Rao’s score test in spatial econometrics

Luc Anselin

Regional Economics Applications Laboratory and Department of Agricultural and Consumer Economics, University of Illinois, Urbana, IL 61801, USA.

Abstract

Rao’s score test provides an extremely useful framework for developing diagnostics against hypotheses that reflect cross-sectional or spatial correlation in regression models, a major focus of attention in spatial econometrics. In this paper, a review and assessment is presented of the application of Rao’s score test against three broad classes of spatial alternatives: spatial autoregressive and moving average processes, spatial error components and direct representation models. A brief review is presented of the various forms and distinctive characteristics of RS tests against spatial processes. New tests are developed against the alternatives of spatial error components and direct representation models. It is shown that these alternatives do not conform to standard regularity conditions for maximum likelihood estimation. In the case of spatial error components, the RS test does have the standard asymptotic properties, whereas Wald and Likelihood Ratio tests do not. Direct representation models yield a situation where the nuisance parameter is only identified under the alternative, such that a Davies-type approximation to the significance level of the RS test is necessary. The performance of both new RS tests is illustrated in a small number of Monte Carlo simulation experiments. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

Spatial econometrics is a subfield of econometrics that deals with the complications caused by spatial interaction (spatial autocorrelation) and spatial structure (spatial heterogeneity) in regression models for cross-sectional and panel data (Paclinck and Klaassen, 1979; Anselin, 1988a). While prominent in the statistical literature, the problem of cross-sectional or spatial dependence has received much less attention in mainstream econometrics. Nevertheless, such models have seen a recent resurgence in empirical economic research, not only in regional and urban economics, where the importance of location is central, but also in local public finance, environmental and
resource economics, international trade and industrial organization, among others (a recent review is given in Anselin and Bera, 1998).

As in mainstream econometrics, the primary focus of attention in spatial econometrics was originally on estimation. This built on classic references in the statistical literature, such as the papers by Whittle (1954), Besag (1974), Ord (1975) and Mardia and Marshall (1984), in which estimators are derived for various spatial autoregressive processes. The maximum likelihood (ML) approach towards estimation has been dominant in this literature for the past 10–20 years (for overviews, see Ripley, 1981; Cliff and Ord, 1981; Upton and Fingleton, 1985; Anselin, 1988a; Haining, 1990; Cressie, 1993; Anselin and Bera, 1998). More recently, alternative estimators have been suggested as well, based on instrumental variables and general methods of moments (e.g., Anselin, 1990; Kelejian and Prucha, 1998, 1999; Conley, 1996).

When estimation is the main focus of interest, specification tests for the presence of spatial autocorrelation are constructed as significance tests on the spatial autoregressive coefficient, using the classic Wald (W) or Likelihood Ratio (LR) principles. A major drawback of this approach is that ML estimation of spatial regression models requires the optimization of a non-linear log-likelihood function that involves a Jacobian term of dimension equal to the size of the data set, which is a non-trivial computational problem (for recent reviews of the issues involved, see Anselin and Hudak (1992), Pace and Barry (1997) and Pace (1997)).

A more explicit attention to specification testing in spatial econometrics is fairly recent, although its statistical origins actually predate the treatment of estimation. Historically, the first test for spatial autocorrelation in a regression model was suggested by Moran (1950a, b). However, in contrast to approaches towards estimation of spatial processes, his test was not developed in an ML framework, nor with a specific spatial process as the alternative hypothesis. Instead, it is a simple test for correlation between nearest neighbors in space, generalizing an idea developed in Moran (1948). In spite of this, Moran’s $I$ test is still by far the best known and most used specification test for spatial autocorrelation in regression models, and it achieves several optimal properties (see Cliff and Ord, 1972; King, 1981). Its major advantage over the Wald and LR tests against specific forms of spatial autocorrelation is its computational simplicity, since only the residuals from an ordinary least-squares (OLS) regression are required and no non-linear optimization needs to be carried out. Of course, the construction of a test statistic from the results of estimation under the null hypothesis is one of the main practical features of Rao’s (1948) score test (RS) as well, and it turns out that Moran’s $I$ is equivalent to an RS test.

