*Econometric Theory*, **26**, 2010, 152–186.
doi:10.1017/S0266466609090641

# POWER PROPERTIES OF INVARIANT TESTS FOR SPATIAL AUTOCORRELATION IN LINEAR REGRESSION

FEDERICO MARTELLOSIO
*University of Reading*

This paper derives some exact power properties of tests for spatial autocorrelation in the context of a linear regression model. In particular, we characterize the circumstances in which the power vanishes as the autocorrelation increases, thus extending the work of Krämer (2005). More generally, the analysis in the paper sheds new light on how the power of tests for spatial autocorrelation is affected by the matrix of regressors and by the spatial structure. We mainly focus on the problem of residual spatial autocorrelation, in which case it is appropriate to restrict attention to the class of invariant tests, but we also consider the case when the autocorrelation is due to the presence of a spatially lagged dependent variable among the regressors. A numerical study aimed at assessing the practical relevance of the theoretical results is included.

## 1. INTRODUCTION

Testing for 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. This paper analyzes exact power properties of tests used for this purpose. We mainly focus on the case of regression models with errors following a first-order simultaneous autoregressive (SAR(1)) process, but we also consider models containing a spatially lagged dependent variable among the regressors. The former models are often referred to as spatial error models, the latter as spatial lag models. For empirical applications of such models in economics, see, for instance, Case (1991), Pinske and Slade (1998), and Bell and Bockstael (2000).

So far, power properties of tests for residual spatial autocorrelation have received much less attention than the power properties of tests for residual serial

I am grateful to the co-editor and two anonymous referees for suggestions that have led to substantial improvements in the paper. I thank Giovanni Forchini, Grant Hillier, Pierre Perron, Benedikt Pötscher, Tony Smith, and Kees Jan van Garderen for helpful comments, and Anurag Banerjee for providing GAUSS routines for the Imhof's procedure. This paper is based on a chapter from my Ph.D. dissertation at the University of Southampton. Address correspondence to Federico Martellosio, School of Economics, University of Reading, URS Building, Whiteknights PO Box 219, Reading RG6 6AW, UK; e-mail: f.martellosio@reading.ac.uk.

 © Cambridge University Press, 2009 0266-4666/10 $15.00

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

autocorrelation, 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, notably King (1981) and Krämer (2005). The former paper establishes that the most popular test for residual spatial autocorrelation, the Cliff-Ord test, is locally best invariant in the case of a Gaussian spatial error model. The latter paper identifies a possible problem with tests of autocorrelation in the context of a Gaussian spatial error model with symmetric weights matrix, generalizing results available for tests of serial autocorrelation (see Krämer, 1985; Zeisel, 1989). In particular, Krämer (2005) considers tests whose associated test statistics can be expressed as ratios of quadratic forms in the regression errors, and shows that there are cases when the power vanishes as the residual autocorrelation increases.

The present paper aims to extend the results in Krämer (2005). We show that the vanishing limiting power problem is not confined to a particular class of tests. The problem arises because a SAR(1) model tends, as the autocorrelation parameter goes to the right boundary of the parameter space, to a family of (improper) distributions supported on a one-dimensional subspace of the sample space. The limiting power disappears whenever the intersection between such a subspace and the critical region has zero one-dimensional Lebesgue measure. In the context of a spatial error model, it is natural to focus on invariant tests (e.g., Lehmann and Romano, 2005). We formulate conditions for the limiting power of any given invariant test to be 0, 1, or in $(0, 1)$. Such conditions require neither Gaussianity nor symmetry of the weights matrix. Allowing for nonsymmetric weights matrices is important, especially because in applications weights matrices are often row-standardized. It turns out that when the weights matrix is row-standardized and the regression contains an intercept, the limiting power of any invariant test for residual spatial autocorrelation is in $(0, 1)$. On the contrary, when the weights matrix is not row-standardized, the limiting power of an invariant test is generally either 0 or 1. An explanation of why some economic phenomena may be better described by means of a non–row-standardized weights matrix is contained in Kelejian and Prucha (2009).

Krämer's results and our extensions are particularly relevant for empirical applications where a dependent variable is highly spatially autocorrelated and the autocorrelation cannot be explained by conditioning on observable factors. For example, this may occur in studies of the term structure of interest rates, where pricing errors are likely to be strongly autocorrelated according to their distance in terms of maturity (e.g., Kennedy, 1994; Goldstein, 2000). In this context, Huse (2006) estimates a spatial error model and finds a very large value of the autocorrelation parameter. Similarly, Gall, Pap, and van Zuijlen (2004) use a spatial autoregression to account for the autocorrelation in the maturity space of forward interest rates, and discuss explicitly the case when the autocorrelation parameter tends to the right boundary of the parameter space. Two examples of economic applications where non-observable factors may induce high residual autocorrelation in the geographic space are hedonic market models for dwelling selling prices

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

(see Militino, Ugarte, and García-Reinaldos 2004), and the analysis of returns for a set of firms within an industry.[1]

To complement the limiting power analysis, we discuss some conditions that are sufficient for unbiasedness of invariant tests for residual spatial autocorrelation and for monotonicity of their power functions. Although such conditions are not necessary, they provide insights into the role played by the regressors and the spatial structure in determining the power of the tests, and they help to understand the causes of undesirable properties of the tests.

The rest of the paper is organized as follows. Section 2 presents the theoretical framework. Our main results are in Section 3. In that section, we first consider in some detail the limiting power of tests for spatial autocorrelation in a spatial error model. Then, we discuss how the results change in the case of some other spatial models, including a spatial lag model and a spatial moving average model. The special case of pure SAR(1) processes is considered separately. To assess the practical relevance of the limiting power analysis, Section 3 reports results from a small numerical study. Section 4 analyzes the conditions for unbiasedness of the tests and monotonicity of their power functions. Section 5 concludes. The Appendices contain some technical material and all proofs.

## 2. FRAMEWORK

Section 2 presents the set-up in which our results will be derived. Section 2.1 defines the testing problem we are concerned with, and Section 2.2 introduces invariant tests for that problem.

### 2.1. The Testing Problem

Consider a fixed and finite set of $n$ observational units, such as the regions of a country, and let $\boldsymbol{y} = (y_1, \ldots, y_n)'$, where $y_i$ denotes the random variable observed at the $i$-th unit. The ordering of the units is arbitrary. We assume that $\boldsymbol{y}$ follows a linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \mathrm{E}(\boldsymbol{u}) = \boldsymbol{0}, \quad \mathrm{var}(\boldsymbol{u}) = \sigma^2 \boldsymbol{\Sigma}(\rho), \tag{1}$$

where $\boldsymbol{X}$ is an $n \times k$ matrix of rank $k < n$, $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters, $\sigma^2 > 0$ is an unknown parameter, and $\rho$ is an unknown parameter belonging to some connected open subset $\Psi$ of $\{\rho : \boldsymbol{\Sigma}(\rho) \text{ is positive definite}\}$. The matrix $\boldsymbol{X}$ contains only exogenous variables; either it is nonstochastic or all the analysis is interpreted as conditional on $\boldsymbol{X}$. As for the distribution of the error term $\boldsymbol{u}$, we only assume that the density of $\boldsymbol{u}$ is positive everywhere on $\mathbb{R}^n$, is larger at $\boldsymbol{0}$ than anywhere else, and is continuous in both $\boldsymbol{y}$ and the parameters $\sigma^2$ and $\rho$.

In the context of model (1), we are interested in testing

$$\mathrm{H}_0 : \rho = 0 \quad \text{versus} \quad \mathrm{H}_a : \rho > 0. \tag{2}$$

Here and throughout, $\rho > 0$ is to be understood as $\rho \in \mathbb{R}^+ \cap \Psi =: \Psi^+$, that is, we leave it implicit that $\rho$ must belong to the parameter space of the model. The

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

choice of a one-sided alternative rather than a two-sided one is dictated by the fact that the former is empirically more relevant for many specifications of $\boldsymbol{\Sigma}(\rho)$.

Throughout the paper we will be mainly concerned with the covariance structure $\boldsymbol{\Sigma}(\rho)$ implied by a first-order *simultaneous autoregressive* (SAR(1)) process (e.g., Whittle, 1954; Cliff and Ord, 1981; Anselin, 1988; Cressie, 1993). Such a process is specified on the basis of a fixed $n \times n$ (spatial) weights matrix $\boldsymbol{W}$, chosen to reflect a priori information on relations among the $n$ observations. Typically, for each $i, j = 1, \ldots, n$, $(\boldsymbol{W})_{ij} = 0$ if $i$ and $j$ are not neighbors according to some metric deemed to be relevant for the phenomenon under analysis, whereas $(\boldsymbol{W})_{ij}$ is set to some nonzero number, possibly reflecting the degree of interaction, otherwise. For instance, when the observational units are the regions of a country, one may set $(\boldsymbol{W})_{ij} = 1$ if two distinct regions $i$ and $j$ share a common boundary, $(\boldsymbol{W})_{ij} = 0$ otherwise. In this paper we assume that a weights matrix (i) has zero entries along its main diagonal, (ii) is entrywise nonnegative, and (iii) is irreducible. Details concerning such assumptions are in Appendix A.

A SAR(1) process for the error vector $\boldsymbol{u}$ is specified by

$$\boldsymbol{u} = \rho \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}, \quad \operatorname{var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{V}, \tag{3}$$

where $\boldsymbol{\varepsilon}$ is a vector of innovations and $\boldsymbol{V}$ is a fixed $n \times n$ symmetric and positive definite matrix. The extension to the case when $\boldsymbol{V}$ depends on unknown parameters will be discussed in Section 3.5.1. Let $\boldsymbol{I}$, or $\boldsymbol{I}_n$, denote the $n \times n$ identity matrix. For testing problem (2), there is no loss of generality in assuming that $\boldsymbol{\Sigma}(0) = \boldsymbol{I}$ (if $\boldsymbol{\Sigma}(0) \neq \boldsymbol{I}$, just premultiply $\boldsymbol{y}$ by $\boldsymbol{\Sigma}^{-1/2}(0)$). Hence, we can take $\boldsymbol{V} = \boldsymbol{I}$.

Provided that $\rho$ is different from the reciprocal of the nonzero real eigenvalues of $\boldsymbol{W}$, equation (3) implies

$$\boldsymbol{\Sigma}(\rho) = \left[(\boldsymbol{I} - \rho \boldsymbol{W}')(\boldsymbol{I} - \rho \boldsymbol{W})\right]^{-1}. \tag{4}$$

For SAR(1) processes, we take $\Psi^+ = (0, \lambda_{\max}^{-1})$, where $\lambda_{\max}$ is the largest positive eigenvalue of $\boldsymbol{W}$. While the condition $\rho < \lambda_{\max}^{-1}$ is not necessary for positive definiteness of (4), it guarantees connectedness of $\Psi^+$. In addition, the alternative hypothesis $\rho \in (0, \lambda_{\max}^{-1})$ represents positive autocorrelation,[2] a much more common phenomenon in practice than negative spatial autocorrelation.

The regression model (1) with disturbances following process (3) is often referred to as a *spatial error model*. There are two important alternatives to a spatial error model: the so-called *spatial lag model*, and the regression model with disturbances following a first-order *conditional autoregressive* (CAR(1)) process. In a spatial lag model, the spatial autocorrelation is introduced by including a spatial lag $\boldsymbol{W}\boldsymbol{y}$ among the regressors. The problem of testing for this type of spatial autocorrelation is different from the testing problem described above, and will be considered separately in Section 3.5.2. A CAR(1) process is a Gaussian model with

$$\boldsymbol{\Sigma}(\rho) = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \boldsymbol{L}, \tag{5}$$

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

where $\boldsymbol{L}$ is a fixed $n \times n$ diagonal matrix such that $\boldsymbol{L}^{-1}\boldsymbol{W}$ is symmetric (see Besag, 1974). Recall that there is no loss of generality in setting $\boldsymbol{\Sigma}(0) = \boldsymbol{I}$, which, in (5), corresponds to $\boldsymbol{L} = \boldsymbol{I}$. Thus, $\boldsymbol{W}$ can be assumed to be symmetric in CAR(1) models (because it must be symmetric when $\boldsymbol{L} = \boldsymbol{I}$ in (5)). In SAR(1) models, on the contrary, we do need to allow for nonsymmetric $\boldsymbol{W}$'s.

For the sake of simplicity, the results in this paper will be stated only for SAR(1) processes and not for CAR(1) processes. Nevertheless, from the proofs it is straightforward to check that the results that hold for a SAR(1) process with a symmetric $\boldsymbol{W}$ (corresponding to $\boldsymbol{\Sigma}(\rho) = (\boldsymbol{I} - \rho\boldsymbol{W})^{-2}$) also hold for a process with $\boldsymbol{\Sigma}(\rho) = (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}$. Under Gaussianity, the latter process can be interpreted as a CAR(1) process; without Gaussianity, it can be interpreted as a general autocorrelation process (see, e.g., Anderson, 1948; Kadiyala, 1970; Kariya, 1980; King, 1980).

## 2.2. The Tests

For the testing problem defined above, it is natural to focus on invariant tests. These are now informally introduced; details on the theory of invariant tests are available in standard references such as Lehmann and Romano (2005). Often a testing problem does not change if any transformation in a certain group is applied to the sample space. In this case, according to the "principle of invariance," the test itself should be invariant under the same group of transformations, that is, it should be 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.

Testing problem (2) is invariant with respect to the group of transformations $\boldsymbol{y} \to \gamma\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\delta}$, with $\gamma \in \mathbb{R}\backslash\{0\}$ and $\boldsymbol{\delta} \in \mathbb{R}^k$ (sometimes the smaller group with $\gamma > 0$ is considered; see Appendix B). By an invariant test for testing problem (2) we mean a test that is invariant under that group. Let $\boldsymbol{C}$ be an $(n-k) \times n$ matrix such that $\boldsymbol{C}\boldsymbol{C}' = \boldsymbol{I}_{n-k}$ and $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{M}_{\boldsymbol{X}} := \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, and let $\|\cdot\|$ denote the Euclidean norm. Fix, without any loss of generality, an arbitrary $\bar{\imath} = 1, \ldots, n$. Then a maximal invariant under the above group is $\boldsymbol{v} := \operatorname{sgn}(y_{\bar{\imath}})\boldsymbol{C}\boldsymbol{y}/\|\boldsymbol{C}\boldsymbol{y}\|$, where $\operatorname{sgn}(y_{\bar{\imath}})$ denotes the sign of $y_{\bar{\imath}}$. In some cases, it is possible to derive a closed-form expression for the density of $\boldsymbol{v}$. For example, if the distribution of $\boldsymbol{u}$ is elliptically symmetric, then it can be shown that the density of $\boldsymbol{v}$, with respect to the normalized Haar measure on the hemisphere $\left\{\boldsymbol{s} \in \mathbb{R}^{n-k} : \|\boldsymbol{s}\| = 1, s_{\bar{\imath}} \geq 0\right\}$, is

$$f(\boldsymbol{v}; \rho) = 2\left|\boldsymbol{C}\boldsymbol{\Sigma}(\rho)\boldsymbol{C}'\right|^{-\frac{1}{2}}\left[\boldsymbol{v}'\left(\boldsymbol{C}\boldsymbol{\Sigma}(\rho)\boldsymbol{C}'\right)^{-1}\boldsymbol{v}\right]^{-\frac{n-k}{2}} \tag{6}$$

(see Kariya, 1980, eqn. (3.7)).

Besides the principle of invariance, there are at least two other reasons why, for our testing problem, it is appropriate to restrict attention to invariant tests. First, the distribution of any invariant test statistic for our testing problem is free of nuisance parameters. This is clearly seen by exploiting the standard result that

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

the distribution of the maximal invariant depends only on the parameter maximal invariant, which in our case is $\rho$. The absence of nuisance parameters means that invariant tests are similar and that the power function of any invariant test does not depend on $\boldsymbol{\beta}$ or $\sigma^2$. Second, expression (6) turns out to be proportional to the Gaussian marginal likelihood of $\rho$ (see Kalbfleisch and Sprott, 1970). Thus, at least under Gaussianity, using an invariant test for our testing problem is equivalent to drawing inference from the marginal rather than the full likelihood of the data. The marginal likelihood has often been found to provide a better basis for inference about $\rho$ than the full likelihood of model (1), especially when $k$ is large with respect to $n$; see, e.g., Tunnicliffe Wilson (1989) and Rahman and King (1997).

In general, despite the elimination of the nuisance parameters achieved by the imposition of invariance, no uniformly most powerful invariant (UMPI) test exists for testing problem (2), not even under Gaussianity (see, e.g., King and Hillier, 1985).[3] Some examples of invariant tests for testing problem (2) are now presented. Throughout the paper, the critical value and the size of a test are denoted by $c$ and $\alpha$, respectively. To avoid trivial cases and unless otherwise specified, $\alpha$ is assumed to be in $(0, 1)$. Note that in view of the similarity of invariant tests, if the distribution of $\boldsymbol{u}$ is fully specified (up to $\sigma^2$ and $\rho$), then the critical value $c$ corresponding to a given size can be obtained accurately by Monte Carlo or other numerical methods.[4] Often, however, critical values are derived from the asymptotic distribution of the test statistic. Asymptotic critical values may generate size distortions, but, on the other hand, are generally obtained from standard distributions and may result in a test that is more robust to different distributions of $\boldsymbol{u}$.

Let $\hat{\boldsymbol{u}}$ be the vector of OLS residuals. Simple tests for our testing problem are those that reject $H_0$ when

$$\frac{\hat{\boldsymbol{u}}'\boldsymbol{Q}\hat{\boldsymbol{u}}}{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}} > c \tag{7}$$

for some fixed matrix $\boldsymbol{Q}$. In particular, when $\boldsymbol{Q}$ equals a spatial weights matrix $\boldsymbol{W}$, we obtain the Cliff-Ord test (see Cliff and Ord, 1981; Kelejian and Prucha, 2001). In some circumstances, a test based on (7) has exact optimality properties. In particular, it is locally best invariant (LBI) if $\boldsymbol{u}$ has an elliptically symmetric distribution, and $\boldsymbol{Q} = d\boldsymbol{\Sigma}(\rho)/d\rho|_{\rho=0}$, for some differentiable $\boldsymbol{\Sigma}(\rho)$ (King and Hillier, 1985; Kariya, 1988). It follows that under the assumption of elliptical symmetry, the Cliff-Ord test is LBI when $\boldsymbol{\Sigma}(\rho)$ is that of a SAR(1) (or CAR(1)) process (see King, 1981).[5] When the regression contains only an intercept, the Cliff-Ord test reduces to the Moran test (Moran, 1950).

