta \in \Gamma} \left[\left(\lambda^{\otimes}_{\Gamma_h}(\beta) \right)^2 \lambda^{\otimes}_{\Gamma_{k_1}}(\beta) \lambda^{\otimes}_{\Gamma_{k_2}}(\beta) \right] \right\}$$

(recall that the eigenvalues $\lambda^{\otimes}_{\Gamma_k}(\beta)$, $k \in H$, $\beta \in \Gamma$, can be derived from the combination of Lemma 4.3.4 and the generating function (4.8)).

To demonstrate the performance of $var(\widetilde{\xi}_h)$ in approximating $var(\xi_h)$, we use the example of a spatial process with a spherical variogram with sill 1, nugget 0 and range r. Such a process is second-order stationary and isotropic with covariances given by

$$c(h) = c(h,r) = \begin{cases} 1 - \frac{1}{2}(3\sqrt{h}/r + (\sqrt{h}/r)^3), & \text{if } 0 \leq h \leq r^2; \\ 0, & \text{if } h > r^2. \end{cases} \tag{4.18}$$

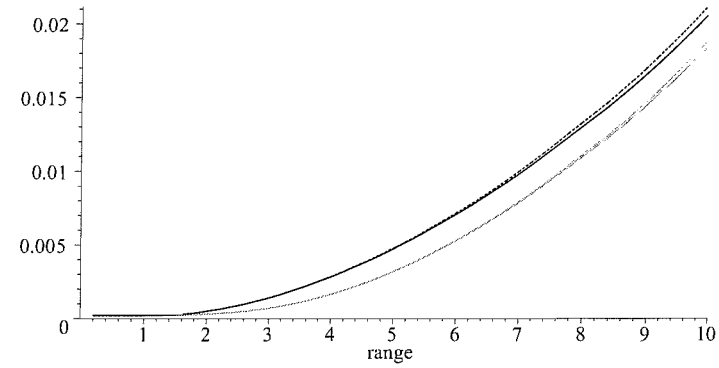
In Figure 4.1(a) $var(\xi_h)$ and $var(\widetilde{\xi}_h)$ are plotted as a function of r for $d = 2$, $n_1 = n_2 = 50$ and $h = 1, 4$ (although it is not visible in the figure, $var(\widetilde{\xi}_h) < var(\xi_h)$ when $r \leq 1$, as predicted by Conjecture 4.4.11). Figure 4.1(b) is the same as Figure 4.1(a), but for the statistic q_h/N_h—the (unbiased) classical estimator of the variogram of the process—rather than for ξ_h.

4.6 Conclusions

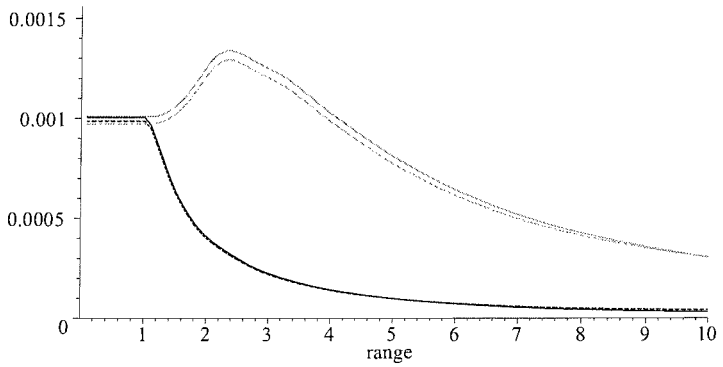
The paper has developed a circular approximation to the spatial design matrices of isotropic processes on d-dimensional uniform grids, and has investigated the performance of the approximation when it is used to compute the cumulants of the quadratic forms q_h^*. Such quadratic forms, suitably normalized, are recurrent in many inferential procedures in the context of isotropic spatial processes.

The approximation is useful when it is most needed, that is, when the uniform grids are large. The saving in computational time can be enormous for i.i.d. Gaussian data, because in that case the cumulants are simple functions of the eigenvalues of the design matrices, and the eigenvalues of the circular spatial design matrices, contrary to the non-circular ones, are known in closed form. We have shown that the same argument applies, under suitable assumptions, to Gaussian second-order stationary and isotropic processes.

Extensions to the quadratic forms q_h in the full spatial design matrices L_h have been considered only briefly. Further work is required before recommending the circular approximation for that case.



(a)



(b)

Figure 4.1: For a spatial process defined on a 2-dimensional square grid of side 50, and with a spherical variogram, (a) displays $var(\xi_h)$ (solid lines) and $var(\widetilde{\xi}_h)$ (dotted lines) as a function of the range, for $h = 1$ (dark lines), and $h = 4$ (light lines); (b) is the same as (a) but for $var(q_h/N_h)$ and $var(\widetilde{q}_h/\widetilde{N}_h)$.