In this paper, a review and assessment is presented of the application of Rao’s score test in spatial econometrics. Specifically, the RS test approach will be considered in the context of three broad classes of ‘spatial’ alternative hypotheses to the classic regression model: spatial autoregressive (AR) and moving average (MA) processes; spatial error components; and direct representation models.

The first type of alternative is the most familiar, popularized by the work of Cliff and Ord (1981). It has received the almost exclusive attention of the literature to date,
arguably due to its similarity to dependent processes in time series. The most important result in this respect was given in a paper by Burridge (1980) where for normal error terms, the equivalence was established of Moran’s $I$ to the RS test against a spatial AR or MA process. Burridge used Silvey’s (1959) form of the test, based on the Lagrange multiplier of a constrained optimization problem, but the equivalence meant that the RS test shared the optimal properties of Moran’s $I$, and vice versa.\(^1\) An explicit score form of Burridge’s test was given in Anselin (1988b) and extended to multiple forms of spatial dependence and spatial heterogeneity. More recently, Rao’s score test has been generalized to a broad range of models and shown to form a solid basis for specification testing in spatial econometrics (see Anselin and Rey, 1991; Anselin and Florax, 1995; Anselin et al., 1996; Anselin and Bera, 1998). In Section 2 of the paper, a brief review is presented of the various forms and distinctive characteristics of RS tests against spatial AR or MA processes. The results are mostly familiar, although they are presented here as special cases of a RS test against a general spatial ARMA($p,q$) process.

In contrast to the review in Section 2, the materials in Sections 3 and 4 are new and deal with alternatives whose non-standard characteristics have not received attention in the literature to date. One is a spatial error component process, in which the regression error term is decomposed into a local and a spatial spillover effect. This specification was suggested by Kelejian and Robinson (1995) as an alternative to the Cliff–Ord model. In a maximum likelihood framework, a natural specification test for this form of ‘spatial autocorrelation’ would consist of an asymptotic significance test on the spatial spillover error component parameter.\(^2\) Upon closer examination however, it turns out that such a Wald test does not satisfy the usual assumption that, under the null hypothesis, the true parameter value for the spillover variance component should be an interior point of the parameter space. Instead, under the null, the parameter is on the boundary of the parameter space, as is the case in most standard (i.e., non-spatial) error component models (e.g., Godfrey, 1988, pp. 6–8). Therefore, in a maximum likelihood context, a W or LR test based on the unrestricted model (i.e., without explicitly accounting for the non-negativity constraint on the variance component) will follow a mixture of chi-squared distributions and not conform to the standard result. In contrast, the RS test is not affected by the fact that the parameter lies on the boundary when the null hypothesis is true, and hence remains a practical specification test. Such an RS test is derived in Section 3 and its performance illustrated in a small number of Monte Carlo simulation experiments.

\(^1\) In Anselin and Kelejian (1997), when a set of assumptions is satisfied pertaining to the heterogeneity of the process and the range of spatial interaction (structure of the spatial weights matrix), the asymptotic equivalence of Moran’s $I$ test to the RS test is also established in a general context, without requiring normality of the regression error terms.

\(^2\) Note that the estimator and specification tests suggested for this model by Kelejian and Robinson (1993, 1995) are not based on ML, but on a generalized method of moments approach (GMM), which does not require normality nor other restrictive assumptions needed for ML estimation.
Another alternative hypothesis of spatial autocorrelation in regression error terms is the so-called direct representation form. In this specification, the error covariance between two observations is a (‘direct’) function of the distance that separates them. This form is commonly used in the physical sciences, such as soil science and geology (see, e.g., Mardia and Marshall, 1984; Mardia, 1990; Cressie, 1993), but has also been implemented by Dubin (1988, 1992) and others in applied econometric work dealing with models of urban real estate markets. Specifically, Dubin (1988) suggests a LR test for spatial autocorrelation based on a model with a negative exponential distance decay function for the error covariance. This specification conforms to another familiar non-standard case of ML estimation, where the nuisance parameter is only identified under the alternative hypothesis. Hence, classical results on the W, LR and RS test are not applicable. So far, this feature has been ignored in the spatial econometric literature. In Section 4, the approximation procedure due to Davies (1977, 1987) is extended to an RS test against this alternative and its performance illustrated in a small number of Monte Carlo simulations.