Other important invariant tests are the likelihood ratio (LR) test (based on the full density of the data) and its "restricted" version based on the density of $\boldsymbol{v}$.[6] We will also consider the tests that, for a fixed $\bar{\rho} > 0$ (with, of course, $\bar{\rho} \in \Psi^+$), reject $H_0$ when

$$\boldsymbol{v}'\left(\boldsymbol{C}\boldsymbol{\Sigma}(\bar{\rho})\boldsymbol{C}'\right)^{-1}\boldsymbol{v} < c. \tag{8}$$

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

If $\boldsymbol{u}$ has an elliptically symmetric distribution, a test based on (8) is point optimal invariant (POI), i.e., it is the most powerful invariant test against the specific alternative hypothesis $\rho = \bar{\rho} > 0$ (see King, 1988). POI tests define the power envelope of invariant tests. More precisely, denoting by $\pi_{\bar{\rho}}(\rho)$ the power of the POI critical region, the power envelope of size-$\alpha$ invariant tests is the function that associates the value $\pi_{\rho}(\rho)$ to each $\rho \geq 0$.

Before we continue, some notation is in order. For a $q \times q$ matrix $\boldsymbol{Q}$, we denote by $\text{col}(\boldsymbol{Q})$ its column space. If $\boldsymbol{Q}$ is symmetric, we denote by $\lambda_1(\boldsymbol{Q}), \ldots, \lambda_q(\boldsymbol{Q})$ its eigenvalues, labeled in nondecreasing order of magnitude; by $m_i(\boldsymbol{Q})$ the multiplicity of $\lambda_i(\boldsymbol{Q})$, for $i = 1, \ldots, q$; by $\boldsymbol{f}_1(\boldsymbol{Q}), \ldots, \boldsymbol{f}_q(\boldsymbol{Q})$ a set of orthonormal (with respect to the Euclidean norm) eigenvectors of $\boldsymbol{Q}$, with the eigenvector $\boldsymbol{f}_i(\boldsymbol{Q})$ being pertinent to the eigenvalue $\lambda_i(\boldsymbol{Q})$; by $E_i(\boldsymbol{Q})$ the eigenspace associated to $\lambda_i(\boldsymbol{Q})$, for $i = 1, \ldots, q$. Note that when $\boldsymbol{W}$ is symmetric, $\lambda_n(\boldsymbol{W}) = \lambda_{\max}$. When $\boldsymbol{W}$ is nonsymmetric, $\lambda_{\max}$ is still well-defined, because $\boldsymbol{W}$ always has a (real) positive eigenvalue by Theorem A.2 All matrices considered in this paper are real.

## 3. LIMITING POWER

In Section 3 we extend the results in Krämer (2005) on the power of tests for residual spatial autocorrelation. Krämer's results are briefly summarized in Section 3.1, whereas our main results are presented in Section 3.2. In Section 3.3 we report results from numerical experiments aimed at assessing the practical relevance of our analysis. In Section 3.4 we discuss the particular case of pure SAR(1) processes. Finally, in Section 3.5 we consider some models that are generalizations of, or alternatives to, a spatial error model.

### 3.1. Previous Results (Krämer, 2005)

Krämer (2005) considers distinguishing $\rho = 0$ from $\rho > 0$ in a spatial error model, under Gaussianity and when $\boldsymbol{W}$ is symmetric. This is a particular case of the testing problem described in Section 2.1. Krämer focuses on test statistics that can be expressed as ratios of quadratic forms in regression errors. More specifically, he considers tests that reject when $\boldsymbol{u}'\boldsymbol{Q}_1\boldsymbol{u}/\boldsymbol{u}'\boldsymbol{Q}_2\boldsymbol{u} > c$ for some $n \times n$ matrices $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$ that in general depend on $\boldsymbol{X}$ and $\boldsymbol{W}$. For example, the Cliff-Ord test and a POI test (8) belong to this class of tests (the former is obtained when $\boldsymbol{Q}_1 = \boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{M}_{\boldsymbol{X}}$ and $\boldsymbol{Q}_2 = \boldsymbol{M}_{\boldsymbol{X}}$, the latter when $\boldsymbol{Q}_1 = -\boldsymbol{C}'\left(\boldsymbol{C}\boldsymbol{\Sigma}(\bar{\rho})\boldsymbol{C}'\right)^{-1}\boldsymbol{C}$ and $\boldsymbol{Q}_2 = \boldsymbol{M}_{\boldsymbol{X}}$).

Henceforth, by "limiting power" of a test for autocorrelation in the context of a spatial autoregression, we mean the limit of the power function as $\rho \to \lambda_{\max}^{-1}$ (from the left). We denote by $\boldsymbol{f}_{\max}$ a particular eigenvector of $\boldsymbol{W}$ pertaining to $\lambda_{\max}$ (the precise definition is in Appendix A). Let $\xi := \boldsymbol{f}_{\max}'\left(\boldsymbol{Q}_1 - c\boldsymbol{Q}_2\right)\boldsymbol{f}_{\max}$. Theorems 1 and 2 in Krämer (2005) state that the limiting power of the above tests is 0 if $\xi < 0$, 1 if $\xi > 0$, and, generally, in $(0, 1)$ if $\xi = 0$. The third case is unlikely to occur because of the assumption that $\boldsymbol{W}$ is symmetric. Let us consider, for example, the Cliff-Ord test. Then, $\xi = 0$ if and only if either $\boldsymbol{f}_{\max} \in \text{col}(\boldsymbol{X})$ or

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

$c = \boldsymbol{f}'_{\max}\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{W}\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{f}_{\max}/\boldsymbol{f}'_{\max}\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{f}_{\max}$. For fixed $\boldsymbol{X}$ and $\boldsymbol{W}$, the latter condition is hugely restrictive, since it requires the critical value $c$ to be equal to a single specific value. The condition $\boldsymbol{f}_{\max} \in \mathrm{col}(\boldsymbol{X})$ is also restrictive: when $\boldsymbol{W}$ is symmetric there is generally no reason why $\boldsymbol{X}$ should be such that $\boldsymbol{f}_{\max} \in \mathrm{col}(\boldsymbol{X})$.[7]

In Section 3.2.2 we shall prove the nontrivial fact that Krämer's conditions can be extended to models with nonsymmetric $\boldsymbol{W}$. In many applications, the weights matrices of SAR(1) models are nonsymmetric; for instance, a row-standardized (so that all its row-sums are equal to 1) matrix is generally nonsymmetric. Note that the condition $\boldsymbol{f}_{\max} \in \mathrm{col}(\boldsymbol{X})$ is satisfied whenever $\boldsymbol{W}$ is row-standardized and an intercept is included in the regression.

## 3.2. Main Results

*3.2.1. The General Case.* Before investigating the case of a spatial error model, it is convenient to consider our testing problem in the context of the general model (1). Let us start from a simple observation. When $\rho \in \Psi^+$, $\boldsymbol{\Sigma}(\rho)$ is positive definite and hence $\boldsymbol{y}$ has positive density over the whole sample space $\mathbb{R}^n$. Thus, for any $\rho \in \Psi^+$, any critical region for testing $\rho = 0$ has probability content (i.e., power) strictly between 0 and 1.

For simplicity, we assume that $\Psi^+$ has a finite right boundary, to be denoted by $a$.[8] The power of a test for $\rho = 0$ as $\rho \to a$ requires more attention. Clearly, the limiting power as $\rho \to a$ depends on the limiting behavior of the density of $\boldsymbol{y}$, but we will see below that important information can be obtained just by looking at the limiting behavior of $\boldsymbol{\Sigma}(\rho)$. There are three possibilities: (i) $\boldsymbol{\Sigma}(a)$ exists and is positive definite; (ii) $\boldsymbol{\Sigma}(a)$ exists and is singular; or (iii) $\boldsymbol{\Sigma}^{-1}(a)$ exists and is singular. By the argument in the previous paragraph, it is clear that in case (i), the limiting power of any critical region must be in $(0, 1)$. Case (ii) applies, for instance, to moving average models, and will be briefly dealt with in Section 3.5.4. Here, we focus on case (iii) and, in particular, on the case $\mathrm{rank}(\boldsymbol{\Sigma}^{-1}(a)) = n - 1$, which, as we will see below, is the relevant one for spatial and stationary time series autoregressive models. For the extension to the more general case when $\mathrm{rank}(\boldsymbol{\Sigma}^{-1}(a)) < n$, see Remark 1 below. We denote by $\mathrm{int}(S)$, $\mathrm{bd}(S)$, and $\mathrm{cl}(S)$, the interior, the boundary, and the closure of a set $S$, respectively. By an invariant critical region we simply mean the subset of the sample space $\mathbb{R}^n$ where an invariant test rejects the null hypothesis.

THEOREM 1. *Consider an invariant critical region $\Phi$ for testing $\rho = 0$ against $\rho > 0$ in model (1). Assume that $\boldsymbol{\Sigma}(\rho)$ is positive definite as $\rho \to a$, and that* $\mathrm{rank}(\boldsymbol{\Sigma}^{-1}(a)) = n - 1$. *The limiting power of $\Phi$ as $\rho \to a$ is:*

- 1 *if* $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a)) \in \mathrm{int}(\Phi)$;
- *in* $(0, 1)$ *if* $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a)) \in \mathrm{bd}(\Phi)$; *or*
- 0 *if* $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a)) \notin \mathrm{cl}(\Phi)$.

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

Theorem 1 asserts that, to some extent, the limiting power of an invariant test is determined by the position of $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a))$ relative to the critical region. The result is quite general, in that it holds for any $\boldsymbol{\Sigma}(\rho)$ satisfying the stated conditions, any $\boldsymbol{X}$, and any invariant test.

The reason why the limiting power may disappear in the context of model (1) is best understood geometrically. If $\boldsymbol{\Sigma}^{-1}(a)$ has reduced rank, then, as $\rho \to a$, the model is not supported on the whole sample space $\mathbb{R}^n$, but only on a subspace thereof (more precisely, the limiting model is a degenerate distribution on a translation of the nullspace of $\boldsymbol{\Sigma}^{-1}(a)$; see the proof of the theorem for details). As a consequence, any critical region that does not (almost surely) intersect such a subspace has vanishing probability content, and hence vanishing power, as $\rho \to a$. On the contrary, a critical region has full limiting power if it contains (almost surely) that subspace.

Let us now concentrate on the case when the limiting power is in $(0, 1)$. This happens when $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a))$ falls in the boundary of $\Phi$. Such a condition may seem very restrictive. In fact, the boundary of an invariant critical region always contains $\mathrm{col}(\boldsymbol{X})$,[9] and therefore the condition is satisfied whenever $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a)) \in \mathrm{col}(\boldsymbol{X})$. This occurs, in particular, in models such that $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a))$ is a vector with identical entries, and such that an intercept is included among the regressors. For example, $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a))$ is a vector with identical entries in the case of covariance stationary AR(1) processes (see Krämer, 1985, and below), or in the case of a SAR(1) process with row-standardized $\boldsymbol{W}$ (see Section 3.2.2). Theorem 3.1 says that in such models the limiting power of any invariant critical region cannot be either 0 or 1, as long as an intercept is included in the regression.

The case of a regression model with AR(1) disturbances represents an important application of Theorem 1. More specifically, consider the error process $u_i = \rho u_{i-1} + \varepsilon_i$, for $i = 1, \ldots, n$, with the initial condition $u_0$ chosen so that the process is covariance stationary (i.e., $\mathrm{E}(u_0) = 0$ and $\mathrm{var}(u_0) = \sigma^2(1 - \rho^2)^{-1}$). Suppose that we are interested in the power of tests for $\rho = 0$ as $\rho$ approaches the unit root. Previous contributions in this context have focused on the power of the Durbin-Watson and some related tests under Gaussianity; see, e.g., Krämer (1985), Zeisel (1989), and Bartels (1992). Since, as is easily shown, the above AR(1) process yields $\mathrm{rank}(\boldsymbol{\Sigma}^{-1}(1)) = n - 1$, Theorem 1 applies, and shows that the results in those papers can be extended to any invariant test for residual serial correlation and to non-Gaussian distributions. Note that the assumption on $u_0$ plays a crucial role here. In general, for an $u_0$ other than the one leading to covariance stationarity (e.g., a fixed $u_0$), $\boldsymbol{\Sigma}^{-1}(1)$ is nonsingular, and hence the limiting power as $\rho \to 1$ is in $(0, 1)$, by the observation at the beginning of this section.

**Remark 1.** Theorem 1 can be generalized to the case when $0 < \mathrm{rank}(\boldsymbol{\Sigma}^{-1}(a)) < n$. Inspection of the proof of the theorem reveals that the formulation of the conditions for the limiting power to be 0, 1, or in $(0, 1)$ would be more complicated in that case. For instance, the condition for a vanishing power should be replaced by

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

the condition that $E_1(\boldsymbol{\Sigma}^{-1}(a)) \cap \mathrm{cl}(\Phi)$ has $\mathrm{rank}(\boldsymbol{\Sigma}^{-1}(a))$-dimensional Lebesgue measure zero.

**Remark 2.** Theorem 1 can be extended to non-invariant critical regions. Again, this would involve a more complicated formulation, because, contrary to an invariant critical region, a non-invariant one may contain only a subset of the limiting support of model (1). In particular, the lack of invariance with respect to transformations $\boldsymbol{y} \to \boldsymbol{y} + \boldsymbol{X\delta}$, $\boldsymbol{\delta} \in \mathbb{R}^k$ would imply that the conditions in the theorem depend on the unknown parameter $\boldsymbol{\beta}$.

*3.2.2. Spatial Error Model.* In the case of the covariance structure $\boldsymbol{\Sigma}(\rho) = [(\boldsymbol{I} - \rho \boldsymbol{W}')(\boldsymbol{I} - \rho \boldsymbol{W})]^{-1}$ of a SAR(1) process, the right boundary $a$ of $\Psi^+$ is $\lambda_{\max}^{-1}$. We now restrict our attention to the limiting power, as $\rho \to \lambda_{\max}^{-1}$, of invariant tests in a spatial error model (defined by equations (1) and (3)).

For a SAR(1) process, $\boldsymbol{\Sigma}^{-1}(\lambda_{\max}^{-1})$ has rank $n - 1$ for any $\boldsymbol{W}$, by Lemma D.4. It follows that Theorem 1 applies to any spatial error model, leading to the following corollary.

COROLLARY 1. *In a spatial error model, the limiting power of an invariant critical region* $\Phi$ *for testing* $\rho = 0$ *against* $\rho > 0$ *is:*

- 1 *if* $\boldsymbol{f}_{\max} \in \mathrm{int}(\Phi)$;
- *in* $(0, 1)$ *if* $\boldsymbol{f}_{\max} \in \mathrm{bd}(\Phi)$; *or*
- 0 *if* $\boldsymbol{f}_{\max} \notin \mathrm{cl}(\Phi)$.

There are three main differences between Corollary 1 and the results summarized in Section 3.1. *First*, Corollary 1 holds for the whole class of invariant tests, which is much larger than the class of tests that can be expressed as ratios of quadratic forms in the regression errors. In particular, Corollary 1 holds regardless of the analytical form of the invariant test statistic, and therefore, it also holds for invariant tests whose test statistics are analytically complicated, or—as it is the case for an LR test—unavailable in closed form. It should be noted that Corollary 1 implies that the zero limiting power phenomenon cannot be attributed to the form of a specific test (contrary to what is argued, for instance, in Krämer, 2005, p. 490). Instead, the phenomenon is due to the fact that a SAR(1) model tends, as $\rho \to \lambda_{\max}^{-1}$, to be supported on a subspace of the sample space, namely the one-dimensional space spanned by $\boldsymbol{f}_{\max}$; see the proof and the discussion of Theorem 1 for details. *Second*, Corollary 1 is not restricted to Gaussian models. *Third*, it does not require symmetry of $\boldsymbol{W}$. This is important, because in SAR(1) models $\boldsymbol{W}$ is very often row-standardized, which generally entails asymmetry. For a row-standardized $\boldsymbol{W}$, $\boldsymbol{f}_{\max}$ is a vector with identical entries. Recall from Section 3.2.1 that whenever a critical region $\Phi$ is invariant, $\mathrm{col}(\boldsymbol{X}) \subset \mathrm{bd}(\Phi)$. Thus, when $\boldsymbol{W}$ is row-standardized and the regression contains an intercept, Corollary 1

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

establishes that the limiting power of any invariant test is in $(0, 1)$. Conversely, when $\boldsymbol{W}$ is not (a scalar multiple of) a row-standardized matrix, the limiting power of an invariant test is generally either 0 or 1, because it is unlikely that $\boldsymbol{f}_{\max}$ falls in $\mathrm{col}(\boldsymbol{X})$ or more generally in $\mathrm{bd}(\Phi)$.[10]

In view of the above observations, one might be tempted to regard row-standardization of $\boldsymbol{W}$ as a simple device to avoid the zero limiting power problem. Such a temptation should be resisted. Instead, as emphasized by Kelejian and Prucha (2009), the decision as to whether or not to row-standardize $\boldsymbol{W}$ should be based on theoretical considerations concerning the particular phenomenon under analysis. It should also be noted that, even if the limiting power cannot be exactly zero when $\boldsymbol{W}$ is row-standardized and the regression contains an intercept, it can still be very low.[11]

The practical usefulness of the conditions stated in Corollary 1 stems from the fact that such conditions are generally simple to check. Let us consider the invariant critical region $\Phi$ that rejects $\rho = 0$ for large values of some univariate statistic $T(\boldsymbol{y})$, i.e.,[12]

$$\Phi = \left\{ \boldsymbol{y} \in \mathbb{R}^n : T(\boldsymbol{y}) > c \right\}. \tag{9}$$

The boundary of such a critical region consists of not only $\{\boldsymbol{y} \in \mathbb{R}^n : T(\boldsymbol{y}) = c\}$ but also, as pointed out in Section 3.2.1, $\mathrm{col}(\boldsymbol{X})$. While it is very unlikely that $\boldsymbol{f}_{\max}$ falls in the former set (in general, there is only one value of $c$ in a continuum of points such that $T(\boldsymbol{f}_{\max}) = c$), we have already noted above that in some important cases $\boldsymbol{f}_{\max} \in \mathrm{col}(\boldsymbol{X})$. Theorem 1 then says that the limiting power of the critical region (9) is: 1 if $T(\boldsymbol{f}_{\max}) > c$; 0 if $T(\boldsymbol{f}_{\max}) < c$; and in $(0, 1)$ if $T(\boldsymbol{f}_{\max}) = c$ or $\boldsymbol{f}_{\max} \in \mathrm{col}(\boldsymbol{X})$. Such conditions are simple to check as long as $\boldsymbol{f}_{\max}$ is known or can be computed efficiently. Consider, for example, the Cliff-Ord test. For the $\boldsymbol{\Sigma}(\rho)$ implied by a SAR(1) process, its limiting power is 1, 0, or in $(0, 1)$, depending on whether $\boldsymbol{f}'_{\max}(\boldsymbol{M_X W M_X} - c\boldsymbol{M_X})\boldsymbol{f}_{\max}$ is, respectively, positive, negative, or 0.[13]

So far we have focused on a fixed test for spatial autocorrelation. Corollary 1 also has consequences for the power envelope $\pi_\rho(\rho)$ of invariant tests. Indeed, since it asserts that any critical region that includes $\boldsymbol{f}_{\max}$ in its interior must have full limiting power, Corollary 1 implies that, under the condition $\boldsymbol{f}_{\max} \notin \mathrm{col}(\boldsymbol{X})$, $\pi_\rho(\rho)$ approaches 1 as $\rho \to \lambda_{\max}^{-1}$.[14] Conversely, under the condition $\boldsymbol{f}_{\max} \in \mathrm{col}(\boldsymbol{X})$, the corollary implies that the limit of $\pi_\rho(\rho)$ as $\rho \to \lambda_{\max}^{-1}$ must be strictly between $\alpha$ and 1. This leads us to the conclusion that the null hypothesis $\rho = 0$ can be distinguished from the limiting alternative $\rho \to \lambda_{\max}^{-1}$ with zero type II error probability if and only if $\boldsymbol{f}_{\max} \notin \mathrm{col}(\boldsymbol{X})$.