Appendix. Proofs

Proof of Proposition 4.3.1 $\widetilde{F}_{\beta}^{\otimes}(\gamma,\delta) = 1$ only if either $|\gamma(i) - \delta(i)| = \beta(i)$ or $n - \beta(i)$ for each $i = 1, ..., d$. Hence, the (γ,δ) element of $\widetilde{A}_h$ is one if and only if $\alpha_{\gamma\delta}$ belongs to the set $\{\epsilon_D - \beta,\ \beta \in \Gamma_h,\ D \in \Delta\}$. That is, there must exist $D \in \Delta$ such that $\|\alpha_{\gamma\delta} - \epsilon_D\|^2 = h$.

Proof of Corollary 4.3.2 Observe that the condition in the corollary implies that there is no pair $(\gamma,\delta) \in \Gamma \times \Gamma$ such that $\|\gamma - \delta\|^2 \neq h$ and $\|\alpha_{\gamma\delta} - \epsilon_D\|^2 = h$ for a $D \in \Delta$, with $D \neq \{0, ..., 0\}$. The sufficiency of the condition is then a simple consequence of Proposition 4.3.1. Its necessity follows from the fact that if there is an $\alpha \in \Gamma_h$ having a component $\alpha(i)$ different from 0 or $n_i/2$, then there is at least one pair (γ,δ) such that $(\widetilde{A}_h)_{\gamma,\delta} \neq (A_h)_{\gamma,\delta}$, again by Proposition 4.3.1.

Proof of Corollary 4.3.3 Clearly, $(\widetilde{A}_h - A_h)_{\gamma,\delta}$ is zero if γ and δ are h-neighbors. If γ and δ are not h-neighbors (i.e., $(A_h)_{\gamma,\delta} = 0$), then it can be deduced from Proposition 4.3.1 that $(\widetilde{A}_h)_{\gamma,\delta} = 0$ if $\alpha_{\gamma\delta}(i) < n_i - \sqrt{h}$, $i = 1, ..., d$, because under such a condition $\|\alpha_{\gamma\delta} - \epsilon_D\|^2 = \sum_{i=1}^{n}(\alpha_{\gamma\delta}(i) - \epsilon_D(i))^2$ is greater than h for any $D \in \Delta$ other than $D = \{0, ..., 0\}$.

Proof of Lemma 4.3.4 Let $x_1, ..., x_n$ be the common eigenvectors of each $\widetilde{F}_r$. Then, for each $\beta \in \Gamma$,

$$\begin{aligned}\left(\sum_{\alpha\in U}\widetilde{F}_{\alpha}^{\otimes}\right)(x_{\beta(1)} \otimes .... \otimes x_{\beta(d)}) &= \sum_{\alpha\in U}(\widetilde{F}_{\alpha(1)}x_{\beta(1)} \otimes ... \otimes \widetilde{F}_{\alpha(d)}x_{\beta(d)}) \\ &= \left(\sum_{\alpha\in U}(\Pi_{i=1}^{d}\lambda_{\alpha(i),\beta(i)})\right)(x_{\beta(1)} \otimes .... \otimes x_{\beta(d)}),\end{aligned}$$

so that the $\lambda_U^{\otimes}(\beta)$, $\beta \in \Gamma$, are eigenvalues of $\widetilde{A}_U$.

Proof of Lemma 4.4.1 *(i)* For $r_1 = 1, 2, ..., \lfloor n/2 \rfloor - 1$, $E(\widetilde{Q}_{r_1}^*) = tr[\widetilde{F}_{r_1}^{(n)}] = 0$ and $var(\widetilde{Q}_{r_1}^*) = 2tr[(\widetilde{F}_{r_1}^{(n)})^2] = 2tr[(C^{r_1} + C^{-r_1})^2] = 2tr[C^{2r_1} + C^{-2r_1} + 2I_n]$, which by equation (4.3), is equal to $2tr[2I_n] = 4n$.