Some concluding remarks are formulated in Section 5.

2. Rao score tests against spatial AR and MA processes

2.1. The null and alternative hypotheses

The point of departure is the classical linear regression model,

\[ y = X\beta + \varepsilon \]  \hspace{1cm} (2.1)

where \( y \) is an \( n \times 1 \) vector of observations on the dependent variable, \( X \) an \( n \times k \) fixed matrix of observations on the explanatory variables, \( \beta \) a \( k \times 1 \) vector of regression coefficients, and \( \varepsilon \) an \( n \times 1 \) vector of disturbance terms, with \( E[\varepsilon] = 0 \) and \( E[\varepsilon\varepsilon'] = \sigma^2 I \), and \( I \) is an \( n \times n \) identity matrix.

Typically, the alternative considered for this model embodies spatial dependence in the form of a (first order) spatial autoregressive or spatial moving average process, similar to the main focus of attention in the time series literature. A generalization of this to higher order processes is the spatial autoregressive-moving average model of order \( p, q \), outlined in Huang (1984), \(^3\) with

\[ y = \rho_1 W_1 y + \rho_2 W_2 y + \cdots + \rho_p W_p y + X\beta + \varepsilon \]  \hspace{1cm} (2.2)

as the mixed regressive (\( X \)), spatial autoregressive process (AR) in the dependent variable \( y \) (or, spatial lag dependence), and

\[ \varepsilon = \lambda_1 W_1 \xi + \lambda_2 W_2 \xi + \cdots + \lambda_q W_q \xi + \xi \]  \hspace{1cm} (2.3)

as the spatial moving average process (MA) in the error terms \( \xi \) (or, spatial error dependence). The parameters \( \rho_h, \ h = 1, \ldots, p \) define the AR process and \( \lambda_g, \ g = 1, \ldots, q \)

\(^3\) Huang’s original model did not contain the regressive part \( X\beta \), which has been inserted here to stay within the context of the linear regression model.
define the MA process. The matrices $W_h$ are $n \times n$ observable spatial weights matrices with positive elements, which represent the ‘degree of potential interaction’ between neighboring locations, in the sense that zero elements in the weights matrix exclude direct interaction. The full array of interaction is obtained as the product of the weights matrix with the spatial coefficient. Typically, the spatial weights matrices are scaled such that the sum of the row elements in each matrix is equal to one. After such row standardization, the weights matrix becomes asymmetric, with elements less than or equal to one. Also, by convention, the diagonal elements of the matrix are set to zero.

The elements of the weights matrix are usually derived from information on the spatial arrangement of the observations, such as contiguity, but more general approaches such as weights based on ‘economic’ distance are possible as well (see Cliff and Ord (1981), Anselin (1980, 1988a), Case et al. (1993), and Anselin and Bera (1998), for discussions of the properties and importance of the spatial weights matrix). When a higher order process is specified, as in (2.2)–(2.3), the matrices $W_h$ typically (but not necessarily) correspond to different orders of contiguity, where care has to be taken to avoid circularity and redundancy in the definition of contiguity (see Anselin and Smirnov, 1996, for an extensive discussion). More precisely, in order to facilitate the interpretation and identifiability of the spatial parameters in (2.2)–(2.3), it is typically assumed that the various weights do not have elements in common, or, for each row $w_{i*}$ of the weights matrix, and for any two orders of contiguity $h; l$, that

$$(w^h_{i*})(w^l_{i*})' = 0.$$  

(2.4)