**Remark 3.** In the context of a spatial error model, consideration of the extreme case $\rho \to \lambda_{\max}^{-1}$ corresponds, in general, to studying power when it is most needed, i.e., when the efficiency of the OLS estimator of $\boldsymbol{\beta}$, relative to (some feasible version of) the GLS estimator, is low. Indeed, for most combinations of

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

$\boldsymbol{X}$ and $\boldsymbol{W}$ and according to most measures of efficiency, the relative efficiency of the OLS estimator is decreasing in $\rho$ (see, e.g., Cordy and Griffith, 1993). There are exceptions: for instance, the measure of efficiency used in Krämer and Donninger (1987) may be increasing in $\rho$ in some circumstances, but see Dielman and Pfaffenberger (1989) for problems with that measure.

**Remark 4.** Corollary 1 admits an interpretation in terms of the Cliff-Ord statistic $\hat{\boldsymbol{u}}'\boldsymbol{W}\hat{\boldsymbol{u}}/\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$. Besides being used to test for spatial autocorrelation, the Cliff-Ord statistic is often regarded as an index of autocorrelation. When $\boldsymbol{W}$ is *symmetric*, $\hat{\boldsymbol{u}}'\boldsymbol{W}\hat{\boldsymbol{u}}/\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$ achieves a maximum at $\hat{\boldsymbol{u}} = \boldsymbol{f}_{\max}$, by Lemma D.1 Thus, Corollary 1 asserts that, for fixed $\boldsymbol{X}$ and fixed symmetric $\boldsymbol{W}$, an invariant critical region has full limiting power only if it contains the points in the sample space that maximize the Cliff-Ord statistic. Indeed, according to the interpretation of the Cliff-Ord statistic as an autocorrelation coefficient, it certainly makes sense to reject $\rho = 0$ when a large value of $\hat{\boldsymbol{u}}'\boldsymbol{W}\hat{\boldsymbol{u}}/\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$ is observed. The situation is, however, less intuitive when $\boldsymbol{W}$ is *nonsymmetric*. In that case, $\hat{\boldsymbol{u}}'\boldsymbol{W}\hat{\boldsymbol{u}}/\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$ is maximized by the eigenvectors of $\boldsymbol{W} + \boldsymbol{W}'$ associated to the largest eigenvalue of $\boldsymbol{W} + \boldsymbol{W}'$. Observe that $\boldsymbol{f}_{\max}$ does not need to be one of such eigenvectors. Hence, for SAR(1) models with nonsymmetric $\boldsymbol{W}$, Corollary 1 implies that an invariant critical region may have vanishing limiting power even if it contains the values of $\boldsymbol{y}$ that maximize the Cliff-Ord statistic. We shall come back to the role played by the symmetry of $\boldsymbol{W}$ in Section 3.4.

### 3.3. Numerical Examples

In Section 3.3 we report results from a small Monte Carlo experiment aimed at illustrating how the matrices $\boldsymbol{X}$ and $\boldsymbol{W}$ affect the exact power of tests for residual spatial autocorrelation. In particular, our objective is to show how sensitive the power can be to $\boldsymbol{X}$, when $\rho$ is large but not necessarily very close to $\lambda_{\max}^{-1}$. For brevity, we restrict our attention to the Cliff-Ord test and to Gaussian models. Related numerical investigations are contained in Krämer (2005).

We consider $10^6$ replications of the $n \times 2$ matrix $\boldsymbol{X} = (\boldsymbol{\iota} : \boldsymbol{z})$, where $\boldsymbol{\iota} := (1, \ldots, 1)'$ and $\boldsymbol{z} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I})$.[15] The weights matrices are derived from the maps of the $n = 17$ counties in Nevada and the $n = 23$ counties in Wyoming; see Figure 1. We consider both a binary $\boldsymbol{W}$, specified according to the queen criterion (i.e., $(\boldsymbol{W})_{ij} = 1$ if two distinct counties $i$ and $j$ share a common boundary or a common point, $(\boldsymbol{W})_{ij} = 0$ otherwise), and its row-standardized version. The average number of neighbors of a county is 4.35 in Nevada and 4.52 in Wyoming, whereas the sparseness of $\boldsymbol{W}$ (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 from the point of view of testing for autocorrelation.

In order to show how sensitive the power of the Cliff-Ord, denoted by $\pi_{CO}(\rho)$, is to $\boldsymbol{X}$, in Table 1 we display the percentage frequency distribution. The size is set

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

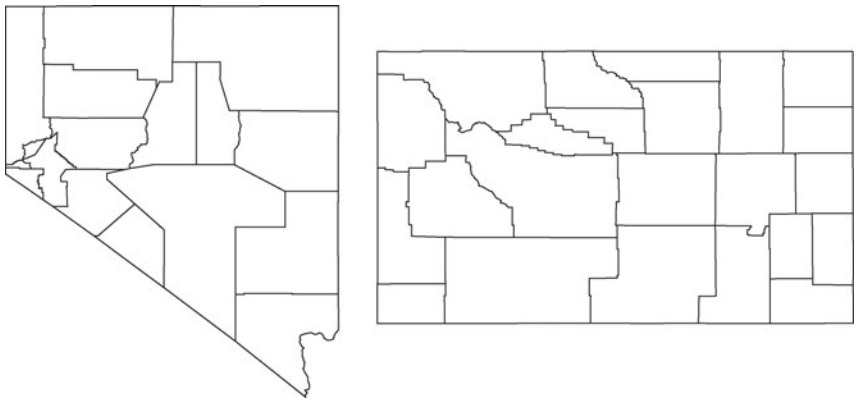


**FIGURE 1.** Maps of the 17 counties in Nevada (left) and the 23 counties in Wyoming (right).

to 0.05, and the power is computed by the Imhof method (Imhof, 1961). We report values for $\rho = 0.9\lambda_{\max}^{-1}$ and $\rho = 0.95\lambda_{\max}^{-1}$. To give an indication of how close such points are to $\lambda_{\max}^{-1}$, the third column of Table 1 gives the average correlation between pairs of neighboring counties (there are 37 such pairs in Nevada and 54 in Wyoming; averages over non-neighbors, not reported, are much lower).[16]

**TABLE 1.** Average correlation between neighbors and percentage frequency distribution of the power $\pi_{CO}(\rho)$ of the Cliff-Ord test, in model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$, where $\boldsymbol{u}$ is a Gaussian SAR(1) process and $\boldsymbol{X}$ contains an intercept and a standard normal variate

| | | | $\pi_{CO}(\rho)$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\rho\lambda_{\max}$ | Average neigh. correlation | 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 $\boldsymbol{W}$ | 0.90 | 0.85 (0.70−0.93) | 0.11 | 0.25 | 28.42 | 71.05 | 0.17 | · | · |
| | 0.95 | 0.95 (0.87−0.98) | 0.29 | 5.75 | 36.29 | 53.43 | 4.11 | 0.13 | · |
| row-st $\boldsymbol{W}$ | 0.90 | 0.88 (0.81−0.93) | · | · | 0.02 | 0.16 | 41.47 | 58.35 | · |
| | 0.95 | 0.96 (0.93−0.98) | · | · | 0.01 | 0.05 | 1.56 | 98.38 | · |
| Wyoming | | | | | | | | | |
| binary $\boldsymbol{W}$ | 0.90 | 0.80 (0.60−0.92) | · | · | · | 0.02 | 0.69 | 99.29 | · |
| | 0.95 | 0.92 (0.77−0.98) | · | · | · | 0.02 | 0.10 | 1.76 | 98.12 |
| row-st $\boldsymbol{W}$ | 0.90 | 0.85 (0.76−0.92) | · | · | · | · | · | 0.50 | 99.50 |
| | 0.95 | 0.95 (0.90−0.97) | · | · | · | · | · | · | 100 |

*Notes*: Minimum and maximum correlations are given in parentheses.
The power is computed by the Imhof method over $10^6$ replications of $\boldsymbol{X}$, with $\alpha = 0.05$.

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

It appears from Table 1 that in the case of Nevada, $\pi_{CO}(\rho)$ depends to a very large extent on $\boldsymbol{X}$, even for values of $\rho$ that are not in a very small neighborhood of $\lambda_{\max}^{-1}$. The dependence is less pronounced in the case of Wyoming.

Before carrying on with our numerical example, it is convenient to introduce a general measure of the difficulty of testing for spatial autocorrelation as $\rho \to \lambda_{\max}^{-1}$. Consider some invariant test. By Corollary 1, whether or not its limiting power vanishes depends on $\alpha$, because whether or not $\boldsymbol{f}_{\max} \in \mathrm{cl}(\Phi)$ depends on the size of $\Phi$. In particular, the limiting power of a test may vanish for some fixed size but be positive for some larger size. In the following definition, by exact test we mean a test whose critical value is selected from the exact null distribution of the test statistic.

DEFINITION 1. *For an exact invariant test of $\rho = 0$ against $\rho > 0$ in a SAR(1) model, $\alpha^*$ is the infimum of the set of values of $\alpha \in (0, 1]$ such that the limiting power does not vanish.*

For fixed $\boldsymbol{X}$, $\boldsymbol{W}$, and $\Phi$ such that $\boldsymbol{f}_{\max} \notin \mathrm{bd}(\Phi)$, $\alpha^*$ is a measure of the distinguishability between the null hypothesis $\rho = 0$ and the limiting alternative $\rho \to \lambda_{\max}^{-1}$.[17] A large $\alpha^*$ indicates that a large critical region is necessary to avoid the zero limiting power problem. We stress that $\alpha^*$ depends on $\boldsymbol{W}$, on the invariant test under consideration, and on $\boldsymbol{X}$ (through $\mathrm{col}(\boldsymbol{X})$, because of the invariance property of the tests). A simple way of computing $\alpha^*$ is provided by Lemma D.2.

We can now go back to our numerical example. Recall that $\mathrm{col}(\boldsymbol{X})$ is contained in the boundary of any invariant critical region. It follows, by Corollary 1, that in our experiment the limiting power is either 0 or 1 when $\boldsymbol{W}$ is binary (as, in that case, $\boldsymbol{f}_{\max} \notin \mathrm{col}(\boldsymbol{X})$ almost surely), whereas it is in $(0, 1)$ when $\boldsymbol{W}$ is row-standardized (as, in that case, $\boldsymbol{f}_{\max}$ is a scalar multiple of $\boldsymbol{\iota}$, and hence is in $\mathrm{col}(\boldsymbol{X})$). Thus, in order to study the zero limiting power phenomenon, we restrict attention to the binary specification of $\boldsymbol{W}$. In columns 2–4 of Table 2 we display some statistics regarding the realizations of $\alpha^*$ over the $10^6$ replications of $\boldsymbol{X}$. Observe that $\alpha^*$ depends to a very large extent on $\mathrm{col}(\boldsymbol{X})$. In the case of Nevada, on

**TABLE 2.** Minimum, maximum, average $\alpha^*$, frequency of zero limiting power, and average shortcoming of the Cliff-Ord test, for a binary $\boldsymbol{W}$

| | | | | | Average shortcoming at $\rho\lambda_{\max} = 0.90$ | | Average shortcoming at $\rho\lambda_{\max} = 0.95$ | |
|---|---|---|---|---|---|---|---|---|
| | Min $\alpha^*$ | Max $\alpha^*$ | Average $\alpha^*$ | Frequency | $\pi_{CO}(\rho) \to 0$ | $\pi_{CO}(\rho) \to 1$ | $\pi_{CO}(\rho) \to 0$ | $\pi_{CO}(\rho) \to 1$ |
| Nevada | $2.8 \cdot 10^{-4}$ | 0.994 | 0.082 (0.061) | 0.77 | 0.20 | 0.16 | 0.32 | 0.24 |
| Wyoming | $8.6 \cdot 10^{-7}$ | 0.430 | $5.6 \cdot 10^{-4}$ ($2.8 \cdot 10^{-3}$) | $5.2 \cdot 10^{-4}$ | 0.15 | 0.03 | 0.26 | 0.02 |

*Note*: Standard deviations of $\alpha^*$ are given in parentheses.

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

average it is necessary to use a critical region of size $\alpha \geq 0.082$ in order to achieve a nonzero limiting power. For one particular value of $\boldsymbol{X}$, $\alpha^*$ was as large as 0.994, meaning that in the presence of such an $\boldsymbol{X}$ the critical region of the Cliff-Ord test has vanishing limiting power unless its size is at least 0.994. Column 5 of Table 2 contains the observed relative frequency of the zero limiting power problem. It is also useful to look at the impact of a zero limiting power on the performance of the Cliff-Ord test at values of $\rho$ that are large but not too close to $\lambda_{\max}^{-1}$. As above, we consider the values $0.9\lambda_{\max}^{-1}$ and $0.95\lambda_{\max}^{-1}$. As a measure of performance, we take the shortcoming, defined as the difference between the power envelope $\pi_\rho(\rho)$ and the power $\pi_{CO}(\rho)$ of the Cliff-Ord test (see, e.g., Lehmann and Romano, 2005, p. 337). The last four columns of Table 2 display the average of the shortcoming over the replications of $\boldsymbol{X}$ yielding a zero limiting power (columns 6 and 8) and the average over the replications yielding a full limiting power (columns 7 and 9). Observe that, on average, an $\boldsymbol{X}$ yielding a zero limiting power causes shortcomings at $\rho = 0.9\lambda_{\max}^{-1}$ and $\rho = 0.95\lambda_{\max}^{-1}$ that are significantly larger than the corresponding shortcomings associated to an $\boldsymbol{X}$ yielding a full limiting power. This suggests that the impact of the zero limiting power problem is not localized only in a very small neighborhood of $\lambda_{\max}^{-1}$.

It is clear from Tables 1 and 2 that the numerical results regarding Nevada and Wyoming are extremely different. In particular, the zero limiting power frequency is very large in the case of Nevada and very small in the case of Wyoming. On repeating our simulations for different weights matrices and different tests, we have found that the zero limiting power frequency is generally very sensitive not only to $\boldsymbol{W}$, but also to $\alpha$, $k$, the choice of a test, and the distribution of $\boldsymbol{X}$. For most matrices $\boldsymbol{W}$ likely to be used in applications and for most distributions of $\boldsymbol{X}$, the zero limiting power frequency is generally small when $n - k$ is large. From a practical perspective, this suggests that the zero limiting power problem is mainly a small sample problem. It should be noticed, however, that for any $\boldsymbol{W}$, the probability of a zero limiting power is positive as long as $\boldsymbol{X}$ is unrestricted, regardless of $n$, and that it is possible to construct matrices $\boldsymbol{W}$ such that, for some distributions of $\boldsymbol{X}$, the probability of a zero limiting power is large even when $n - k$ is large. Examples of such matrices are 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) or a very dense matrix. When $\boldsymbol{W}$ is defined on a regular grid, one can study how the zero limiting power frequency depends on $n$ explicitly (cf. Table 1 of Krämer, 2005).

To summarize, the main conclusion of our numerical study is that, in some cases, the probability that the limiting power of the Cliff-Ord test vanishes may well be nonnegligible. This obviously induces a large dependence of the power of the Cliff-Ord test on $\boldsymbol{X}$ as $\rho \to \lambda_{\max}^{-1}$, but the numerical results indicate that both the power and the shortcoming may still depend to a large extent on $\boldsymbol{X}$ for values of $\rho$ in a rather large neighborhood of $\lambda_{\max}^{-1}$. As mentioned in Remark 3, this is cause of concern, because such values may induce a large inefficiency of the ordinary least squares estimator of $\boldsymbol{\beta}$.

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

### 3.4. Pure SAR(1) Model

We have seen above that tests for autocorrelation in the context of a regression model with SAR(1) disturbances do not necessarily achieve full power as $\rho \to \lambda_{\max}^{-1}$. This is so even when the tests have some finite sample optimality properties, as in the case of LBI and POI tests. It is natural to wonder whether such a phenomenon should be completely ascribed to the presence of regressors. To investigate this issue, we now consider pure SAR(1) processes, that is, SAR(1) processes with $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{0}$. The extension to models with known mean can be trivially obtained by taking $\boldsymbol{y}$ to be a demeaned variable.