(ii) The covariance between $\widetilde{Q}_{r_1}^*$ and $\widetilde{Q}_{r_2}^*$ is $2tr[\widetilde{F}_{r_1}^{(n)}\widetilde{F}_{r_2}^{(n)}]$, which does not vanish if and only if there is at least one integer $0 \leq i \leq n-1$ such that $|i - r_1| = |i - r_2|$. This is clearly impossible if $r_1 \neq r_2$, and $r_1, r_2 = 1, 2, ..., \lfloor n/2 \rfloor - 1$;

(iii) The density of $\widetilde{Q}_{r_1}^*$ is symmetric about zero if and only if all its odd cumulants vanish, that is, if and only if $tr[(\widetilde{F}_{r_1}^{(n)})^p] = 0$ for all odd p. Since, for $r_1 = 1, 2, ..., \lfloor n/2 \rfloor - 1$, $\widetilde{F}_{r_1}^{(n)} = C^{r_1} + C^{-r_1}$, such a condition is equivalent to $\sum_{k=0}^{p}\binom{p}{k}tr[C^{r_1(2k-p)}] = 0$ for all odd p. By equation (4.2), $tr[C^{r_1}]$ is equal to n if $a_{r_1} = 0$, to 0 otherwise, for each r_1. Thus, by equation (4.3), the density of $\widetilde{Q}_{r_1}^*$ is symmetric about zero if and only if $a_{r_1(2k-p)} \neq 0$, $k = 0, 1, ..., p$. If n is even and r_1 is odd, $a_{r_1(2k-p)} \neq 0$, $k = 0, 1, ..., p$, because $r_1(2k-p)$ is odd. If n is an even multiple of r_1, $a_{r_1(2k-p)} \neq 0$, for $k = 0, 1, ..., p$, because $2k - p$ is odd. Sufficiency of the condition in the Proposition is thus proved; its necessity follows from the fact that in all the remaining cases, there exists at least one triple (p, r_1, k) such that $a_{r_1(2k-p)} = 0$. This is so because if n is odd $a_{r_1(2k-p)} = 0$ for $k = p = n$, and if r_1 is an even multiple of n, $a_{r_1(2k-p)}$ is always 0.

Proof of Lemma 4.4.2 *(i)* For $h_1 = 1, 2, ..., \sum_{i=1}^{d}(n_i - 1)^2$, $E(\widetilde{q}_{h_1}^*) = 0$ because $E(\widetilde{q}_{h_1}^*) = tr[\widetilde{A}_{h_1}]$ and no $\alpha \in \Gamma(n_1, ..., n_d)$ can be an h_1-neighbor of itself. The variance is $var(\widetilde{q}_{h_1}^*) = 2tr[\widetilde{A}_{h_1}^2] = 2\widetilde{m}_h N = 4\widetilde{N}_h$.

(ii) The covariance between two distinct quadratic forms $\widetilde{q}_{h_1}^*$ and $\widetilde{q}_{h_2}^*$ is $2tr[\widetilde{A}_{h_1}\widetilde{A}_{h_2}]$. Now, the diagonal element of $\widetilde{A}_{h_1}\widetilde{A}_{h_2}$ indexed by $\gamma \in \Gamma$ is $\sum_{\delta\in\Gamma}(\widetilde{A}_{h_1})_{\gamma,\delta}(\widetilde{A}_{h_2})_{\delta,\gamma}$, and thus

$cov(\widetilde{q}^*_{h_1}, \widetilde{q}^*_{h_2}) = 0$ unless there is at least one pair γ, δ and two (different) subsets D_1 and D_2 of Δ such that $\|\alpha - \epsilon_{D_1}\|^2 = h_1$ and $\|\alpha - \epsilon_{D_2}\|^2 = h_2$, where $\alpha(i) = |\gamma(i) - \delta(i)|$, $i = 1, ..., d$. That is, two distinct quadratic forms $\widetilde{q}^*_{h_1}$ and $\widetilde{q}^*_{h_2}$ are correlated if and only if there exists an $\alpha \in \Gamma$ at distance h_1 from one vertex of $\Gamma(n_1 + 1, ..., n_d + 1)$ and at distance h_2 from another. It is easily deduced that a necessary condition for $\widetilde{q}^*_{h_1}$ and $\widetilde{q}^*_{h_2}$ to be correlated is that there exist $\alpha \in \Gamma_{h_1}$ and $\beta \in \Gamma_{h_2}$ such that, for at least one $i = 1, ..., d$, $\alpha(i) + \beta(i) = n_i$. As a consequence, $\widetilde{q}^*_{h_1}$ and $\widetilde{q}^*_{h_2}$ are uncorrelated if no $\alpha \in \Gamma_{h_1} \cup \Gamma_{h_2}$ contains an element $\alpha(i) \geq n_i/2$, which completes the proof of the lemma.