When duplicate non-zero elements are present in weights matrices associated with different orders in a higher order process, the interpretation of the coefficients $\rho_h$ or $\lambda_g$ as measuring the ‘pure’ contribution of the $h$th ($g$th) order is not straightforward. Blommestein (1985) shows how ignoring these redundancies affects ML estimation. The parallel to time series analysis is obvious: observations that are ‘shifted’ in time by different units cannot coincide. While in a strict sense, a proper 2SLS estimator may yield a consistent estimator, such an estimator must incorporate the constraints on the parameters that are implied by the redundancy in the weights. For example, consider a second order process,

$$y = \rho_1 W_1 y + \rho_2 W_2 y + X\beta + \varepsilon$$  

(2.5)

with $W_1 = W_{11} + W_0$ and $W_2 = W_{22} + W_0$, where $W_0$ is a weights matrix with the common elements between $W_1$ and $W_2$. Taking into account the overlap between the weights, the model can also be expressed as

$$y = \rho_1 W_{11} y + \rho_2 W_{22} y + (\rho_1 + \rho_2)W_0 y + X\beta + \varepsilon,$$  

(2.6)

which illustrates the potential problem of identification.

It is the inclusion of the spatial weights that render the spatial models to depart from the standard linear model, thereby limiting the applicability of standard estimation procedures based on ordinary least squares (OLS). Specifically, OLS in the presence of spatially lagged dependent variables ($Wy$) is a biased and inconsistent estimator,
whereas the presence of spatially dependent error terms \((W\zeta)\) renders it inefficient and yields a biased estimator for the error variance covariance matrix (an extensive treatment is given in Anselin (1988a)). In addition to these inferential differences between the two specifications, there are important distinctions between the interpretation of ‘substantive’ spatial dependence (lag) and spatial dependence as a ‘nuisance’ (error) (see Anselin, 1988a; Manski, 1993).

While the general SARMA \((p,q)\) model as such has not seen much application, several special cases have been considered extensively in the literature, such as,

(a) mixed regressive, (first order) spatial autoregressive model (Ord, 1975):
\[
y = \rho Wy + X\beta + \epsilon, \tag{2.7}
\]
(b) biparametric spatial autoregressive model (Brandsma and Ketellapper, 1979):
\[
y = \rho_1 W_1 y + \rho_2 W_2 y + X\beta + \epsilon, \tag{2.8}
\]
(c) higher order spatial autoregressive model (Blommestein, 1983, 1985):
\[
y = \rho_1 W_1 y + \rho_2 W_2 y + \cdots + \rho_p W_p y + X\beta + \epsilon, \tag{2.9}
\]
(d) (first order) spatial moving average error process (Haining, 1988):
\[
\epsilon = \lambda W\zeta + \zeta. \tag{2.10}
\]

In addition, a (first order) spatial autoregressive error process is commonly considered as well (Ord, 1975), or,
\[
\epsilon = \phi W\epsilon + \zeta. \tag{2.11}
\]

Both (2.10) and (2.11) result in a non-spherical error covariance matrix, although with very different degrees of non-diagonality.\(^4\) For the spatial MA process, the error variance is
\[
E[\epsilon\epsilon'] = \sigma^2[(I + \lambda W)(I + \lambda W)'] = \sigma^2[I + \lambda(W + W') + \lambda^2 WW'] \tag{2.12}
\]
with \(\sigma^2\) as the variance for the error term \(\zeta\). This yields non-zero elements for the first order \((W, W')\) and second order \((WW')\) neighbors only. In contrast, for the spatial AR process, the error variance is
\[
E[\epsilon\epsilon'] = \sigma^2[(I - \phi W)(I - \phi W)]^{-1}, \tag{2.13}
\]
which, for a row-standardized weights matrix and for \(|\phi| < 1\), has declining values for the off-diagonal elements with increasing orders of contiguity and thus contains many more non-zero elements than (2.10).\(^5\)

\(^4\) Note that a spatial AR error process is sometimes considered in combination with a spatial AR process in the dependent variable (e.g., Case, 1991). As shown in Anselin (1988a), Anselin and Bera (1998), and Kelejian and Prucha (1998), unless care is taken in the specification of the spatial weights and the exogenous variables \((X)\), some parameters in this model may be unidentified.