Recall from Section 3.2.2 that, for an invariant critical region $\Phi$ of the form (9), $\mathrm{bd}(\Phi) = \mathrm{col}(\boldsymbol{X}) \cup \{\boldsymbol{y} \in \mathbb{R}^n : T(\boldsymbol{y}) = c\}$. In the case of a pure model, $\mathrm{col}(\boldsymbol{X}) = \varnothing$. Thus, it is very unlikely that $\boldsymbol{f}_{\max} \in \mathrm{bd}(\Phi)$, as this would require the critical value $c$ to be precisely equal to $T(\boldsymbol{f}_{\max})$. Neglecting this possibility, the limiting power of $\Phi$ can only be 0 or 1, by Corollary 1. One might hope that, at least when $\Phi$ is LBI or POI, the limiting power is 1 for any size $\alpha$. This is not the case, as shown by the following proposition.

PROPOSITION 1. *Consider testing $\rho = 0$ against $\rho > 0$ in a pure SAR(1) model. The limiting power of the Cliff-Ord test or of a test (8) is 1 irrespective of $\alpha$ if and only if $\boldsymbol{f}_{\max}$ is an eigenvector of $\boldsymbol{W}'$.*

The tests considered in Proposition 1 are obtained from (7) and (8) when $\boldsymbol{X} = \boldsymbol{O}$, and thus reject when $\boldsymbol{y}'\boldsymbol{W}\boldsymbol{y}/\boldsymbol{y}'\boldsymbol{y} > c$ and $\boldsymbol{y}'(\boldsymbol{I} - \bar{\rho}\boldsymbol{W}')(\boldsymbol{I} - \bar{\rho}\boldsymbol{W})\boldsymbol{y}/\boldsymbol{y}'\boldsymbol{y} < c$. Recall that such tests are, respectively, LBI and POI when the distribution of $\boldsymbol{y}$ is elliptically symmetric.

The condition in Proposition 1 is always satisfied when $\boldsymbol{W}$ is symmetric. Hence, for a pure SAR(1) model with symmetric $\boldsymbol{W}$, the Cliff-Ord test and a test (8) always achieve full limiting power, regardless of their size. Conversely, when $\boldsymbol{W}$ is nonsymmetric, the condition in Proposition 1 is generally not met; for details, see Appendix C, where, in particular, it is shown that the condition is never met when $\boldsymbol{W}$ is a nonsymmetric matrix obtained by row-standardization of a symmetric matrix. Thus, in a pure SAR(1) model with nonsymmetric $\boldsymbol{W}$, there generally are values of $\alpha$ such that the limiting power of the Cliff-Ord test or of a test (8) vanishes. A simple example follows.

#### Example 1

A random variable is observed at $n$ units placed along a line and, in the context of a pure Gaussian SAR(1) 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 among first-order neighbors is stronger in one direction than in the other. Accordingly, $\boldsymbol{W}$ is taken so that $(\boldsymbol{W})_{ij}$, for $i, j = 1, \ldots, n$, is equal to some fixed positive scalar $w \neq 1$ if $i - j = 1$, to 1 if $j - i = 1$, and to 0 otherwise. In Figure 2, we plot the power function of the Cliff-Ord test, and the envelope

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

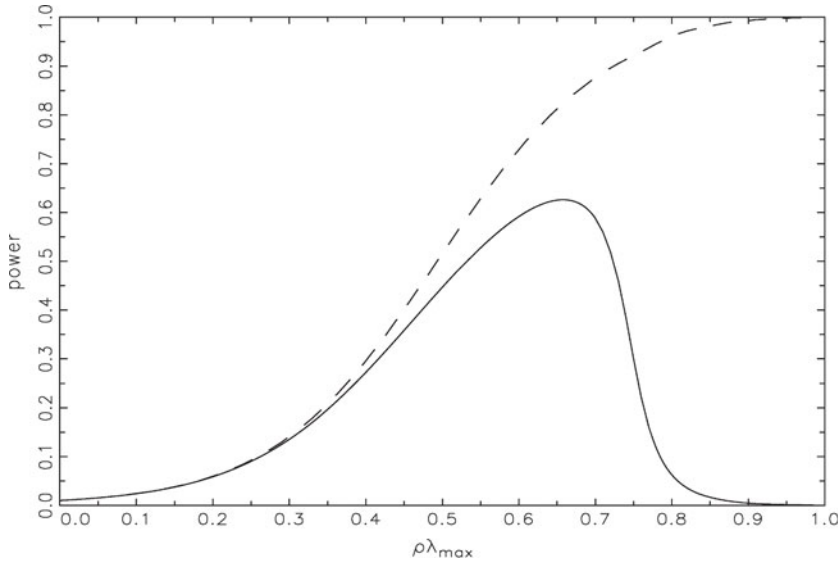



**FIGURE 2.** The power function of the Cliff-Ord test (solid line) and the envelope $\pi_\rho(\rho)$ (dashed line) for the pure SAR(1) model described in Example 1.

$\pi_\rho(\rho)$ for $n = 6$, $w = 10$, and $\alpha = 0.01$. The power has been computed numerically, via the Imhof method, and is plotted against $\rho\lambda_{\max}$, which ranges between 0 and 1.

Although it is based on an artificial $\boldsymbol{W}$, Figure 2 shows that the performance of a test for spatial autocorrelation may be extremely disappointing even in models that are not contaminated by regressors. It is of some interest to investigate which matrices $\boldsymbol{W}$ are particularly problematic from the point of the limiting power of tests of autocorrelation in pure SAR(1) models. The measure $\alpha^*$ introduced in Definition 1 can be used to this purpose, as we discuss next.

**Example 2**

Consider the testing problem in Example 1. The measure $\alpha^*$ is, as one would expect, decreasing in $n$ and increasing in $|w - 1|$ (this can be shown by using the Imhof method or other accurate numerical approximations to the null distribution of the statistic $\boldsymbol{y}'\boldsymbol{W}\boldsymbol{y}/\boldsymbol{y}'\boldsymbol{y}$). For the particular case of Figure 1, $\alpha^*$ is about 0.056, i.e., any critical region of size less than 0.056 has vanishing limiting power. To give another example, if $n = 30$ and $w = 50$, then $\alpha^*$ is about 0.063. Interestingly, if one "closes the line" (by setting $(\boldsymbol{W})_{1n} = w$ and $(\boldsymbol{W})_{n1} = 1$), then $\boldsymbol{W}$ becomes a scalar multiple of a doubly stochastic matrix, and consequently $\alpha^* = 0$ by the combination of Proposition 1 and Lemma C.1.

Numerical experiments not reported here suggest that the message delivered by Example 2 is very general. Namely, for a fixed $n$, large values of $\alpha^*$ are

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

typically associated to weights matrices $\boldsymbol{W}$ such that $(\boldsymbol{W})_{ij}/(\boldsymbol{W})_{ji}$ is large for at least one pair $(i, j)$. When $\boldsymbol{W}$ is the row-standardized version of a $(0, 1)$ matrix (i.e., a matrix containing only zeros and ones), $(\boldsymbol{W})_{ij}/(\boldsymbol{W})_{ji}$ cannot be larger than the ratio, say $r$, of the largest to the smallest row-sum of the $(0, 1)$ matrix, for any $i, j = 1, \ldots, n$. This implies that the asymmetry introduced by the popular practice of row-standardizing a $(0, 1)$ symmetric matrix does not yield large values of $\alpha^*$ in pure SAR(1) models. The largest possible value of $r$ over all $n \times n$ $(0, 1)$ symmetric matrices is $n - 1$, achieved by the adjacency matrix of a star graph. Even in the case of a star graph, the value of $\alpha^*$ associated to the corresponding row-standardized $\boldsymbol{W}$ is very small, and decreasing in $n$; for the Cliff-Ord test, $\alpha^* < 0.01$ whenever $n > 6$; that is, the limiting power of the Cliff-Ord test test is 1 as long as $n > 6$ and $\alpha > 0.01$.

We have thus found that, although asymmetry of $\boldsymbol{W}$ may cause the limiting power of POI and LBI tests to disappear when $\boldsymbol{X} = \mathbf{O}$, this typically occurs only for very small values of $\alpha$ or $n$ if the asymmetry of $\boldsymbol{W}$ is due to row-standardization of a $(0, 1)$ matrix. As we have seen in Section 3.2.2, the situation is very different when conditioning on regressors. In that case, the limiting power of POI and LBI may vanish even for large $\alpha$ or large $n$, regardless of $\boldsymbol{W}$.

## 3.5. Other Spatial Models

In this section we discuss how the analysis in Section 3.2 can be extended to study the limiting power of tests for autocorrelation in some more general models, and in some different spatial models.

*3.5.1. Nuisance Parameters in the Innovation Variance Matrix.* In applications, it is often useful to allow the innovation variance matrix $\sigma^2\boldsymbol{V}$ in (3) to depend on a vector of parameters $\boldsymbol{\theta}$. A particular case is when $\boldsymbol{V}(\boldsymbol{\theta})$ is diagonal, so that $\boldsymbol{\theta}$ controls the heteroskedasticity of the innovations $\boldsymbol{\varepsilon}$.[18]

For the problem of testing $\rho = 0$ in a spatial error model, $\boldsymbol{\theta}$ is a nuisance parameter that cannot be eliminated by invariance. Consequently, invariant tests are not similar, and their whole power function depends on $\boldsymbol{\theta}$. This is not a problem for the validity of Theorem 1. Indeed, it is straightforward to check that Theorem 1 continues to hold, provided that $\boldsymbol{\Sigma}(\rho)$ is replaced by $\boldsymbol{\Sigma}(\rho, \boldsymbol{\theta}) := (\boldsymbol{I} - \rho\boldsymbol{W})^{-1}\boldsymbol{V}(\boldsymbol{\theta})(\boldsymbol{I} - \rho\boldsymbol{W}')^{-1}$. The problem is that, in general, $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(\lambda_{\max}^{-1}, \boldsymbol{\theta}))$ depends on $\boldsymbol{\theta}$, and hence the conditions stated by the theorem cannot be checked. It should also be noted that, when $\boldsymbol{V}$ depends on $\boldsymbol{\theta}$, Corollary 1 does not obtain.

Theorem 1, however, may still provide useful information. This occurs whenever one is able to identify a set $\Theta$ such that one or more of the conditions in the theorem are satisfied for all $\boldsymbol{\theta} \in \Theta$. Suppose, for example, that for given $\boldsymbol{W}$, $\boldsymbol{X}$, and $\Phi$, one finds that $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(\lambda_{\max}^{-1}, \boldsymbol{\theta})) \in \text{int}(\Phi)$ for all $\boldsymbol{\theta} \in \Theta$. Then, Theorem 1 guarantees that the limiting power of $\Phi$ is 1 as long as $\boldsymbol{\theta} \in \Theta$. Identification of a set $\Theta$ may involve computational difficulties that will not be discussed here.

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

*3.5.2. Spatial Lag Model.* An alternative to a spatial error model is the so-called spatial lag model

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},\ \mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0},\ \mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}. \tag{10}$$

Model (10) is also known as a mixed regressive, spatial autoregressive model, and is widely used both in the applied and the theoretical spatial econometric literature; see, e.g., Ord (1975), Anselin (1988), and Lee (2002). In (10) we have taken $\mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}$, which can be done without any loss of generality as long as $\sigma^{-2}\mathrm{var}(\boldsymbol{\varepsilon})$ is fixed; for the case when $\mathrm{var}(\boldsymbol{\varepsilon})$ depends on extra parameters, the same considerations as in Section 3.5.1 apply.

Similarly to the case of a spatial error model, in the context of model (10) we are concerned with testing $\rho = 0$ against $\rho \in \Psi^+ = (0, \lambda_{\max}^{-1})$. For any $\rho \in \Psi^+$, the two models imply the same variance matrix, but different expectations for $\boldsymbol{y}$ ($\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X}\boldsymbol{\beta}$ in a spatial error model, $\mathrm{E}(\boldsymbol{y}) = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1}\boldsymbol{X}\boldsymbol{\beta}$ in a spatial lag model). As a consequence, the problem of testing for a spatially lagged dependent variable is quite different from that of testing for residual spatial autocorrelation. In particular, the former testing problem is not invariant under the transformations $\boldsymbol{y} \to \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\delta}$, $\boldsymbol{\delta} \in \mathbb{R}^k$. Thus, when testing $\rho = 0$ in model (10), there is no reason to focus on tests that are invariant under those transformations. The analog of Corollary 1 for a spatial lag model is the following result.

PROPOSITION 2. *Consider testing $\rho = 0$ against $\rho > 0$ in model (10) by means of a critical region $\Upsilon$ that is invariant under $\boldsymbol{y} \to \gamma \boldsymbol{y}$, $\gamma \in \mathbb{R}\backslash\{0\}$. The limiting power as $\rho \to \lambda_{\max}^{-1}$ is:*

- 1 *if* $\boldsymbol{f}_{\max} \in \mathrm{int}(\Upsilon)$;
- *in* $(0, 1)$ *if* $\boldsymbol{f}_{\max} \in \mathrm{bd}(\Upsilon)$; *or*
- 0 *if* $\boldsymbol{f}_{\max} \notin \mathrm{cl}(\Upsilon)$.

Contrary to Corollary 1, Proposition 2 does not require invariance with respect to the transformations $\boldsymbol{y} \to \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\delta}$, $\boldsymbol{\delta} \in \mathbb{R}^k$. In Section 3.2.1 we have seen that, in the context of a spatial error model, any critical region $\Phi$ that is invariant under such transformations has limiting power in $(0, 1)$ whenever $\boldsymbol{W}$ is row-standardized and the regression contains an intercept. This was due to the fact that $\boldsymbol{f}_{\max} \in \mathrm{col}(\boldsymbol{X}) \subset \mathrm{bd}(\Phi)$. In contrast, for a critical region $\Upsilon$ that is not invariant under the transformations $\boldsymbol{y} \to \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\delta}$, $\mathrm{col}(\boldsymbol{X})$ is generally not a subset of $\mathrm{bd}(\Upsilon)$. Proposition 2 then implies that the limiting power of a critical region $\Upsilon$ for testing $\rho = 0$ against $\rho > 0$ in model (10) is typically either 0 or 1, even when $\boldsymbol{W}$ is row-standardized and the regression contains an intercept.

*3.5.3. Spatial Autoregressive Model with Autoregressive Disturbances.* The tests for residual spatial autocorrelation or for spatial lag dependence considered so far can be generalized to tests in model

$$\boldsymbol{y} = \psi \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \qquad \boldsymbol{u} = \rho \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}, \qquad \mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}, \qquad \mathrm{var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{V}(\boldsymbol{\theta}) \tag{11}$$

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

(e.g., Anselin, 1988; Case, 1991). That is, one may be interested in testing $\rho = 0$ allowing for $\psi \neq 0$ (as, for instance, in Kelejian and Prucha, 2001), or in testing $\psi = 0$ allowing for $\rho \neq 0$.

Model (11) implies that $\text{var}(\boldsymbol{y})$ tends to a singular matrix as $\rho \to \lambda_{\max}^{-1}$ or $\psi \to \lambda_{\max}^{-1}$. Thus, the main argument of Section 3.2.1 continues to apply, both when testing $\rho = 0$ and when testing $\psi = 0$: as $\rho \to \lambda_{\max}^{-1}$ (resp. $\psi \to \lambda_{\max}^{-1}$), model (11) tends to be supported on a subspace of the sample space, and hence any critical region for $\rho = 0$ against $\rho > 0$ (resp. $\psi = 0$ against $\psi > 0$) that intersects such a subspace only on a set of measure zero will have vanishing limiting power. However, large values of $\psi$ or $\rho$ are less likely to occur in model (11) than in spatial error or spatial lag models.

*3.5.4. Spatial MA(1) Model.* A spatial first-order moving average (SMA(1)) process for the regression errors $\boldsymbol{u}$ (e.g., Anselin, 1988) is

$$\boldsymbol{u} = \boldsymbol{\varepsilon} + \rho \boldsymbol{W}\boldsymbol{\varepsilon}, \qquad \text{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}, \qquad \text{var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}.$$

Let us assume that $\boldsymbol{W}$ admits at least one (real) negative eigenvalue,[19] and let $\lambda_{\min}$ denote the smallest negative eigenvalue of $\boldsymbol{W}$. The largest interval of values of $\rho$ containing the origin such that the model is invertible, i.e., $\boldsymbol{I} + \rho \boldsymbol{W}$ is invertible, is $(-\lambda_{\max}^{-1}, -\lambda_{\min}^{-1})$. The results in Section 3.2 can be extended to show that, in the context of a SMA(1) model, the limiting power of a test for $\rho = 0$ does not necessarily achieve full power as $\rho \to -\lambda_{\min}^{-1}$.

For our purposes, the main difference from the case of a SAR(1) process is that, for a SMA(1) process, $\boldsymbol{\Sigma}(\rho)$, rather than $\boldsymbol{\Sigma}^{-1}(\rho)$, is well-defined and singular at $\rho = a$ (with $a = \lambda_{\max}^{-1}$ for a SAR(1) process, $a = -\lambda_{\min}^{-1}$ for a SMA(1) process). Now, a model with singular $\boldsymbol{\Sigma}(a)$ tends, as $\rho \to a$, to be supported on the subspace of $\mathbb{R}^n$ orthogonal to the nullspace of $\boldsymbol{\Sigma}(a)$.[20] Let us denote such a subspace by $\Lambda$, and its dimension by $\dim(\Lambda)$. Then, any critical region for $\rho = 0$ has vanishing limiting power if its intersection with $\Lambda$ has $\dim(\Lambda)$-dimensional Lebesgue measure zero. For a SMA(1) process, $\Lambda$ is the orthogonal complement of the eigenspace of $\boldsymbol{W}$ associated to $\lambda_{\min}$, and hence $\dim(\Lambda) = n - m_{\min}(\boldsymbol{W})$, where $m_{\min}(\boldsymbol{W})$ denotes the geometric multiplicity of $\lambda_{\min}$. Recall that the limiting support of a SAR(1) model is one-dimensional. On the other hand the limiting support $\Lambda$ of a SMA(1) model has generally higher dimension, because $n - m_{\min}(\boldsymbol{W}) > 1$, except for very special cases. Since a critical region has vanishing limiting power when it does not intersect (almost surely) the limiting support, it can be argued that the zero limiting power problem is more relevant for a SAR(1) process than for a SMA(1) process.