Proof of Lemma 4.4.3 From (4.13), $\kappa_p(q^*_h) = 2^{p-1}(p-1)!tr[A^p_h]$ and $\kappa_p(\widetilde{q}^*_h) = 2^{p-1}(p-1)!tr[\widetilde{A}^p_h]$. The first part of the lemma is then trivial and follows from the fact that $\widetilde{A}_h$ and A_h are nonnegative matrices and $\widetilde{A}_h = A_h + R_h$, with R_h a nonnegative matrix. The case $p = 1$ is also trivial, since in that case $E(\widetilde{q}^*_h)$ and $E(q^*_h)$ are both zero. Next, observe that the (γ, γ) entry of A^p_h is

$$(A^p_h)_{\gamma,\gamma} = \sum_{\delta_1,...,\delta_{p-1} \in \Gamma} \{(A_h)_{\gamma,\delta_1}(A_h)_{\delta_1,\delta_2}...(A_h)_{\delta_{p-2},\delta_{p-1}}(A_h)_{\delta_{p-1},\gamma}\}, \tag{4.19}$$

which is non-zero if and only if there is at least one $(p-1)$-tuple $(\delta_1, ..., \delta_{p-1})$ such that

$$\|\gamma - \delta_1\|^2 = \|\delta_1 - \delta_2\|^2 =, ..., = \|\delta_{p-2} - \delta_{p-1}\|^2 = \|\delta_{p-1} - \gamma\|^2 = h. \tag{4.20}$$

Similarly,

$$(\widetilde{A}^p_h)_{\gamma,\gamma} = \sum_{\delta_1,...,\delta_{p-1} \in \Gamma} \{(\widetilde{A}_h)_{\gamma,\delta_1}(\widetilde{A}_h)_{\delta_1,\delta_2}...(\widetilde{A}_h)_{\delta_{p-2},\delta_{p-1}}(\widetilde{A}_h)_{\delta_{p-1},\gamma}\}, \tag{4.21}$$

which, in view of Proposition 4.3.1, is non-zero if and only if there is at least one $(p-1)$-tuple $(\delta_1, ..., \delta_{p-1})$ such that

$$\exists D_j \in \Delta : \left\|\alpha_j - \epsilon_{D_j}\right\|^2 = h, \text{ for } j = 1, ..., p, \tag{4.22}$$

where the α_j's are defined by $\alpha_j(i) = |\delta_{j-1}(i) - \delta_j(i)|$, $i = 1, ..., d$, with $\delta_0 = \delta_p = \gamma$. Call a distance h feasible if Γ contains at least one pair of h-neighbors; then, for even p, $k_p(\widetilde{q}^*_h) = 0$ is necessary and sufficient for $k_p(q^*_h) = 0$ simply because if h is not feasible then both $(A^p_h)_{\gamma,\gamma}$ and $(\widetilde{A}^p_h)_{\gamma,\gamma}$ are zero for any $\gamma \in \Gamma$, while if h is feasible then both $\kappa_p(\widetilde{q}^*_h)$ and $\kappa_p(q^*_h)$ are positive (because for any pair of h-neighbors (γ, δ), the $(p-1)$-tuple $(\delta, \gamma, \delta, ..., \gamma, \delta)$ satisfies both (4.20) and (4.22)). To prove the part of the Lemma relative to Condition C, we show that, under Condition C, $k_p(\widetilde{q}^*_h) \neq 0$ implies $k_p(q^*_h) \neq 0$. First, observe that $\widetilde{A}^p_h(\gamma, \gamma)$ does not depend on γ, as is easily seen by considering formula (4.4) plus the fact that the product

of any two $F_r^{(n)}$ matrices has constant diagonal. If $k_p(\widetilde{q}_h^*) \neq 0$, then for each $\gamma \in \Gamma$, there is at least one $(p-1)$-tuple $(\delta_1, ..., \delta_{p-1})$ satisfying (4.22). This clearly implies that, under Condition C, it is always possible to find a p-tuple $(\gamma, \delta_1, ..., \delta_{p-1})$ such that (4.22) is satisfied with $D_j = \{0, ..., 0\}$, for $j = 1, ..., p$. But this in turn implies that the same p-tuple satisfies (4.20), and hence that $k_p(q_h^*) \neq 0$, as was to be proved. To complete the proof, we need to show that when $p \geq 3$ and without Condition C, $k_p(q_h^*)$ may be zero even if $k_p(\widetilde{q}_h^*)$ is not. This is so because even if there is not a γ and a $(p-1)$-tuple $(\delta_1, ..., \delta_{p-1})$ that satisfy (4.20), it may still be possible, by taking at least one of the D_j's to be different from $\{0, ..., 0\}$, to find a γ and a $(p-1)$-tuple that satisfy (4.22). For example, for $d = 2$ and $h = 20$, the sequences $\delta_1 = (4, 2)$ and $\delta_2 = (2, 6)$ satisfy (4.22) when $\gamma = (0, 0)$, and thus $k_3(\widetilde{q}_{20}^*) > 0$, although $k_3(q_{20}^*)$ must be 0, since, as it is well-known (e.g., Beeson, 1992), there is no equilateral triangle with vertices on a 2-dimensional planar lattice.