\(^5\) Under these conditions, \((I - \phi W)^{-1} = I + \phi W + \phi^2 W^2 + \phi^3 W^3 + \cdots\), where each power of the weights matrix corresponds to a higher order of contiguity. For an illustration and more extensive discussion, see Haining (1988) and Anselin and Bera (1998).
Rao score tests against the various alternative hypotheses can be computed from the results of OLS estimation of the model under the null, i.e., the classic regression specification (2.1).

2.2. Score test against a SARMA\((p, q)\) process

An RS test with a SARMA\((p, q)\) process as the alternative hypothesis is based on the \(k + 1 + p + q\) by 1 parameter vector \(\theta\) arranged as \(\theta = [\beta', \sigma^2, \rho_1, \ldots, \rho_p, \lambda_1, \ldots, \lambda_q]'\). The null hypothesis is formulated as \(H_0: \rho_h = 0, \forall h\) and \(\lambda_g = 0, \forall g\). This allows tests against the various specific alternatives, such as first order AR or MA processes, to be found as special cases. For ease of notation, set

\[
A = I - \rho_1 W_1 - \rho_2 W_2 - \cdots - \rho_p W_p,  \tag{2.14}
\]

\[
B = I + \lambda_1 W_1 + \lambda_2 W_2 + \cdots + \lambda_q W_q \tag{2.15}
\]

and thus \(|A|\) and \(|B|\) are Jacobian terms in a transformation corresponding to the AR and MA process, respectively. For these Jacobian terms to exist, the matrices \(A\) and \(B\) must be non-singular, which yields constraints on the admissible parameter space.

Assuming normality, the log-likelihood can be expressed as

\[
L = -\frac{n}{2} \ln(2\pi I) - \frac{n}{2} \ln(\sigma^2) - \ln|B| + \ln|A| \tag{2.16}
- \frac{1}{2\sigma^2} (Ay - X\beta)'(BB')^{-1}(Ay - X\beta).
\]

A Rao score test is obtained in the usual fashion as \(RS = d'(\theta_0)J(\theta_0)^{-1}d(\theta_0)\), with the score and information matrix evaluated under the null. The first two elements of

---

The derivation which follows is a generalization of Anselin (1988a, Chapter 6), but using an MA process for the error terms. In a strict sense, this model does not satisfy the standard regularity conditions underlying maximum likelihood estimation (e.g., as spelled out in Rao (1973, Chapter 5)), nor those for ML estimation for dependent and heterogeneous processes considered in the time series literature (e.g., White, 1984; Pötscher and Prucha, 1997). The main problem is that the required central limit theorem(s) must take into account triangular arrays, due to the fact that the structure of the weights matrix \(W\) depends on the sample size \(n\) [for a recent treatment incorporating a central limit theorem for triangular arrays, see Kelejian and Prucha (1998, 1999)]. The ‘standard’ regularity conditions therefore do not apply directly to spatial stochastic process models of the general SARMA type (and associated special cases). To date, a comprehensive, formal treatment of this issue remains to be completed. However, ‘intuitively’ most weights matrices used in practice (e.g., those based on contiguity) satisfy ‘sufficient conditions’ that imply constraints on the extent of the spatial dependence (covariance) and the degree of heterogeneity (higher order moments), which is the ‘spirit’ in which the formal properties are obtained in the time series literature. In addition, considerable simulation experiments also suggest that the maximization of log-likelihood (2.16) indeed yields estimators whose properties match those suggested by the maximum likelihood results. Hence, taking into account the caveats expressed above, one may reasonably proceed as if estimation and specification testing follow the asymptotic properties established for ML estimation and testing in other types of dependent stochastic processes, although in a strict sense this has not been formally established.
the score vector \( d = \frac{\partial L}{\partial \theta} \) follow, with \( \varepsilon = Ay - X\beta \), as

\[
\frac{\partial L}{\partial \beta} = \frac{1}{\sigma^2} X'(BB')^{-1}\varepsilon
\]

(2.17)

\[
\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\varepsilon'(BB')^{-1}\varepsilon.
\]

(2.18)

The elements of the score vector corresponding to the AR parameters \( \rho_h \) are, for each \( h = 1, \ldots, p \),

\[
\frac{\partial L}{\partial \rho_h} = -\text{tr} A^{-1}W_h + \frac{1}{\sigma^2}\varepsilon'(BB')^{-1}W_hy
\]