## 4. UNBIASEDNESS AND MONOTONICITY

So far, we have analyzed power properties of tests for problem (2) as $\rho$ approaches the right extreme of $\Psi^+$. In Section 4, we turn to global power properties of the

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

tests, i.e., properties that hold for any $\rho \in \Psi^+$. One crucial property is unbiasedness. A stronger property is that the power function of the test is monotonic in $\rho$. We already know that tests for our testing problem may not satisfy such properties, because a zero limiting power implies both biasedness and nonmonotonicity. Below, we study conditions that guarantee unbiasedness and monotonicity. The conditions are not necessary, but, as we shall see, (i) they are important to understand the structure of the testing problem under analysis; and (ii) in the case of spatial autoregressions, they admit a simple interpretation. To achieve analytical tractability, we assume Gaussianity, and we focus on LBI and POI tests.

Before focusing on the case of a spatial error model, we study the unbiasedness of LBI and POI tests for a general $\boldsymbol{\Sigma}(\rho)$ in regression model (1). LBI or POI tests for problem (2) are certainly unbiased if $\boldsymbol{X}$ and $\boldsymbol{\Sigma}(\rho)$ are such that a UMPI test exists.[21] As we have already pointed out in Section 2.2, this is a very restrictive condition. We now formulate two conditions that guarantee unbiasedness of LBI and POI tests, even when a UMPI test does not exist. Following Horn and Johnson (1985), a commuting family of matrices is a finite or infinite set of matrices that are pairwise commutative (under standard multiplication).

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

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

For a fixed covariance structure $\boldsymbol{\Sigma}(\rho)$, Condition A may or may not be satisfied. The restriction implied by Condition A in the particular case of a SAR(1) model is clarified in Lemma 1 below. A well-known property of commuting symmetric matrices is that they share the same eigenvectors. Thus, when Condition A holds, Condition B does not depend on $\widetilde{\rho}$. Condition B, in any of its many equivalent formulations, has often been used in the theoretical analysis of regression models with nonspherical errors since Anderson (1948). In applications, Condition B is unlikely to be satisfied exactly, but in some circumstances it may be satisfied approximately; see Durbin (1970) for the case of serial correlation, and the end of this section for the case of spatial autoregressions. There is some evidence in the literature that the power properties of tests for $\rho = 0$ when Condition B holds approximately are similar to the power properties when Condition B holds exactly; e.g., Tillman (1975, p. 971). It is worth remarking that Condition B is trivially satisfied by pure models ($k = 0$).

We denote by $\mathrm{col}^{\perp}(\boldsymbol{X})$ the orthogonal complement of $\mathrm{col}(\boldsymbol{X})$. We are now in a position to prove the following result.

PROPOSITION 3. *Consider testing* $\rho = 0$ *against* $\rho > 0$ *in model (1). Assume that* $\boldsymbol{u}$ *has a Gaussian distribution, and that Conditions A and B hold. Then, LBI and POI tests are unbiased. The unbiasedness is strict except when* $\mathrm{col}^{\perp}(\boldsymbol{X})$ *is*

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

*contained in one of the eigenspaces of* $\boldsymbol{\Sigma}(\rho)$*, in which case the power is* $\alpha$ *for any* $\rho > 0$.

Although they guarantee unbiasedness of the tests considered in Proposition 3, Conditions A and B are not sufficient for the monotonicity in $\rho$ of the power functions of those tests (even when $\boldsymbol{X} = \mathbf{O}$). This is simply because, starting from a $\boldsymbol{\Sigma}(\rho)$ satisfying Condition A, a reparametrization $\rho \to f(\rho)$ may destroy the monotonicity of the power function without causing Condition A to fail. Monotonicity of the power function is a much stronger property than unbiasedness and may or may not be desirable depending on the specification of $\boldsymbol{\Sigma}(\rho)$. In general, it is desirable whenever $\rho$ is interpreted as an autocorrelation parameter, as in a SAR(1) model. Next, we discuss the impact of Conditions A and B on the monotonicity of the power function of tests for autocorrelation in a spatial error model. Such a discussion is relevant also outside a formal hypothesis testing setting, because nonmonotonicity of the power function makes it difficult to interpret the underlying test statistic—the Cliff-Ord statistic, say—as an index of spatial autocorrelation. Indeed, one would expect that the probability of an index of autocorrelation being greater than some constant (i.e., the power of the associated test) is nondecreasing in $\rho$ over the interval $(0, \lambda_{\max}^{-1})$ (as all correlations between pairs of variables $y_i$ and $y_j$).

For a SAR(1) process, Condition A boils down to a condition on $\boldsymbol{W}$. A matrix is said to be *normal* if it commutes with its transpose (e.g., Horn and Johnson, 1985, p. 100).

LEMMA 1. *For a SAR(1) process, Condition A is satisfied if and only if* $\boldsymbol{W}$ *is normal. In particular, Condition A is not satisfied if* $\boldsymbol{W}$ *is a nonsymmetric matrix obtained by row-standardization of a symmetric matrix.*

The requirement of matrix normality is very restrictive also for more general nonsymmetric weights matrices, whereas it is trivially satisfied when $\boldsymbol{W}$ is symmetric. Thus, from a practical point of view, the class of SAR(1) processes satisfying Condition A is essentially equivalent to the class of SAR(1) processes with symmetric $\boldsymbol{W}$. For this reason, we now focus on the case of a symmetric $\boldsymbol{W}$.

PROPOSITION 4. *Consider testing* $\rho = 0$ *against* $\rho > 0$ *in a Gaussian spatial error model with symmetric* $\boldsymbol{W}$*. Assume that Condition B holds (for one value, and hence all values,* $\widetilde{\rho} > 0$*). Then, the power functions of the LBI and POI tests are increasing in* $\rho$*. They are strictly increasing except when* $\mathrm{col}^{\perp}(\boldsymbol{X})$ *is contained in one of the eigenspaces of* $\boldsymbol{W}$*, in which case the power is* $\alpha$ *for any* $\rho > 0$.

Proposition 4 can be related to the analysis in Section 3. Of course, monotonicity of the power function implies that the limiting power cannot be smaller than $\alpha$. In fact, the following stronger result can be proved.

PROPOSITION 5. *Consider testing* $\rho = 0$ *against* $\rho > 0$ *in a spatial error model with symmetric* $\boldsymbol{W}$ *by means of the Cliff-Ord test or a test (8). Assume that*

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

*Condition B holds (for one value, and hence all values $\widetilde{\rho} > 0$), and that the critical value $c$ is not equal to $T(\boldsymbol{f}_{\max})$, where $T(\boldsymbol{y})$ is the test statistic. Then, the limiting power is in $(\alpha, 1)$ if $\boldsymbol{f}_{\max} \in \mathrm{col}(\boldsymbol{X})$ and $\mathrm{col}^{\perp}(\boldsymbol{X})$ is not contained in any of the eigenspaces of $\boldsymbol{W}$; it is 1 if $\boldsymbol{f}_{\max} \notin \mathrm{col}(\boldsymbol{X})$; and it is $\alpha$ otherwise.*

Observe that Proposition 5 does not require Gaussianity, so that the Cliff-Ord test or a test (8) are not necessarily LBI and POI. Under the conditions stated in the proposition, for the Cliff-Ord test and tests (8) to have full limiting power it is sufficient that $\boldsymbol{f}_{\max} \notin \mathrm{col}(\boldsymbol{X})$. It is worth pointing out that Proposition 5, as Propositions 3 and 4, holds for any $\alpha$. Extensions of these three propositions to a spatial lag model are possible, along the same lines as in Section 3.5.2.

We conclude Section 4 by describing cases in which Condition B holds in applications of spatial autoregressions, exactly or at least approximately. The CAR(1) model specified by equations (1) and (5) satisfies Condition B exactly when the mean is assumed to be unknown but constant across observations, and $\boldsymbol{W}$ is the row-standardized version of a symmetric matrix.[22] If other regressors are included alongside the intercept in a CAR(1) or SAR(1) model, it is unlikely that Condition B is satisfied, unless $\boldsymbol{W}$ is symmetric and the number of eigenspaces of $\boldsymbol{W}$ (and hence of $\boldsymbol{\Sigma}(\rho)$) is small relative to $n$. This typically occurs when $\boldsymbol{W}$ is invariant under a large group of permutations of its index set (see Biggs, 1993). For example, a weights matrix $\boldsymbol{W}$ with constant off-diagonal entries and zero diagonal entries has only two eigenspaces: the line spanned by $\boldsymbol{\iota}$ and the hyperplane orthogonal to it (weights matrices of this type have been considered, for instance, in Kelejian and Prucha, 2002, and Baltagi, 2006). Hence, for such a $\boldsymbol{W}$, Condition B is satisfied if the entries of each regressor other than the intercept sum to zero.

Turning to the circumstances when Condition B can be expected to hold at least approximately, let us consider, for simplicity, a spatial error model with symmetric $\boldsymbol{W}$ and with only one regressor, denoted by $\boldsymbol{x} = (x_1, \ldots, x_n)'$. Call $j$ a neighbor of $i$ if $(\boldsymbol{W})_{ij} > 0$, and let $\bar{x}_i := \sum_{j \neq i} (\boldsymbol{W})_{ij} x_j$. For a given $\boldsymbol{W}$, the ratio $x_i / \bar{x}_i$ may be regarded as a measure of "similarity" between unit $i$ and its neighbors (as far as $\boldsymbol{x}$ is concerned). Now, in SAR(1) models with symmetric $\boldsymbol{W}$ the eigenvectors of $\boldsymbol{\Sigma}(\rho)$ are the same as those of $\boldsymbol{W}$, so Condition B is met if and only if $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{W}$, i.e., $\boldsymbol{W}\boldsymbol{x} = \lambda \boldsymbol{x}$, for some scalar $\lambda$. Since the $i$th row of the equation $\boldsymbol{W}\boldsymbol{x} = \lambda \boldsymbol{x}$ is $\bar{x}_i = \lambda x_i$, for $i = 1, \ldots, n$, it follows that Condition B is equivalent to the condition that $x_i / \bar{x}_i$ does not depend on $i$. This suggests that Condition B is approximately met, and hence the power of LBI and POI tests has desirable properties when $\boldsymbol{x}$ is such that the degree of similarity between $i$ and its neighbors does not vary substantially with $i$.[23] On the contrary, Condition B is far from being satisfied when different clusters of neighbors have very different degrees of similarity.

## 5. CONCLUSION

This paper has investigated some exact properties of tests for spatial autocorrelation in the context of a linear regression model. We have mainly focused on

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

the case when the disturbances follow a spatial autoregressive process. We have studied the limiting power of tests as the autocorrelation parameter goes to the right boundary of the parameter space, and we have discussed conditions for unbiasedness and monotonicity of the power function of the tests. The results call for caution in interpreting the outcome of tests for spatial autocorrelation. In some circumstances, it may prove very difficult to detect spatial autocorrelation by means of a certain test when the autocorrelation is in fact large. Our results are also relevant outside a formal hypothesis testing framework, because nonmonotonicity of the power function implies that the underlying test statistic, the Cliff-Ord statistic, say, cannot be properly interpreted as an autocorrelation index.

For a fixed weights matrix and a fixed test, it is possible to characterize the matrices $\boldsymbol{X}$ such that the limiting power vanishes. This will be the object of a separate paper. Another possible extension of our results would be to allow for misspecification of $\boldsymbol{W}$ (cf. Kelejian and Prucha, 2001, p. 225). For instance, it would be of interest to study the power properties of a Cliff-Ord test based on some $\boldsymbol{W}$, when the data generating process is a spatial autoregression based on a different weights matrix.

## NOTES

1. In some noneconomic fields, such as image analysis and agriculture, it is well established that simultaneous or conditional autoregressions often lead to a very large value of the autocorrelation parameter (e.g., Besag and Kooperberg, 1995; Bhattacharyya, Richardson, and Franklin, 1997). When data are observed over a regular lattice, this is usually interpreted as an indication of a type of nonstationarity similar to that due to near unit roots in time series. Extensions of this concept of nonstationarity have been attempted also for the case of irregular lattices (e.g., Fingleton, 1999).

2. This can be easily seen writing $(\boldsymbol{I}-\rho\boldsymbol{W})^{-1}=\sum\rho^r\boldsymbol{W}^r$, for $|\rho|<\lambda_{\max}^{-1}$. The expansion shows that when $\rho\in(0,\lambda_{\max}^{-1})$, $\operatorname{cov}(y_i,y_j)>0$ for any $i,j=1,\ldots,n$, and that when $\rho\in(-\lambda_{\max}^{-1},0)$, the covariances may be positive or negative, but not all of them are positive in any left neighborhood of $\lambda_{\max}^{-1}$.

3. One interesting exception is a CAR(1) model satisfying Condition B of Section 4; see King (1988, p. 187).

4. For example, suppose that the distribution of $\boldsymbol{u}$ is elliptically symmetric. Then, according to (6), under $H_0$ $\boldsymbol{v}$ is uniformly distributed on a hemisphere. It follows that in the presence of any test statistic that can be expressed as a quadratic form in $\boldsymbol{v}$ (as, for instance, in (7) and (8)), critical values can be obtain by resorting to one of the many numerical approximations available for the distribution of a quadratic form in a vector uniformly distributed on a sphere.

5. Similarly, in the case of an AR(1) model, a test based on the serial correlation coefficient for $\hat{\boldsymbol{u}}$ is LBI and the Durbin-Watson test is approximately LBI; e.g., Kariya (1988).

6. For a proof that an LR test based on the full likelihood of $\boldsymbol{y}$ is invariant, see, e.g., Cox and Hinkley (1974, p. 173). The restricted LR test is invariant by definition.

7. One exception is a symmetric $k$-nearest-neighbors weights matrix. A $k$-nearest-neighbors weights matrix is a $(0,1)$ matrix specified by taking the same number, $k$, of neighbors for each unit. In this case, $\boldsymbol{f}_{\max}$ is a vector of identical entries, and hence it belongs to $\operatorname{col}(\boldsymbol{X})$ whenever the regression contains an intercept.

8. The results to follow can be trivially extended to study the limit of the power as $\rho\to\infty$ when $\Psi^+=(0,\infty)$, provided that we intepret $\boldsymbol{\Sigma}(a)$ as $\lim_{\rho\to\infty}\boldsymbol{\Sigma}(\rho)$.

9. A critical region $\Phi$ is invariant if $\boldsymbol{y}\in\Phi$ implies $\gamma\boldsymbol{y}+\boldsymbol{X\delta}\in\Phi$, for any $\gamma\in\mathbb{R}\backslash\{0\}$ and any $\boldsymbol{\delta}\in\mathbb{R}^k$. Thus, if $\Phi$ is invariant, then $\operatorname{col}(\boldsymbol{X})\in\operatorname{bd}(\Phi)$ (i.e., any $n$-ball centered at some point $\boldsymbol{y}\in\operatorname{col}(\boldsymbol{X})$

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

contains at least one point in $\Phi$ and at least one point not in $\Phi$), since otherwise $\alpha$ would be either 0 or 1.

10. Here it is irrelevant whether $\boldsymbol{W}$ refers to a model before or after normalization to $\boldsymbol{\Sigma}(0) = \boldsymbol{I}$ imposed in Section 2.1, because the condition $\boldsymbol{f}_{\max} \in \mathrm{col}(\boldsymbol{X})$ is invariant under any invertible linear transformation of $\boldsymbol{y}$, when $\boldsymbol{y}$ follows a spatial error model.

11. Analyzing the circumstances in which power is low but positive goes beyond the scope of this paper.

12. Here and throughout, we do not distinguish notationally between a random variable and its realizations.

13. Since the Cliff-Ord test statistic can be expressed as a ratio of quadratic forms in regression errors, such conditions reduce, in the case of a Gaussian SAR(1) model with symmetric $\boldsymbol{W}$, to the condition stated in Krämer (2005).

14. A practical recommendation, thus, is to always check that the conclusion of a single test is robust over different tests.

15. Because of its invariance property, the power of the Cliff-Ord test depends on $\boldsymbol{X}$ only through $\mathrm{col}(\boldsymbol{X})$. Thus, it would be natural to draw $\boldsymbol{X}$ from $\mathrm{N}(\boldsymbol{O}, \boldsymbol{I}_n \otimes \boldsymbol{I}_k)$, as this would imply that $\mathrm{col}(\boldsymbol{X})$ is uniformly distributed on the Grassmann manifold $G_{k,n}$ (see James, 1954). In our simulations, we have modified such a distribution to take into account the fact that, in practice, an intercept is always included in the regression.

16. As $\rho \to \lambda_{\max}^{-1}$, $\mathrm{corr}(y_i, y_j) \to 1$, for any $i, j$ and $\boldsymbol{W}$. This follows easily from observing that: (i) a SAR(1) model tends, as $\rho \to \lambda_{\max}^{-1}$, to be concentrated on a one-dimensional subspace of the sample space (see Section 3.2.2); (ii) $\mathrm{corr}(y_i, y_j) > 0$ when $\rho > 0$.

17. By Corollary 1, when $\boldsymbol{f}_{\max} \in \mathrm{bd}(\Phi)$, $\alpha^*$ is always zero and hence uninformative. In order to study power when $\boldsymbol{f}_{\max} \in \mathrm{bd}(\Phi)$, one could define $\alpha^*$ 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.

18. Recall from Section 2.1 that the heteroskedasticity induced by a fixed $\boldsymbol{V} \neq \boldsymbol{I}$ can be eliminated by normalizing to $\boldsymbol{V} = \boldsymbol{I}$, and hence does not pose any problem from the point of view of applying the results in Section 3.2. Small (1993) considers the limiting power of the Durbin-Watson test when the innovation variance matrix of an AR(1) model is nonspherical but fixed.

19. Such an assumption is satisfied in virtually all weights matrices used in applications. In particular, it is certainly satisfied when $\boldsymbol{W}$ is symmetric or, more generally, similar to a symmetric matrix. In the latter case, which occurs for instance when $\boldsymbol{W}$ is row-standardized, all eigenvalues of $\boldsymbol{W}$ are real, and consequently one of them must be negative, because $\mathrm{tr}(\boldsymbol{W}) = 0$ by assumption and $\lambda_{\max} > 0$ by Theorem A.2

20. This can be shown by straightforward extension of the proof of Theorem 1. Another example of an econometric model such that $\boldsymbol{\Sigma}(\rho)$ tends to be singular as $\rho \to a$ is a fractionally integrated white noise, with $\rho$ being the differencing parameter and $a = 1/2$ (see Kleiber and Krämer, 2005).