Proof of Proposition 4.4.4 Under Condition B, we have from the generating function (4.11) that

$$\widetilde{m}_h = \sum_{\alpha \in \Gamma_h} 2^{d - k_\alpha(0)}.$$

On the other hand, from (4.12) we obtain

$$\begin{aligned} 2N_h &= \sum_{\alpha \in \Gamma_h} \Pi_{i=1}^d \left\{ \left(2 - \delta_{\alpha(i),0}\right) (n_i - \alpha(i)) \right\} \\ &= \sum_{\alpha \in \Gamma_h} \left\{ 2^{d - k_\alpha(0)} \Pi_{i=1}^d (n_i - \alpha(i)) \right\}. \end{aligned}$$

The expression in the proposition follows simply by taking $2\widetilde{N}_h = N\widetilde{m}_h$.

Proof of Corollary 4.4.5 Assume that $n_1 = ... = n_d = n$, and, for any $\omega \in \Omega_h$, let $v(\omega) = \prod_{j=0}^{n-1} k_\omega(j)!$, where $k_\omega(j)$ denotes the multiplicity of j in ω. Then, the numerator in (4.14) is $n^d \sum_{\omega \in \Omega_h} \{d! 2^{-\omega(0)} / \nu(\omega)\}$, and the denominator is $\sum_{\omega \in \Omega_h} \{d! 2^{-\omega(0)} \Pi_{i=1}^d (n - \omega(i)) / \nu(\omega)\}$. The corollary follows straightforwardly on assuming $\Omega_h = \{\omega_h\}$.

Proof of Corollary 4.4.6 For $h = 1, 2, 3$, $|\Omega_h| = 1$ for any dimension $d \geq h$, because any decomposition in d squares of h consists of h ones and $d - h$ zeros. The corollary obtains by applying Corollary 4.4.5.

Proof of Corollary 4.4.7 If only non-diagonal directions are considered, $k_\alpha(0) = d - 1$, for any $\alpha \in \Gamma_h$. Then, the expression in the corollary follows from (4.14), on noting that the only non-zero element of $\alpha \in \Gamma_h$ must be $\sqrt{h}$.

Proof of Proposition 4.4.8 From expressions (4.11) and (4.12) it is clear that $2\widetilde{N}_h = N\widetilde{m}_h$ and $2N_h$ are symmetric polynomials in $n_1, ..., n_d$. Then the rational function $\eta_{2,h} =$

$\widetilde{N}_h/N_h$ admits a basis in terms of the elementary symmetric functions of $n_1, ..., n_d$ (e.g., Macdonald, 1995). Expressed in such a basis, $\eta_{2,h}$ can be differentiated easily with respect to $n_1, ..., n_d$, keeping N fixed. Imposing that the Jacobian is zero and verifying that the obtained solution is a minimum completes the proof of the lemma.

Proof of Lemma 4.4.9 Clearly, $\kappa_p(\widetilde{q}_h^*) \geq 0$ and $\kappa_p(q_h^*) \geq 0$, because $\widetilde{A}_h$ and A_h are nonnegative matrices, and $\kappa_p(\widetilde{q}_h^*) \geq \kappa_p(q_h^*)$, because $\widetilde{A}_h = A_h + R_h$, with R_h a nonnegative matrix. We have to show that when $\kappa_p(\widetilde{q}_h^*) > 0$, $\kappa_p(\widetilde{q}_h^*)$ is strictly greater than $\kappa_p(q_h^*)$. This amounts to proving that (a) $(\widetilde{A}_h^p)_{\gamma,\gamma} \geq (A_h^p)_{\gamma,\gamma}$, for each $\gamma \in \Gamma$, and (b) there is at least one $\gamma \in \Gamma$ such that $(\widetilde{A}_h^p)_{\gamma,\gamma} > (A_h^p)_{\gamma,\gamma}$ (when $\kappa_p(\widetilde{q}_h^*) > 0$). Assertion (a) follows from the fact that if a $(p-1)$-tuple $(\delta_1, ..., \delta_{p-1})$ satisfies expression (4.20) given in the proof of Lemma 4.4.3 for a given γ, then it also satisfies expression (4.22) given in the same proof for the same γ, with each D_j equal to $\{0, ..., 0\}$. A sequence γ always satisfying (b) is the origin $(0, ..., 0)$; to see why this is so, let us denote by α^D the sequence with elements $\alpha^D(i) = |\epsilon_D - \alpha(i)|$, for a sequence $\alpha \in \Gamma$ and a set $D \in \Delta$. Then, it is clear that if $\gamma = (0, ..., 0)$, then, for each $(\delta_1, ..., \delta_{p-1})$ satisfying (4.20), any of the 2^d tuples $(\delta_1^D, ..., \delta_{p-1}^D)$, for a $D \in \Delta$, satisfies the conditions (4.22), with each D_j equal to D. This completes the proof of the lemma.

Proof of Lemma 4.4.10 The lemma is a straightforward consequence of expressions (4.19)-(4.22) given in the proof of Lemma 4.4.3.

Proof of Lemma 4.4.12 Same as the previous proof.

Chapter 5

Concluding Remarks

This thesis represents our contribution to the econometric and statistical literature on parametric spatial processes defined on regular or irregular lattices. In this final chapter, we summarize our main findings and indicate directions for further work.

Chapter 1 has provided a novel interpretation of the correlation structure of spatial autoregressive models. The interpretation, based on some formal graph theory, explains several known peculiarities of the correlations implied by the models when they are constructed over irregular lattices. It has also allowed to prove new results regarding the second-order properties of the models. In addition, the adoption of a graph theoretic perspective has enabled us to clarify the role played by the presence of symmetries and regularities in the set of observational units.

The most natural extension of Chapter 1 is to use the graph theoretic framework that has been developed there to study the properties of inferential procedures in the context of spatial autoregressive models. Recently, a graph theoretic approach which has some connections to ours has been used in Martin (2005a and 2005b) to devise a solution to some computational problems arising when estimating CAR and SAR models by maximum likelihood.

Chapter 2 has investigated exact power properties of invariant tests for spatial autocorrelation in the context of the linear regression model. We have studied how the power is affected by the matrix of regressor and by the assumed spatial structure. Some relevant results of the paper are: we have given conditions for unbiasedness and monotonicity of the power function of the tests; we have shown that, for any assumed spatial structure, a matrix of regressor exists such that the limiting power of locally best tests and point optimal tests vanishes; we have argued that there is a fundamental difference between the case of symmetric and non-symmetric weights matrices as far as the power of the tests is concerned.

Chapter 2 is prone to several extensions, many of which have already been indicated. It seems possible that the results of the paper can be adapted to the context of a time series first-order autoregressive (AR(1)) model, in particular to investigate how the power of invariant tests of unit roots depends on the regressors. Such an adaptation should be based on approximating the covariance matrix of an AR(1) model with a matrix belonging to the span of I and another fixed matrix, as done in the seminal work by Anderson (1948). Additionally, the possibility of extending the framework developed in the paper to problems in estimation based on the density of the maximal invariant (i.e., estimation based on the marginal likelihood) deserves attention in further work.

Chapter 3 has provided a complete structural representation of the spatial design matrices of isotropic processes on uniform grids, and has given generating functions which make the computation of the matrices straightforward. The generating functions for the spatial design matrices also induce generating functions for the cumulant of the quadratic forms associated to such matrices. A particular consequence of these results is that the variance of the classical variogram estimator can be computed exactly. This has been done explicitly in the paper for some popular assumptions on the actual variogram.

One question that provides an interesting connection between Chapter 2 and Chapter 3, and that arises quite naturally when studying the properties of the statistics q_h and q_h^*, is which of the two statistics is better, in some sense to be specified, to test for spatial autocorrelation in the context of a linear regression model. Recall that $\bar{q}_h = q_h/(z'z)$ is the generalization of a von Neumann statistic, while $\bar{q}_h^* = q_h^*/(z'z)$ is the generalization of a serial correlation coefficient. Curiously, the dominant test for serial correlation in regression errors—the Durbin-Watson test—is based on $\bar{q}_h$, whereas the dominant test in higher dimension ($d > 1$)—the Cliff-Ord test—is based on $\bar{q}_h^*$. The question referred to above is, in more precise terms, for which class of models, and which values of their parameters, tests based on q_h are more (or less) powerful than tests based on q_h^*.

Chapter 4 has developed a circular approximation to the spatial design matrices. We have shown that the approximation is useful to obtain the cumulants of the quadratic forms associated to the spatial design matrices, in that it reduces dramatically the required computational effort.

An important extension of Chapter 4 would be to perform a large numerical study of the performance of the circular approximation for both q_h and q_h^*, and for a variety of isotropic spatial processes on grids of dimension 2 and 3 and with popular covariance functions or

variograms. While—given the results in Chapter 4—it is known that the approximation improves as each side n_i of the grid increases, it would certainly be useful for practitioners to have an idea of the magnitude of the errors involved when, for instance, approximating the variance of the classical variogram estimator, for a given spatial process on a given grid.

References

Ali, M.M., 1987. Durbin-Watson and generalized Durbin-Watson tests for autocorrelations and randomness, Journal of Business and Economic Statistics 5, 195-203.

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

Anderson, T.W., 1948. On the theory of testing serial correlation, Skandinavisk Aktuarietidskrift 31, 88-116.