21. If a UMPI test exists, then the power function of POI and LBI tests is the envelope $\pi_\rho(\rho)$. But, as long as $\boldsymbol{\Sigma}(\rho_1) = \boldsymbol{\Sigma}(\rho_2)$ implies $\rho_1 = \rho_2$, we have $\pi_\rho(\rho) > \alpha$ for any $\rho > 0$, by Theorem 1 in Kadiyala (1970), or, more directly, by Corollary 3.2.1 in Lehmann and Romano (2005).

22. To see this, write $\boldsymbol{W} = \boldsymbol{D}^{-1}\boldsymbol{A}$, where $\boldsymbol{A}$ is a symmetric weights matrix and $\boldsymbol{D}$ is the diagonal matrix with $(\boldsymbol{D})_{ii} = \sum_{j=1}^{n} (\boldsymbol{A})_{ij}$, $i = 1, \ldots, n$. On setting $\boldsymbol{L} = \boldsymbol{D}^{-1}$ and normalizing to $\boldsymbol{\Sigma}(0) = \boldsymbol{I}$, the mean of the model becomes proportional to $\boldsymbol{D}^{1/2}\boldsymbol{\iota}$, and the variance matrix becomes $\sigma^2\boldsymbol{\Sigma}(\rho) = \sigma^2\left(\boldsymbol{I} - \rho\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}\right)^{-1}$. Since $\boldsymbol{\iota}$ is an eigenvector of $\boldsymbol{D}^{-1}\boldsymbol{A}$, it follows that $\boldsymbol{D}^{1/2}\boldsymbol{\iota}$ is an eigenvector of $\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$ and hence of $\boldsymbol{\Sigma}(\rho)$, which implies that Condition B is satisfied.

23. It should be noted that the fact that $x_i/\bar{x}_i$ is not very sensitive to $i$ does not imply that $\boldsymbol{x}$ is highly autocorrelated. To see this, suppose, for simplicity, that $\mathrm{E}(\boldsymbol{x}) = \boldsymbol{0}$, so that the Cliff-Ord statistic associated to $\boldsymbol{x}$ is $\boldsymbol{x}'\boldsymbol{W}\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{x}$. Then, under Condition B, $\boldsymbol{x} = \boldsymbol{f}_j(\boldsymbol{W})$ and $\boldsymbol{x}'\boldsymbol{W}\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{x} = \lambda_j(\boldsymbol{W})$ for some $j = 1, \ldots, n$. Note that $\lambda_1(\boldsymbol{W}) \leq \boldsymbol{x}'\boldsymbol{W}\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{x} \leq \lambda_n(\boldsymbol{W})$, so whether $\boldsymbol{x}$ is highly autocorrelated or not (according to the Cliff-Ord statistic) depends on $j$.

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

## REFERENCES

Anderson, T.W. (1948) On the theory of testing serial correlation. *Skandinavisk Aktuarietidskrift* 31, 88–116.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models.* Kluwer.

Baltagi, B.H. (2006) Random effects and spatial autocorrelation with equal weights. *Econometric Theory* 22, 973–984.

Bartels, R. (1992) On the power function of the Durbin-Watson test. *Journal of Econometrics* 51, 101–112.

Bell, K.P. & N.E. Bockstael (2000) Applying the generalized-moments estimation approach to spatial problems involving microlevel data. *Review of Economics and Statistics* 82, 72–82.

Berenblut, I.I. & G.I. Webb (1973) A new test for autocorrelated errors in the linear regression model. *Journal of the Royal Statistical Society* B 35, 33–50.

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. & C. Kooperberg (1995) On conditional and intrinsic autoregression. *Biometrika* 82, 733–746.

Bhattacharyya, B.B., G.D. Richardson, & L.A. Franklin (1997) Asymptotic inference for near unit roots in spatial autoregression. *Annals of Statistics* 25, 1709–1724.

Biggs, N.L. (1993) *Algebraic Graph Theory*, 2nd ed. Cambridge University Press.

Case, A. (1991) Spatial patterns in household demand. *Econometrica* 59, 953–966.

Cliff, A.D. & J.K. Ord (1981) *Spatial Processes: Models and Applications*. Pion.

Cox, D.R. & D.V. Hinkley (1974) *Theoretical Statistics.* Chapman & Hall.

Cordy, C.B. & D.A. Griffith (1993) Efficiency of least squares estimators in the presence of spatial autocorrelation. *Communications in Statistics: Simulation and Computation* 22, 1161–1179.

Cressie, N. (1993) *Statistics for Spatial Data*, revised ed. Wiley.

Cvetković, D.M., M. Doob, & H. Sachs (1980) *Spectra of Graphs.* Academic Press.

Dielman, D.E. & R.C. Pfaffenberger (1989) Efficiency of ordinary least squares for linear models with autocorrelation. *Journal of the American Statistical Association* 84, 248.

Durbin, J. (1970) An alternative to the bounds test for testing for serial correlation in least-squares regression. *Econometrica* 38, 422–29.

Fingleton, B. (1999) Spurious spatial regression: Some Monte Carlo results with a spatial unit root and spatial cointegration. *Journal of Regional Science* 39, 1–19.

Florax, R.J.G.M. & T. de Graaff (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*, pp. 29–65. Springer.

Gall, J., G. Pap, & M. van Zuijlen (2004) Maximum likelihood estimator of the volatility of forward rates driven by geometric spatial AR sheet. *Journal of Applied Mathematics* 4, 293–309.

Gantmacher, F.R. (1974) *The Theory of Matrices*, vol. II. Chelsea.

Goldstein, R. (2000) The term structure of interest rates as a random field. *Review of Financial Studies* 13, 365–384.

Golub, G.H. & C.F. Van Loan (1996) *Matrix Computations*, vol. II, 3rd ed. Johns Hopkins University Press.

Hardy, G., J.E. Littlewood, & G. Pólya (1952) *Inequalities*, 2nd ed. Cambridge University Press.

Horn, R. & C.R. Johnson (1985) *Matrix Analysis*. Cambridge University Press.

Huse, C., (2006) Term Structure Modelling with Spatial Dependence and Observable State Variables. Paper presented at the International Workshop on Spatial Econometrics and Statistics, Rome.

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.

Kadiyala, K.R. (1970) Testing for the independence of regression disturbances. *Econometrica* 38, 97–117.

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

Kalbfleisch, J.D. & D.A. Sprott (1970) Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society* B 32, 175–208.
Kariya, T. (1980) Locally robust tests for serial correlation in least squares regression. *Annals of Statistics* 8, 1065–1070.
Kariya, T. (1988) The class of models for which the Durbin-Watson test is locally optimal. *International Economic Review* 29, 167–175.
Kelejian, H.H. & I.R. Prucha (2001) On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219–257.
Kelejian, H.H. & I.R. Prucha (2002) 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Regional Science and Urban Economics* 32, 691–707.
Kelejian, H.H. & I.R. Prucha (2009) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, forthcoming.
Kennedy, D. (1994) The term structure of interest rates as a Gaussian random field. *Mathematical Finance* 4, 247–258.
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. & G.H. Hillier (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. & W. Krämer (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.
Krämer, W. & C. Donninger (1987) Spatial autocorrelation among errors and relative efficiency of OLS in the linear regression model. *Journal of the American Statistical Association* 82, 577–579.
Lee, L.F. (2002) Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory* 18, 252–277.
Lehmann, E.L. & J. Romano (2005) *Testing Statistical Hypotheses*, 3rd ed. Springer.
Militino, A.F., Ugarte, M.D., & L. García-Reinaldos (2004) Alternative models for describing spatial dependence among dwelling selling prices. *Journal of Real Estate Finance and Economics* 29, 193–209.
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. & J.P. LeSage (2002) Semiparametric maximum likelihood estimates of spatial dependence. *Geographical Analysis* 34, 76–90.
Paulauskas, V. (2007) On unit roots for spatial autoregressive models. *Journal of Multivariate Analysis* 98, 209–226.
Pinske, J. & M.E. Slade (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85, 125–154.
Rahman, S. & M.L. King (1997) Marginal-likelihood score-based tests of regression disturbances in the presence of nuisance parameters. *Journal of Econometrics* 82, 81–106.
Small, J.P. (1993) The limiting power of point optimal autocorrelation tests. *Communications in Statistics: Theory and Methods* 22, 2463–2470.
Tillman, J.A. (1975) The power of the Durbin-Watson test. *Econometrica* 43, 959–974.
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.
Whittle, P. (1954) On stationary processes in the plane. *Biometrika* 41, 434–449.

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

Zeisel, H. (1989) On the power of the Durbin-Watson test under high autocorrelation. *Communications in Statistics: Theory and Methods* 18, 3907–3916.

# APPENDIX A: The Weights Matrices

Appendix A discusses the conditions we have imposed on the weights matrices $\boldsymbol{W}$. In Section 2.1, we have assumed that (i) $(\boldsymbol{W})_{ii} = 0$, for $i = 1, \ldots, n$; (ii) $(\boldsymbol{W})_{ij} \geq 0$, for $i, j = 1, \ldots, n$; and (iii) $\boldsymbol{W}$ is an irreducible matrix. Condition (i) is assumed for SAR(1) models only for convenience, whereas it is required for the validity of CAR(1) models (see, e.g., Besag, 1974). Condition (ii) is not required by the definition of the models, but is virtually always satisfied in empirical applications. Moving to condition (iii), let us first define an irreducible matrix (e.g., Gantmacher, 1974, Ch. 13).

DEFINITION A.1. *A square matrix $\boldsymbol{Q}$ is said to be reducible if there exists a permutation matrix $\boldsymbol{P}$ such that $\boldsymbol{P}'\boldsymbol{Q}\boldsymbol{P}$ can be written in the form $\begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \mathbf{O} & \boldsymbol{S} \end{bmatrix}$, where $\boldsymbol{R}$ and $\boldsymbol{S}$ are square matrices; otherwise, $\boldsymbol{Q}$ is said to be irreducible.*

Irreducibility of $\boldsymbol{W}$ is a natural assumption in a spatial context, because it is equivalent to the condition that the graph with adjacency matrix $\boldsymbol{W}$ (that is, the graph with $n$ vertices and an edge from vertex $i$ to vertex $j$ if and only $(\boldsymbol{W})_{ij} \neq 0$) has a path from any vertex $i$ to any vertex $j$ (see, e.g., Cvetković, Doob, and Sachs, 1980, p. 18). The weights matrices $\boldsymbol{W}$ used in spatial autoregressions typically satisfy this property. In some cases, reducible block diagonal weights matrices with irreducible blocks are used; e.g., Case (1991). Extensions to cover such matrices are straightforward, but not pursued in the present paper for the sake of simplicity. More generally, our analysis can be easily extended to cover the class of weights matrices such that their spectral radius is an eigenvalue with algebraic multiplicity one; by Theorem A.2 below, such a class is larger than the class of matrices satisfying conditions (ii) and (iii). Observe that condition (iii) implies that (noncircular) AR(1) models are not in our class of SAR(1) processes: since time dependence is specified unilaterally, the matrix $\boldsymbol{W}$ necessary to write an AR(1) process as in equation (3) would be triangular and hence reducible.

Assumptions (ii) and (iii) have the advantage of making the following result available (e.g., Horn and Johnson, 1985, Thm. 8.4.4).

THEOREM A.2 (Perron-Frobenius Theorem). *An entrywise nonnegative and irreducible square matrix always has a positive eigenvalue that has algebraic multiplicity one, is associated to an entrywise positive eigenvector, and is nonsmaller than the moduli of all other eigenvalues.*

In Section 2.1, $\lambda_{\max}$ has been defined as the largest positive eigenvalue of $\boldsymbol{W}$. By Theorem A.2, it would have been equivalent to define it as the spectral radius of $\boldsymbol{W}$. In addition, Theorem A.2 implies that (a) $\lambda_{\max}$ has algebraic (and hence geometric) multiplicity 1; (b) there exists a unique entrywise positive and normalized (according to the Euclidean norm) eigenvector associated to $\lambda_{\max}$. Such an eigenvector is denoted by $\boldsymbol{f}_{\max}$. It is worth pointing out that the value $\rho = \lambda_{\max}^{-1}$ can be interpreted as the analog of a *unit root* in an AR(1) model (e.g., Fingleton, 1999; Paulauskas, 2006).

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

# APPENDIX B: Definition of Invariant Tests

Let $F_{\boldsymbol{X}}$ be the group of transformations $\boldsymbol{y} \to \gamma \boldsymbol{y} + \boldsymbol{X\delta}$, for $\gamma \in \mathbb{R}\backslash\{0\}$ and $\boldsymbol{\delta} \in \mathbb{R}^k$, and let $F_{\boldsymbol{X}}^+$ be the smaller group obtained when $\gamma > 0$. In the present paper, invariant tests are defined with respect to $F_{\boldsymbol{X}}$, as, for instance, in Berenblut and Webb (1973), because this simplifies the statement of our results. Some authors (e.g., King, 1988) define invariance of tests for autocorrelation in linear regression with respect to $F_{\boldsymbol{X}}^+$. The distinction between invariance under $F_{\boldsymbol{X}}$ and invariance under $F_{\boldsymbol{X}}^+$ is not substantive, because tests that are invariant under $F_{\boldsymbol{X}}^+$ but not under $F_{\boldsymbol{X}}$ are never used in practice. Under $F_{\boldsymbol{X}}^+$, a maximal invariant is $\boldsymbol{Cy}/\|\boldsymbol{Cy}\|$, and invariant critical regions are defined on the unit $(n-k)$-sphere (rather than on a hemisphere, as in Section 2.2). The class of critical regions that are invariant under $F_{\boldsymbol{X}}$ is equivalent to the class of critical regions that are invariant under $F_{\boldsymbol{X}}^+$ and are centrally symmetric (i.e., they contain a vector $\boldsymbol{t} \in \mathbb{R}^n$ if and only if they contain $-\boldsymbol{t}$).

# APPENDIX C: The Condition in Proposition 1

In this appendix we discuss the necessary and sufficient condition in Proposition 1. The condition is that $\boldsymbol{f}_{\max}$ is an eigenvector of $\boldsymbol{W}'$, or, equivalently, that $\lambda_{\max}$ is perfectly well-conditioned (see, e.g., Golub and Van Loan, 1996, p. 323). The requirement that an eigenvalue is perfectly well-conditioned is trivially satisfied in the case of a symmetric matrix, but is well-known to be very restrictive otherwise. Observe that, for any given choice of the neighborhood structure of a set of observational units (i.e., any choice of the ordered pairs $(i, j)$ such that $(\boldsymbol{W})_{ij} = 0$), it is always possible to set the nonzero elements of $\boldsymbol{W}$ in such a way that $\lambda_{\max}$ is perfectly well-conditioned. (Start from a nonsymmetric matrix $\boldsymbol{W}$ and apply the similarity transformation $\boldsymbol{P}^{-1}\boldsymbol{WP}$, where $\boldsymbol{P}$ is a diagonal matrix with $(\boldsymbol{P})_{ii} = [(\boldsymbol{f}_{\max})_i/(\boldsymbol{l}_{\max})_i]^{1/2}$ and $\boldsymbol{l}_{\max}$ denotes the left eigenvector of $\boldsymbol{W}$ associated to $\lambda_{\max}$. Clearly, the left and the right eigenvectors of $\boldsymbol{P}^{-1}\boldsymbol{WP}$ associated to $\lambda_{\max}$ are identical, and hence $\lambda_{\max}$ is perfectly well-conditioned.) In general, however, the choice of weights yielding a perfectly well-conditioned $\lambda_{\max}$ does not correspond to any relevant notion of distance among the observational units. The restrictiveness of the condition that $\boldsymbol{f}_{\max}$ is an eigenvector of $\boldsymbol{W}'$ is very transparent in the case of a row-stochastic $\boldsymbol{W}$, as emphasized by the following two lemmas. A matrix is said to be row-stochastic if all its row-sums are 1, doubly stochastic if both itself and its transpose are row-stochastic.

LEMMA C.1. *Assume that $\boldsymbol{W}$ is row-stochastic. Then, $\boldsymbol{f}_{\max}$ is an eigenvector of $\boldsymbol{W}'$ if and only if $\boldsymbol{W}$ is doubly stochastic.*

**Proof.** If $\boldsymbol{W}$ is row-stochastic, $\boldsymbol{f}_{\max}$ has identical entries. Hence, when $\boldsymbol{W}$ is row-stochastic, $\boldsymbol{f}_{\max}$ is an eigenvector of $\boldsymbol{W}'$ if and only if the columns of $\boldsymbol{W}$, as its rows, sum to 1, that is, if and only if $\boldsymbol{W}$ is doubly stochastic. ■

Clearly, the requirement that a (nonsymmetric) weights matrix is doubly stochastic is very restrictive (this can formally 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 Pace and LeSage (2002) are not relevant here because they are symmetric.

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

LEMMA C.2. *If $\boldsymbol{W}$ is obtained by row-standardization of a symmetric matrix and is nonsymmetric, then $\boldsymbol{f}_{\max}$ is not an eigenvector of $\boldsymbol{W}'$.*

**Proof.** Assume that $\boldsymbol{W}$ is nonsymmetric and can be written as $\boldsymbol{D}^{-1}\boldsymbol{A}$, where $\boldsymbol{A}$ is some symmetric and nonnegative matrix and $\boldsymbol{D} \neq \boldsymbol{I}$ is the diagonal matrix with $(\boldsymbol{D})_{ii} = \sum_{j=1}^{n} (\boldsymbol{A})_{ij}, i = 1, \ldots, n$ ($\boldsymbol{D}$ is invertible because $\boldsymbol{W}$, and hence $\boldsymbol{A}$, is irreducible). By Lemma C.1, to prove that $\boldsymbol{f}_{\max}$ is not an eigenvector of $\boldsymbol{W}'$, it is sufficient to show that $\boldsymbol{D}^{-1}\boldsymbol{A}$ cannot be doubly stochastic. Suppose that $\boldsymbol{D}^{-1}\boldsymbol{A}$ is doubly stochastic. Then, $(\boldsymbol{D}^{-1}\boldsymbol{A})'\boldsymbol{\iota} = \boldsymbol{\iota}$. Premultiplying by $\boldsymbol{D}^{-1}$, we obtain $\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\boldsymbol{\iota} = \boldsymbol{D}^{-1}\boldsymbol{\iota}$, which implies that $\boldsymbol{D}^{-1}\boldsymbol{\iota}$ is an eigenvector of $\boldsymbol{D}^{-1}\boldsymbol{A}$ associated to the eigenvalue 1. But $\boldsymbol{\iota}$ is another eigenvector of $\boldsymbol{D}^{-1}\boldsymbol{A}$ associated to the eigenvalue 1. Observe that the eigenvalue 1 has algebraic multiplicity one by Theorem A.2 because it is the largest positive eigenvalue of $\boldsymbol{D}^{-1}\boldsymbol{A}$ (e.g., Gantmacher 1974, p. 83). Since $\boldsymbol{D}^{-1}\boldsymbol{\iota}$ and $\boldsymbol{\iota}$ are linearly independent, it is impossible that they are both eigenvectors of $\boldsymbol{D}^{-1}\boldsymbol{A}$ associated to 1, completing the proof. ■

In view of Proposition 1, Lemma C.2 establishes that whenever the weights matrix of a SAR(1) model is obtained by row-standardization of a symmetric matrix and is nonsymmetric, there are always values of $\alpha$ such that the limiting power of the Cliff-Ord test or of a test (8) vanishes, even if the model is not contaminated by regressors.

# APPENDIX D: Proofs for Section 3

First, we state four auxiliary lemmas. The first one is a standard result from matrix algebra (see, e.g., Horn and Johnson, 1985).

LEMMA D.1 (Rayleigh-Ritz Theorem). *For a $q \times q$ symmetric matrix $\boldsymbol{Q}$, $\lambda_1(\boldsymbol{Q})\boldsymbol{x}'\boldsymbol{x} \leq \boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x} \leq \lambda_q(\boldsymbol{Q})\boldsymbol{x}'\boldsymbol{x}$, for all $\boldsymbol{x} \in \mathbb{R}^q$. The equalities on the left and on the right are attained if and only if $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{Q}$ associated to, respectively, $\lambda_1(\boldsymbol{Q})$ and $\lambda_q(\boldsymbol{Q})$.*

LEMMA D.2. *Consider a model $G(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2[(\boldsymbol{I} - \rho\boldsymbol{W}')(\boldsymbol{I} - \rho\boldsymbol{W})]^{-1})$, where $G(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ denotes some multivariate distribution with mean $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Gamma}$. When an invariant critical region for testing $\rho = 0$ against $\rho > 0$ is in form (9), and is such that $\boldsymbol{f}_{\max}$ is not contained in its boundary,*

$$\alpha^* = \Pr(T(\boldsymbol{z}) > T(\boldsymbol{f}_{\max});\ \boldsymbol{z} \sim G(\boldsymbol{0}, \boldsymbol{I})). \tag{D.1}$$

**Proof.** Consider an invariant critical region $\Phi$ in form (9), and assume that $\boldsymbol{f}_{\max} \notin \mathrm{bd}(\Phi)$. According to Corollary 1, the limiting power of $\Phi$ is 1 if $T(\boldsymbol{f}_{\max}) > c$. Thus, by Definition 1, $\alpha^* = \Pr(T(\boldsymbol{y}) > T(\boldsymbol{f}_{\max}); \boldsymbol{y} \sim G(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{I}))$. Such an expression simplifies to (D.1), due to the invariance of $\Phi$ with respect to the transformations $\boldsymbol{y} \to \gamma\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\delta}$, with $\gamma \in \mathbb{R}\backslash\{0\}$ and $\boldsymbol{\delta} \in \mathbb{R}^k$. ■

LEMMA D.3. *Consider a test that, in the context of a spatial error model with symmetric $\boldsymbol{W}$, rejects $\rho = 0$ for small values of a statistic $\boldsymbol{v}'\boldsymbol{B}\boldsymbol{v}$, where $\boldsymbol{B}$ is an $(n-k) \times (n-k)$ known symmetric matrix that does not depend on $\alpha$, and $\boldsymbol{v}$ is as defined in Section*

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

*2.2. Provided that* $\boldsymbol{f}_{\max} \notin \mathrm{bd}(\Phi)$, $\alpha^* = 0$ *if and only if* $\boldsymbol{C}\boldsymbol{f}_{\max} \in E_1(\boldsymbol{B})$, *and* $\alpha^* = 1$ *if and only if* $\boldsymbol{C}\boldsymbol{f}_{\max} \in E_{n-k}(\boldsymbol{B})$.

**Proof.** By Lemma D.2, we have that, when $\boldsymbol{C}\boldsymbol{f}_{\max} \neq \boldsymbol{0}$, $\alpha^* = 0$ if and only if $\boldsymbol{C}\boldsymbol{f}_{\max} / \left\|\boldsymbol{C}\boldsymbol{f}_{\max}\right\| = \arg\max_{\boldsymbol{v}'\boldsymbol{v}=1}\{\boldsymbol{v}'\boldsymbol{B}\boldsymbol{v}\}$, and $\alpha^* = 1$ if and only if $\boldsymbol{C}\boldsymbol{f}_{\max} / \left\|\boldsymbol{C}\boldsymbol{f}_{\max}\right\| = \arg\min_{\boldsymbol{v}'\boldsymbol{v}=1}\{\boldsymbol{v}'\boldsymbol{B}\boldsymbol{v}\}$. The result follows by Lemma D.1 ■

LEMMA D.4. *For any weights matrix* $\boldsymbol{W}$, $\mathrm{rank}((\boldsymbol{I} - \lambda_{\max}^{-1}\boldsymbol{W}')(\boldsymbol{I} - \lambda_{\max}^{-1}\boldsymbol{W})) = n - 1$.

**Proof.** The lemma follows from observing that, by Theorem A.2, $\mathrm{rank}(\boldsymbol{I} - \lambda_{\max}^{-1}\boldsymbol{W}') = n - 1$, and that $\mathrm{rank}(\boldsymbol{Q}'\boldsymbol{Q}) = \mathrm{rank}(\boldsymbol{Q})$ for any matrix $\boldsymbol{Q}$ (e.g., Horn and Johnson, 1985, p. 13). ■

Next, we prove all results in Section 3.

**Proof of Theorem 1.** For any $0 < \rho \leq a$, let $\boldsymbol{\varepsilon} := \boldsymbol{\Sigma}^{-1/2}(\rho)\boldsymbol{u}$, where $\boldsymbol{\Sigma}^{1/2}(\rho)$ is the (unique) square root of $\boldsymbol{\Sigma}(\rho)$, and let $g(\boldsymbol{\varepsilon})$ denote the density of $\boldsymbol{\varepsilon}$. Rewrite model (1) as $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\Sigma}^{1/2}(\rho)\boldsymbol{\varepsilon}$. By the change of variables theorem, the density of $\boldsymbol{y}$ is

$$f(\boldsymbol{y}) = \left|\det(\boldsymbol{\Sigma}^{-1/2}(\rho))\right| g(\boldsymbol{\Sigma}^{-1/2}(\rho)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})). \tag{D.2}$$

Expression (D.2) is useful to establish the limiting behavior of $f(\boldsymbol{y})$ as $\rho \to a$. Observe that the term $\left|\det(\boldsymbol{\Sigma}^{-1/2}(\rho))\right|$ vanishes as $\rho \to a$, because $\mathrm{rank}(\boldsymbol{\Sigma}^{-1}(a)) = n - 1$. Next, recall that in Section 2.1 we have assumed that the density of $\boldsymbol{u}$ is larger at $\boldsymbol{0}$ than anywhere else. Hence, as $\rho \to a$, the term $g(\boldsymbol{\Sigma}^{-1/2}(\rho)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}))$ is maximized at the points $\boldsymbol{y}$ such that $\boldsymbol{\Sigma}^{-1/2}(a)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) = \boldsymbol{0}$, or, equivalently, $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \in \mathcal{N}(\boldsymbol{\Sigma}^{-1/2}(a))$, where $\mathcal{N}(\boldsymbol{Q})$ denotes the nullspace of a matrix $\boldsymbol{Q}$. Note that $\mathcal{N}(\boldsymbol{\Sigma}^{-1/2}(\rho)) = \mathcal{N}(\boldsymbol{\Sigma}^{-1}(a)) = E_1(\boldsymbol{\Sigma}^{-1}(a))$. Combining the above observations, we obtain that, as $\rho \to a$, $f(\boldsymbol{y})$ tends to a degenerate density supported on the set $\Lambda := \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} \in E_1(\boldsymbol{\Sigma}^{-1}(a))\}$, i.e., on the translation by $\boldsymbol{X}\boldsymbol{\beta}$ of the subspace $E_1(\boldsymbol{\Sigma}^{-1}(a))$. Now, the limiting power of a critical region for testing $\rho = 0$ is the probability content of that region under $f(\boldsymbol{y})$, as $\rho \to a$. Thus, the limit of the power function depends on the position of the critical region in $\mathbb{R}^n$ relative to $\Lambda$. In particular, if $\mathrm{int}(\Phi) \setminus E_1(\boldsymbol{\Sigma}^{-1}(a))$ has measure zero, then the limiting power must be 1; if $E_1(\boldsymbol{\Sigma}^{-1}(a)) \cap \mathrm{cl}(\Phi)$ has measure zero, then the limiting power must vanish. Such conditions can be simplified, because, by the invariance of $\Phi$ with respect to the transformations $\boldsymbol{y} \to \gamma\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\delta}$, $\gamma \in \mathbb{R}\setminus\{0\}$, $\boldsymbol{\delta} \in \mathbb{R}^k$, we can focus on an arbitrary point in $E_1(\boldsymbol{\Sigma}^{-1}(a))$. The first and the third case in the statement of the theorem are obtained by taking any of the two points identified by $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a))$. To complete the proof, we only need to establish what happens when $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(a)) \in \mathrm{bd}(\Phi)$. In that case, the limiting power depends on how the distribution of $\boldsymbol{y}$ tends to the degenerate distribution on $\Lambda$, which in turn depends on $n$, $\boldsymbol{X}$, and $\boldsymbol{W}$. For a given triplet $n$, $\boldsymbol{X}$, and $\boldsymbol{W}$, the power approaches, as $\rho \to a$, some value in $(0, 1)$; the extremes 0 and 1 are excluded because, by assumption, $0 < \alpha < 1$, and the density of $\boldsymbol{u}$, and hence that of $\boldsymbol{y}$, is continuous in $\rho$. ■

**Proof of Corollary 1.** For a SAR(1) process, $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(\lambda_{\max}^{-1})) = \boldsymbol{f}_{\max}$ because $\boldsymbol{\Sigma}^{-1}(\lambda_{\max}^{-1})\boldsymbol{f}_{\max} = (\boldsymbol{I} - \lambda_{\max}^{-1}\boldsymbol{W}')(\boldsymbol{I} - \lambda_{\max}^{-1}\boldsymbol{W})\boldsymbol{f}_{\max} = (1 - \lambda_{\max}^{-1}\lambda_{\max})(\boldsymbol{I} - \lambda_{\max}^{-1}\boldsymbol{W}')\boldsymbol{f}_{\max} = 0$. In addition, $\boldsymbol{\Sigma}(\rho)$ is positive definite as $\rho \to \lambda_{\max}^{-1}$, and $\mathrm{rank}(\boldsymbol{\Sigma}^{-1}(\lambda_{\max}^{-1})) = n - 1$ by Lemma D.4. The corollary then follows from Theorem 1. ■

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

**Proof of Proposition 1.** Observe that if $\boldsymbol{f}_{\max}$ is an eigenvector of $\boldsymbol{W}'$, then it must be associated to $\lambda_{\max}$. To see this, call $\varphi$ the eigenvalue of $\boldsymbol{W}'$ associated to $\boldsymbol{f}_{\max}$. Then $\boldsymbol{W}'\boldsymbol{f}_{\max} = \varphi\boldsymbol{f}_{\max}$, and hence $\boldsymbol{f}'_{\max}\boldsymbol{W}'\boldsymbol{f}_{\max} = \varphi$. But since $\boldsymbol{W}\boldsymbol{f}_{\max} = \lambda_{\max}\boldsymbol{f}_{\max}$, it also holds that $\boldsymbol{f}'_{\max}\boldsymbol{W}\boldsymbol{f}_{\max} = \boldsymbol{f}'_{\max}\boldsymbol{W}'\boldsymbol{f}_{\max} = \lambda_{\max}$. Thus $\varphi = \lambda_{\max}$. Now let $\boldsymbol{\Gamma}(\rho) := [(\boldsymbol{I} - \rho\boldsymbol{W}')(\boldsymbol{I} - \rho\boldsymbol{W})]^{-1}$. When $\boldsymbol{X} = \boldsymbol{O}$, the Cliff-Ord test and test (8) reject for small values of, respectively, $\boldsymbol{v}'(-\boldsymbol{W})\boldsymbol{v}$ and $\boldsymbol{v}'\boldsymbol{\Gamma}^{-1}(\bar{\rho})\boldsymbol{v}$. By Lemma D.3 with $\boldsymbol{B} = \boldsymbol{\Gamma}^{-1}(\bar{\rho})$, in order to prove that the limiting power of test (8) is 1 for any $\alpha$, we need to show that $\boldsymbol{W}'\boldsymbol{f}_{\max} = \lambda_{\max}\boldsymbol{f}_{\max}$ is necessary and sufficient for $\boldsymbol{f}_{\max} \in E_n(\boldsymbol{\Gamma}(\bar{\rho}))$. Clearly, if this holds for any $\bar{\rho} > 0$, it holds for $\bar{\rho} \to 0$ too, establishing also the part of the proposition regarding the Cliff-Ord test. Starting from the necessity, we show that if $\boldsymbol{f}_{\max} \in E_n(\boldsymbol{\Gamma}(\bar{\rho}))$, then $\boldsymbol{f}_{\max}$ is an eigenvector of $\boldsymbol{W}'$ (the corresponding eigenvalue being $\lambda_{\max}$ by the above argument). If $\boldsymbol{\Gamma}(\bar{\rho})\boldsymbol{f}_{\max} = \lambda_n(\boldsymbol{\Gamma}(\bar{\rho}))\boldsymbol{f}_{\max}$, then $\boldsymbol{\Gamma}^{-1}(\bar{\rho})\boldsymbol{f}_{\max} = \lambda_n^{-1}(\boldsymbol{\Gamma}(\bar{\rho}))\boldsymbol{f}_{\max}$. From the latter equation we have $(1 - \bar{\rho}\lambda_{\max})(\boldsymbol{I} - \rho\boldsymbol{W}')\,\boldsymbol{f}_{\max} = \lambda_n^{-1}(\boldsymbol{\Gamma}(\bar{\rho}))\boldsymbol{f}_{\max}$, which requires $\boldsymbol{f}_{\max}$ to be an eigenvector of $\boldsymbol{I} - \rho\boldsymbol{W}'$ and hence of $\boldsymbol{W}'$. Turning to the sufficiency, note that if $\boldsymbol{W}'\boldsymbol{f}_{\max} = \lambda_{\max}\boldsymbol{f}_{\max}$, then $\boldsymbol{f}_{\max}$ is an eigenvector of $\boldsymbol{\Gamma}(\bar{\rho})$ for any $\bar{\rho} > 0$. Also, observe that $\boldsymbol{\Gamma}(\bar{\rho})$ is entrywise positive for any $\bar{\rho} > 0$ (see, e.g., Gantmacher, 1974, p. 69). Since an entrywise positive matrix is nonnegative irreducible, it follows, by Theorem A.2, that any vector in $E_n(\boldsymbol{\Gamma}(\bar{\rho}))$ is entrywise positive (or entrywise negative), for any $\bar{\rho} > 0$. But $\boldsymbol{f}_{\max}$ is also entrywise positive by Theorem A.2 applied to $\boldsymbol{W}$. Hence, $\boldsymbol{f}_{\max}$ must be in $E_n(\boldsymbol{\Gamma}(\bar{\rho}))$ for any $\bar{\rho} > 0$, because otherwise, by the symmetry of $\boldsymbol{\Gamma}(\bar{\rho})$, it should be orthogonal to an entrywise positive vector, which is impossible. This completes the proof of the proposition. ■

**Proof of Proposition 2.** By straightforward modification of the argument in the proof of Theorem 1, model (10) tends, as $\rho \to \lambda_{\max}^{-1}$, to be supported on the subspace $E_1(\boldsymbol{\Sigma}^{-1}(\lambda_{\max}^{-1}))$, the only difference with respect to the case of model (1) being that no translation by $\boldsymbol{X}\boldsymbol{\beta}$ is necessary. Thus, for model (10), Theorem 1 does not require invariance of $\Phi$ with respect to $\boldsymbol{y} \to \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\delta}$, $\boldsymbol{\delta} \in \mathbb{R}^k$. The proof is completed as in Corollary 1, because the variance matrix of model (10) is the same as that of a spatial error model. ■

# APPENDIX E: Proofs for Section 4

We need the following four auxiliary lemmas. For the first one, see, e.g., Horn and Johnson (1985).