Anderson, T.W., 1955. The integral of a symmetric unimodal function, Proceedings of the American Mathematical Society 6, 170-176.

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

Anselin, L., Bera, A., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics, in: A. Ullah, D.E.A., Giles (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, New York, pp. 237-289.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers: The Netherlands.

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

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

Bartels, R., 1992. On the power function of the Durbin-Watson test, Journal of Econometrics 51, 101-112.

Bartlett, M.S., 1978. Nearest neighbour models in the analysis of field experiments, Journal of the Royal Statistical Society B 40, 147-174.

Beeson, M.J., 1992. Triangles with vertices on lattice points, American Mathematical Monthly 99, 243-252.

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., 1975. Statistical analysis of non-lattice data, The Statistician 24, 179-195.

Besag, J.E., 1981. On a system of two-dimensional recurrence equations, Journal of the Royal Statistical Society B 43, 302-309.

Besag, J.E., Moran, P.A.P., 1975. On the estimation and testing of spatial interaction in Gaussian lattice processes, Biometrika 62, 555-562.

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

Besag, J.E., York, J.C. and Mollié, A., 1991. Bayesian image restoration, with two applications in spatial statistics, Annals of the Institute of Statistical Mathematics 43, 1–20.

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.

Brook, D., 1964. On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbor systems, Biometrika 51, 481-483.

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., 1972. Testing for spatial autocorrelation among regression residuals, Geographical Analysis 4, 267-84.

Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. London: Pion.

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

Conley, T.G., 1999. GMM estimation with cross sectional dependence, Journal of Econometrics 92, 1-45.

Conley, T.G., Topa, G., 2002. Socio-economic distance and spatial patterns in unemployment, Journal of Applied Econometrics 17, 303-327.

Cox, D.R.. Hinkley, D.V., 1974. Theoretical Statistics, Chapman & Hall, London.

Cressie, N., 1985. Fitting varogram models by weighted least squares, Mathematical Geology 17, 563-586.

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

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

Diggle, P.J., 1988. An approach to the analysis of repeated measurements, Biometrics 44, 959-971.

Durbin, J., 1970. An alternative to the bounds test for testing for serial correlation in least-squares regression, Econometrica 38, 422-29.

Durbin, J., Watson, G.S., 1951. Testing for serial correlation in least squares regression II, Biometrika 38, 159-178.

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

Florax, R.J.G.M., de Graaff, T., 2004. The performance of diagnostic tests for spatial dependence in linear regression models: a meta-analysis of simulation studies, in: L. Anselin, R.J.G.M. Florax, S. J. Rey (Eds.), Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer, Heidelberg, pp. 29-65.

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

Geary, R.C., 1954. The contiguity ratio and statistical mapping, The Incorporated Statistician 5, 115-145.

Genton, M.G., 1998. Variogram fitting by generalized least squares using an explicit formula for the covariance structure, Mathematical Geology 30, 323-345.

Genton, M.G., 1999. The correlation structure of the sample autocovariance function for a particular class of time series with elliptically contoured distribution, Statistics and Probability Letters 41, 131-137.

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

Golub, G.H., and Van Loan, C.F., 1996. Matrix Computations, third edition. The Johns Hopkins University Press, London.

Gorsich, D.J., Genton, M.G., Strang, G., 2002. Eigenstructures of spatial design matrices, Journal of Multivariate Analysis 80, 138-165.

Hardy, G.H., Littlewood, J.E., Pólya, G., 1952. Inequalities, second edition. Cambridge University Press, Cambridge.

Hardy, G.H., Wright, E.M., 1979. An Introduction to the Theory of Numbers, fifth edition. Oxford University Press, Oxford.

Henshaw, R.C., 1996. Testing single-equation least squares regression models for autocorrelated disturbances, Econometrica 34, 646-660.

Hillier, G.H., 1987. Classes of similar regions and their power properties for some econometric testing problems, Econometric Theory 3, 1– 44.

Hillier, G.H., 2001. The density of a quadratic form in a vector uniformly distributed on the n-sphere, Econometric Theory, 17, 1-28.

Hillier, G.H., Martellosio, F., 2006. Spatial design matrices and associated quadratic forms: structure and properties, Journal of Multivariate Analysis, 97, 1-18.

Horn, R., Johnson, C.R., 1985. Matrix Analysis. Cambridge University Press, Cambridge.

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

Imhof, J.P., 1961. Computing the distribution of quadratic forms in normal variables, Biometrika 48, 419-26.

James, A.T., 1954. Normal multivariate analysis and the orthogonal group, Annals of Mathematical Statistics, 25, 40-75.

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

James, A.T., 1964. Distributions of matrix variates and latent roots derived from normal samples, Annals of Mathematical Statistics 35, 475-501.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, London.

Kadiyala, K.R., 1970. Testing for the independence of regression disturbances, Econometrica 38, 97-117.

Kariya, T., 1977. A robustness property of the tests for serial correlation, Annals of Statistics 5, 1212-1220.

Kariya, T., 1980. Locally robust tests for serial correlation in least squares regression, Annals of Statistics 8, 1065-1070.

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., 2001. On the asymptotic distribution of the Moran I test statistic with applications, Journal of Econometrics 104, 219–257.

Kendall, M.G., Stuart, A., 1969. The Advanced Theory of Statistics, Vol. 1, Distribution Theory, Griffin and Co., London.

King, M.L., 1980. Robust tests for spherical symmetry and their application to least squares regression, Annals of Statistics 8, 1265-1271.

King, M.L., 1981. A small sample property of the Cliff-Ord test for spatial autocorrelation, Journal of the Royal Statistical Society B 43, 263 - 4.

King, M.L., 1988. Towards a theory of point optimal testing, Econometric Reviews 6, 169–255.

King, M.L. and Hillier, G.H, 1985. Locally best invariant tests of the error covariance matrix of the linear regression model, Journal of the Royal Statistical Society B 47, 98-102.

Kleiber, C., Krämer, W., 2005, Finite-sample power of the Durbin-Watson test against fractionally integrated disturbances, Econometrics Journal 8, 406-417.

Krämer, W., 1985. The power of the Durbin-Watson test for regressions without an intercept, Journal of Econometrics 28, 363-370.

Krämer, W., 2005. Finite sample power of Cliff–Ord-type tests for spatial disturbance correlation in linear regression, Journal of Statistical Planning and Inference 128, 489-496.

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

Lehmann, E.L. and Romano, J., 2005. Testing Statistical Hypotheses, third edition. Springer, New York.

Li, Z., Hall. F., Zhang, F., 1997. Sign patterns of nonnegative normal matrices, Linear Algebra and Its Applications 254, 335-354.

Macdonald, I.G., 1995. Symmetric Functions and Hall Polynomials, second edition. Oxford University Press.

Marcus, M., Minc, H., 1969. A Survey of Matrix Theory and Matrix Inequalities. Dover, New York.

Martellosio, F., 2005. The correlation structure of spatial autoregressions, manuscript, University of Southampton.

Martin, R.J., 1986. A note on the asymptotic eigenvalues and eigenvectorsof the dispersion matrix of a second-order stationary process on a d-dimensional lattice, Journal of Applied Probability 23, 529-535.

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

Martin, R.J., 2005a. Some results on graph spectra, with applications to geographic spatial modelling, preprint available at www.maths.otago.ac.nz/?martin.

Martin, R.J., 2005b. Some fast methods for fitting some one-parameter spatial models, preprint available at www.maths.otago.ac.nz/?martin.

Mohar, B., 1997. Some applications of Laplace eigenvalues of graphs, in: G. Hahn and G. Sabidussi (Eds.), Graph Symmetry: Algebraic Methods and Applications, vol. 497 of NATO ASI Series C, Kluwer, Dordrecht, pp. 227-275.

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

Ord, J.K., 1975. Estimation methods for models of spatial interaction, Journal of the American Statistical Association 70, 120-6.

Pace, R.K, LeSage, J.P., 2002. Semiparametric maximum likelihood estimates of spatial dependence, Geographical Analysis 34, 76-90.

Paelinck, J., Klaassen, L., 1979. Spatial Econometrics. Saxon house, Farnborough.

Rahman, S., King, M.L., 1997. Marginal-likelihood score-based tests of regression disturbances in the presence of nuisance parameters, Journal of Econometrics 82, 81-106.

Rao, C.R., 1973. Linear Statistical Inference and Its Applications, second edition. Wiley, New York.

Ross, I.C., Harary, F., 1952. On the determination of redundancies in sociometric chains, Psychometrika 17, 195-209.

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.

Tiefelsdorf, M., Boots, B. 1995. The exact distribution of Moran's I, Environment and Planning A 27, 985-999.

Tillman, J.A., 1975. The power of the Durbin-Watson test, Econometrica 43, 959-74.

Tunnicliffe Wilson, G., 1989. On the use of marginal likelihood in time series model estimation, Journal of the Royal Statistical Society B 51, 15-27.

von Neumann, J., Kent, R.H., Bellinson, H.R., Hart, B.I., 1941. The mean square successive differences, Annals of Mathematical Statistics 12, 153-162.

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.

Wilf, H.S., 1994. generatingfunctionology, second edition, Academic Press Inc.

Zeisel, H., 1989. On the power of the Durbin-Watson test under high autocorrelation, Communications in Statistics: Theory and Methods 18, 3907-3916.