LEMMA E.1 (Poincaré Separation Theorem). *Let $\boldsymbol{Q}$ be an $n \times n$ symmetric matrix, and $\boldsymbol{C}$ an $(n-k) \times n$ matrix such that $\boldsymbol{C}\boldsymbol{C}' = \boldsymbol{I}_{n-k}$. Then, $\lambda_l(\boldsymbol{Q}) \le \lambda_l(\boldsymbol{C}\boldsymbol{Q}\boldsymbol{C}') \le \lambda_{k+l}(\boldsymbol{Q})$ for $l = 1, \ldots, n-k$.*

LEMMA E.2. *Let $\boldsymbol{R}(\bar{\rho}) := \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\bar{\rho})\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\bar{\rho})$, for any $0 < \bar{\rho} < \lambda_{\max}^{-1}$. Then,*

$$\boldsymbol{v}'\left(\boldsymbol{C}\boldsymbol{\Sigma}(\bar{\rho})\boldsymbol{C}'\right)^{-1}\boldsymbol{v} = \frac{\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}(\bar{\rho})\boldsymbol{R}(\bar{\rho})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}}.$$

**Proof.** Immediate from Lemma 2 of King (1980). ■

LEMMA E.3. *Let $0 \le a_1 \le a_2 \le \ldots \le a_r$, $b_1 \ge b_2 \ge \ldots \ge b_r \ge 0$, and $p_i > 0$, $i = 1, \ldots, r$. Then $\sum_{i=1}^r p_i a_i \sum_{i=1}^r p_i b_i \ge \sum_{i=1}^r p_i \sum_{i=1}^r p_i a_i b_i$, with equality if and only if all $a_i$ or all $b_i$ are equal.*

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

**Proof.** The lemma is a particular case of Tchebychef's inequality (e.g., Hardy, Littlewood, and Pólya, 1952, Thm. 43). ■

LEMMA E.4. *Consider, in the context of a spatial error model, testing $\rho = 0$ against $\rho > 0$ by means of the Cliff-Ord test or of a test (8). Provided that $\boldsymbol{f}_{\max} \notin \text{col}(\boldsymbol{X})$, $\alpha^* = 0$ if $E_{n-k}(\boldsymbol{C}\boldsymbol{\Sigma}(\rho)\boldsymbol{C}')$ does not depend on $\rho$ for $\rho > 0$.*

**Proof.** In this proof, all limits are taken as $\rho \to \lambda_{\max}^{-1}$, and $\boldsymbol{\Sigma}(\rho)$ denotes the variance matrix of a SAR(1) model. Consider the spectral decomposition $\boldsymbol{\Sigma}(\rho) = \sum_{i=1}^{n} \lambda_i(\boldsymbol{\Sigma}(\rho))\, \boldsymbol{f}_i(\boldsymbol{\Sigma}(\rho))\boldsymbol{f}_i'(\boldsymbol{\Sigma}(\rho))$. By Lemma D.4, $\lim \lambda_n(\boldsymbol{\Sigma}(\rho)) = +\infty$ and $-\infty < \lim \lambda_i(\boldsymbol{\Sigma}(\rho)) < +\infty$, for $i = 1, \ldots, n-1$. Thus, $\lim \lambda_n^{-1}(\boldsymbol{\Sigma}(\rho))\boldsymbol{C}\boldsymbol{\Sigma}(\rho)\boldsymbol{C}' = \boldsymbol{C}\left[\lim \boldsymbol{f}_n(\boldsymbol{\Sigma}(\rho))\boldsymbol{f}_n'(\boldsymbol{\Sigma}(\rho))\right]\boldsymbol{C}'$. Observe that $\lim \boldsymbol{f}_n(\boldsymbol{\Sigma}(\rho)) = \boldsymbol{f}_{\max}$ because, as we have established in the proof of Corollary 1, $\boldsymbol{f}_1(\boldsymbol{\Sigma}^{-1}(\lambda_{\max}^{-1})) = \boldsymbol{f}_{\max}$. So,

$$\lim \lambda_n^{-1}(\boldsymbol{\Sigma}(\rho))\boldsymbol{C}\boldsymbol{\Sigma}(\rho)\boldsymbol{C}' = \boldsymbol{C}\boldsymbol{f}_{\max}\, \boldsymbol{f}_{\max}'\boldsymbol{C}'.$$

Let us assume that $\boldsymbol{f}_{\max} \notin \text{col}(\boldsymbol{X})$. Then $\text{rank}(\boldsymbol{C}\boldsymbol{f}_{\max}\boldsymbol{f}_{\max}'\boldsymbol{C}') = 1$. The nonzero eigenvalue of $\boldsymbol{C}\boldsymbol{f}_{\max}\, \boldsymbol{f}_{\max}'\boldsymbol{C}'$ is $\bar{\lambda} := \boldsymbol{f}_{\max}'\boldsymbol{M}_{\boldsymbol{X}}\, \boldsymbol{f}_{\max}$ and is associated to the eigenvector $\boldsymbol{C}\boldsymbol{f}_{\max}$ because

$$(\boldsymbol{C}\boldsymbol{f}_{\max}\, \boldsymbol{f}_{\max}'\boldsymbol{C}')\boldsymbol{C}\boldsymbol{f}_{\max} = \boldsymbol{C}\boldsymbol{f}_{\max}\, \boldsymbol{f}_{\max}'\boldsymbol{M}_{\boldsymbol{X}}\, \boldsymbol{f}_{\max} = \bar{\lambda}\boldsymbol{C}\boldsymbol{f}_{\max}.$$

Since the eigenvalues of a matrix are continuous in the matrix entries, we also have that $\boldsymbol{C}\boldsymbol{f}_{\max}$ spans the one-dimensional eigenspace of $\lim \boldsymbol{C}\boldsymbol{\Sigma}(\rho)\boldsymbol{C}'$ associated to the eigenvalue $\lim[\lambda_n(\boldsymbol{\Sigma}(\rho))\bar{\lambda}] = +\infty$, where the positive sign follows from the fact that $\bar{\lambda} > 0$, by Lemma E.1 So far, we have thus established that if $\boldsymbol{f}_{\max} \notin \text{col}(\boldsymbol{X})$, then $\lim E_{n-k}(\boldsymbol{C}\boldsymbol{\Sigma}(\rho)\boldsymbol{C}')$ is one-dimensional and contains $\boldsymbol{C}\boldsymbol{f}_{\max}$. Now, if $E_{n-k}(\boldsymbol{C}\boldsymbol{\Sigma}(\rho)\boldsymbol{C}')$ does not depend on $\rho$ for $\rho > 0$, it must be spanned by $\boldsymbol{C}\boldsymbol{f}_{\max}$ for any $\rho > 0$. Lemma D.3 with $\boldsymbol{B} = (\boldsymbol{C}\boldsymbol{\Sigma}(\bar{\rho})\boldsymbol{C}')^{-1}$ then implies that, for a test (8), $\alpha^* = 0$. Since this property holds for any $\bar{\rho} > 0$, it also holds for the Cliff-Ord test, completing the proof. ■

Two particular cases that are easily seen to satisfy the condition in Lemma E.4 are: (i) $\boldsymbol{W}$ symmetric and $\boldsymbol{X} = \boldsymbol{O}$; and (ii) $\boldsymbol{W}$ symmetric and $\boldsymbol{f}_{\max} \perp \text{col}(\boldsymbol{X})$. We can now prove all results in Section 4.

**Proof of Proposition 3.** Under Gaussianity, a POI critical region is defined by (8). By Lemma E.2, for any size $\alpha$, its power can be written as

$$\pi_{\bar{\rho}}(\rho) = \Pr\left(\frac{\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}(\bar{\rho})\boldsymbol{R}(\bar{\rho})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}} < c;\ \boldsymbol{y} \sim \text{N}\left(\boldsymbol{X}\beta, \sigma^2\boldsymbol{\Sigma}(\rho)\right)\right). \tag{E.1}$$

Since the the critical region is invariant, (E.1) simplifies to

$$\pi_{\bar{\rho}}(\rho) = \Pr\left(\frac{\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}(\bar{\rho})\boldsymbol{R}(\bar{\rho})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y}} < c;\ \boldsymbol{y} \sim \text{N}(\boldsymbol{0}, \boldsymbol{\Sigma}(\rho))\right).$$

Under Conditions A and B, $\boldsymbol{R}(\bar{\rho}) = \boldsymbol{M}_{\boldsymbol{X}}$ and $\boldsymbol{\Sigma}(\bar{\rho})\boldsymbol{X} = \boldsymbol{X}\boldsymbol{Q}$ for any $\bar{\rho} > 0$ and for some invertible matrix $\boldsymbol{Q}$. It follows that, under the two conditions, the matrices $\boldsymbol{\Sigma}^{-1}(\bar{\rho})$ and $\boldsymbol{M}_{\boldsymbol{X}}$ commute for any $\bar{\rho} > 0$. Hence,

$$\pi_{\bar{\rho}}(\rho) = \Pr\left(\frac{\boldsymbol{z}'\boldsymbol{\Sigma}(\rho)\boldsymbol{\Sigma}^{-1}(\bar{\rho})\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{\Sigma}(\rho)\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z}} < c;\ \boldsymbol{z} \sim \text{N}(\boldsymbol{0}, \boldsymbol{I})\right).$$

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

Under Conditions A and B, the matrix $\boldsymbol{M_X}$ has a zero eigenvalue with eigenspace spanned by the $k$ eigenvectors of $\boldsymbol{\Sigma}(\rho)$ that are in col($\boldsymbol{X}$), and an eigenvalue equal to 1 with eigenspace spanned by the remaining eigenvectors of $\boldsymbol{\Sigma}(\rho)$. Let $H$ denote the set of indexes $i$ of the $n-k$ eigenvalues $\lambda_i(\boldsymbol{\Sigma}(\rho))$ associated to a set of linearly independent eigenvectors of $\boldsymbol{\Sigma}(\rho)$ that are not in col($\boldsymbol{X}$). Note that, when Condition A holds, $H$ does not depend on $\rho$. Under Conditions A and B, the power of a POI critical region can therefore be expressed as

$$\pi_{\bar{\rho}}(\rho) = \Pr\left( \frac{\sum_{i\in H} \lambda_i(\boldsymbol{\Sigma}(\rho))\lambda_i^{-1}(\boldsymbol{\Sigma}(\bar{\rho}))z_i^2}{\sum_{i\in H} \lambda_i(\boldsymbol{\Sigma}(\rho))z_i^2} < c \right), \tag{E.2}$$

and its size as

$$\alpha = \Pr\left( \frac{\boldsymbol{z}'\boldsymbol{\Sigma}^{-1}(\bar{\rho})\boldsymbol{M_X}\boldsymbol{z}}{\boldsymbol{z}'\boldsymbol{M_X}\boldsymbol{z}} < c \right) = \Pr\left( \frac{\sum_{i\in H} \lambda_i^{-1}(\boldsymbol{\Sigma}(\bar{\rho}))z_i^2}{\sum_{i\in H} z_i^2} < c \right). \tag{E.3}$$

Now, by Lemma E.3 with $a_i = \lambda_i(\boldsymbol{\Sigma}(\rho))$, $b_i = \lambda_i^{-1}(\boldsymbol{\Sigma}((\bar{\rho}))$, and $p_i = z_i^2$, for $i \in H$, we have

$$\sum_{i\in H} \lambda_i(\boldsymbol{\Sigma}(\rho))z_i^2 \sum_{i\in H} \lambda_i^{-1}(\boldsymbol{\Sigma}(\bar{\rho}))z_i^2 \geq \sum_{i\in H} z_i^2 \sum_{i\in H} \lambda_i(\boldsymbol{\Sigma}(\rho))\lambda_i^{-1}(\boldsymbol{\Sigma}(\bar{\rho}))z_i^2, \tag{E.4}$$

for any $\boldsymbol{z} \in \mathbb{R}^n$. Rearranging the terms of the above inequality, we obtain that, for any $\bar{\rho} > 0$, any $\rho > 0$, and any $\alpha$, the statistic appearing in expression (E.2) is stochastically larger (e.g., Lehmann and Romano, 2005, p. 70) than that appearing in expression (E.3). We have thus established that $\pi_{\bar{\rho}}(\rho) \geq \alpha$. The latter inequality is strict, except when (E.4) holds with equality, which, by Lemma E.3, occurs if and only if all the $\lambda_i(\boldsymbol{\Sigma}(\rho))$, $i \in H$, are the same. But, by the definition of $H$ plus the fact that the $f_i(\boldsymbol{\Sigma}(\rho))$, $i = 1, \ldots, n$, are mutually orthogonal, the condition that all the $\lambda_i(\boldsymbol{\Sigma}(\rho))$, $i \in H$, are the same is equivalent to the condition that $\text{col}^{\perp}(\boldsymbol{X}) \subseteq E_i(\boldsymbol{\Sigma}(\rho))$, for some $i = 1, \ldots, n$. This completes the proof of the part of the proposition relative to POI tests. Finally, observe that if the proposition holds for any POI test, i.e., for any $\bar{\rho} > 0$, then it must also hold for the LBI test (i.e., test (7) with $\boldsymbol{Q} = d\boldsymbol{\Sigma}(\rho)/d\rho|_{\rho=0}$), since the latter test is the limiting case of a POI test as $\bar{\rho} \to 0$. ■

**Proof of Lemma 1.** Clearly, two invertible matrices commute if and only if their inverses commute. Thus, Condition A requires that $\boldsymbol{\Sigma}^{-1}(\rho_1)$ and $\boldsymbol{\Sigma}^{-1}(\rho_2)$ commute for any $\rho_1, \rho_2 > 0$. For a SAR(1) model, $\boldsymbol{\Sigma}^{-1}(\rho) = \boldsymbol{I} - \rho(\boldsymbol{W}' + \boldsymbol{W}) + \rho^2\boldsymbol{W}'\boldsymbol{W}$, and hence Condition A is verified if and only if $\boldsymbol{W}\boldsymbol{W}' = \boldsymbol{W}'\boldsymbol{W}$, i.e., $\boldsymbol{W}$ is normal. To complete the proof, it remains to show that a nonsymmetric $\boldsymbol{W}$ that is obtained by row-standardization of a symmetric matrix cannot be normal. This follows from Lemma C.2, because if $\boldsymbol{W}$ were normal, then $\boldsymbol{f}_{\max}$ should be an eigenvector of $\boldsymbol{W}'$ associated to $\lambda_{\max}$. ■

**Proof of Proposition 4.** Assume that $\boldsymbol{W}$ is symmetric, and let, for notational convenience, $\tau_i(\rho) := 1 - \rho\lambda_i(\boldsymbol{W})$, $i = 1, \ldots, n$. When $0 \leq \rho < \lambda_{\max}$, we have $\lambda_i(\boldsymbol{\Sigma}(\rho)) = \tau_i^{-2}(\rho)$, $i = 1, \ldots, n$. Inserting such expressions in equation (E.2), we obtain that the power function of the POI test is nondecreasing in $\rho$ if and only if the statistic

$$t_{\bar{\rho}}(\rho) := \left\{ \sum_{i\in H} \frac{1}{\tau_i^2(\rho)} z_i^2 \right\}^{-1} \sum_{i\in H} \frac{\tau_i^2(\bar{\rho})}{\tau_i^2(\rho)} z_i^2 \tag{E.5}$$

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press

is nonincreasing in $\rho$ for any vector $z \in \mathbb{R}^n$. The condition $dt_{\bar{\rho}}(\rho)/d\rho \leq 0$ gives

$$\sum_{i\in H} \frac{1}{\tau_i^2(\rho)} z_i^2 \sum_{i\in H} \frac{2\lambda_i(\boldsymbol{W})\tau_i^2(\bar{\rho})}{\tau_i^3(\rho)} z_i^2 - \sum_{i\in H} \frac{\tau_i^2(\bar{\rho})}{\tau_i^2(\rho)} z_i^2 \sum_{i\in H} \frac{2\lambda_i(\boldsymbol{W})}{\tau_i^3(\rho)} z_i^2 \leq 0,$$

which can be rewritten as

$$2 \sum_{i,j\in H} a_{ij} z_i^2 z_j^2 \leq 0, \tag{E.6}$$

with

$$a_{ij} := \lambda_j(\boldsymbol{W}) \frac{\tau_j^2(\bar{\rho}) - \tau_i^2(\bar{\rho})}{\tau_i^2(\rho)\tau_j^3(\rho)}.$$

Note that for each $i, j \in H$ such that $i \neq j$,

$$a_{ij} + a_{ji} = \left[\frac{\lambda_j(\boldsymbol{W})}{\tau_j(\rho)} - \frac{\lambda_i(\boldsymbol{W})}{\tau_i(\rho)}\right] \frac{\tau_j^2(\bar{\rho}) - \tau_i^2(\bar{\rho})}{\tau_i^2(\rho)\tau_j^2(\rho)} = \left[\lambda_j(\boldsymbol{W}) - \lambda_i(\boldsymbol{W})\right] \frac{\tau_j^2(\bar{\rho}) - \tau_i^2(\bar{\rho})}{\tau_i^3(\rho)\tau_j^3(\rho)} \leq 0,$$

with equality if and only if $\lambda_i(\boldsymbol{W}) = \lambda_j(\boldsymbol{W})$, because $\tau_i(\rho)$ and $\tau_j(\rho)$ are positive, and $\tau_j^2(\bar{\rho}) - \tau_i^2(\bar{\rho}) = [1 - \bar{\rho}\lambda_j(\boldsymbol{W})]^2 - [1 - \bar{\rho}\lambda_i(\boldsymbol{W})]^2$ is positive, negative, or zero, if $\lambda_j(\boldsymbol{W}) - \lambda_i(\boldsymbol{W})$ is, respectively, negative, positive, or zero. Since, in addition, $a_{ii} = 0$, for each $i \in H$, it follows that (E.6) holds with equality if and only if all the $\lambda_i(\boldsymbol{\Sigma}(\rho))$, $i \in H$, are the same. This is the same condition that leads to equality in (E.4), so the proof can be completed in exactly the same way as the proof of Proposition 3. ■

**Proof of Proposition 5.** Assume that $\boldsymbol{W}$ is symmetric, so that the variance matrix of a SAR(1) model is $\boldsymbol{\Gamma}(\rho) := (\boldsymbol{I} - \rho\boldsymbol{W})^{-2}$ (having taken $\sigma^2 = 1$, without loss of generality). Since the $\boldsymbol{f}_i(\boldsymbol{W})$, $i = 1, \ldots, n$, are orthogonal, it follows that, under Condition B, if $\boldsymbol{f}_i(\boldsymbol{W}) \notin \mathrm{col}(\boldsymbol{X})$, then $\boldsymbol{f}_i(\boldsymbol{W}) \in \mathrm{col}^{\perp}(\boldsymbol{X})$, for any $i = 1, \ldots, n$. Thus, when $\boldsymbol{f}_i(\boldsymbol{W}) \notin \mathrm{col}(\boldsymbol{X})$, $i = 1, \ldots, n$, we have $\boldsymbol{C}\boldsymbol{\Gamma}(\rho)\boldsymbol{C}'\boldsymbol{C}\boldsymbol{f}_i(\boldsymbol{W}) = \boldsymbol{C}\boldsymbol{\Gamma}(\rho)\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{f}_i(\boldsymbol{W}) = \boldsymbol{C}\boldsymbol{\Gamma}(\rho)\boldsymbol{f}_i(\boldsymbol{W}) = \lambda_i(\boldsymbol{\Gamma}(\rho))\boldsymbol{C}\boldsymbol{f}_i(\boldsymbol{W})$. That is, $\boldsymbol{C}\boldsymbol{f}_i(\boldsymbol{W})$, $i \in H$, with $H$ as defined in the proof of Proposition 3, are $n - k$ orthogonal eigenvectors of $\boldsymbol{C}\boldsymbol{\Gamma}(\rho)\boldsymbol{C}'$. This implies, in particular, that $E_{n-k}(\boldsymbol{C}\boldsymbol{\Gamma}(\rho)\boldsymbol{C}')$ does not depend on $\rho$. Then, the proposition follows by Corollary 1, Lemma E.4, and Proposition 4. ■

https://doi.org/10.1017/S0266466609090641 Published online by Cambridge University Press