(2.19)

with \( \text{tr} \) as the matrix trace operator. The elements of the score vector for the MA parameters \( \lambda_g \) are, for each \( g = 1, \ldots, q \),

\[
\frac{\partial L}{\partial \lambda_g} = -\text{tr} B^{-1}W_g + \frac{1}{\sigma^2}\varepsilon'(BB')^{-1}W_gb^{-1}\varepsilon.
\]

(2.20)

Under the null hypothesis, \( A = B = I \), which, with Eqs. (2.17) and (2.18) set to zero, yields the OLS estimator for \( \beta \) and \( \hat{\sigma}^2 = \varepsilon'e/n \), with \( e \) as the vector of OLS residuals. The score elements for the spatial autoregressive parameters become, for each \( h = 1, \ldots, p \):

\[
\frac{\partial L}{\partial \rho_h} \bigg|_{H_0} = -\text{tr} W_h + \frac{1}{\sigma^2}\varepsilon'W_hy,
\]

(2.21)

or, since \( \text{tr} W_h = 0 \) by convention,

\[
\frac{\partial L}{\partial \rho_h} \bigg|_{H_0} = \frac{1}{\sigma^2}\varepsilon'W_hy.
\]

(2.22)

Similarly, under the null, the score elements for the spatial moving average parameters become, for each \( g = 1, \ldots, q \):

\[
\frac{\partial L}{\partial \lambda_g} \bigg|_{H_0} = \frac{1}{\sigma^2}\varepsilon'W_ge.
\]

(2.23)

A more involved issue is the derivation of the information matrix \( J_{\theta\theta} = -E[\partial^2 L / \partial \theta \partial \theta'] \). As shown in Anselin (1988a) for a first-order SAR model with a spatial autoregressive error, this information matrix is not block diagonal between the parameters of the model \( (\beta, \rho) \) and those of the error covariance \( (\sigma^2, \lambda) \). This result also holds for the more general model considered here. For ease of notation, consider the information matrix partitioned into four blocks: \( J_{11} \), for the diagonal block of \( k + 1 \) by \( k + 1 \) elements corresponding to \( \beta \) and \( \sigma^2 \); \( J_{22} \) for the diagonal block of \( p + q \) by \( p + q \) elements corresponding to the parameters \( \rho_h \) and \( \lambda_g \); and \( J_{12} \) (and its transpose \( J_{21} \)) for the off-diagonal block. The first block is a familiar result:

\[
J_{11} = \begin{bmatrix}
\frac{1}{\sigma^2} X'(BB')^{-1}X & 0 \\
0 & \frac{n}{2\sigma^4}
\end{bmatrix}.
\]

(2.24)
The elements of $J_{22}$ follow as\footnote{The results are obtained using the same approach as in Anselin (1988a, Chapter 6). The expected values of the elements of the Hessian are obtained by using the following intermediate results: $y = A^{-1}X\hat{\beta}$, $E[\hat{\beta}] = 0$, $E[y] = A^{-1}X\hat{\beta}$, $E[y'y'] = (A^{-1}X\hat{\beta})(A^{-1}X\hat{\beta})' + \sigma^2 A^{-1}(BB')(A^{-1})'$.}

$$J_{\rho_{hj}} = \text{tr} A^{-1}W_jA^{-1}W_h + \frac{1}{\sigma^2} \text{tr} W_j'(BB')^{-1}W_h(A^{-1}X\hat{\beta})(A^{-1}X\hat{\beta})'$$

$$+ \text{tr} W_j'(BB')^{-1}W_hA^{-1}(BB')(A^{-1})'$$

(2.25)

for $h = 1, \ldots, p$ and $j = 1, \ldots, p$;

$$J_{\lambda_i\lambda_i} = \text{tr}[W_i(B' + BW_i)'(BB')^{-1}W_iB^{-1}$$

(2.26)

for $g = 1, \ldots, q$ and $l = 1, \ldots, q$; and

$$J_{\rho_{h\lambda}} = \text{tr}[W_h(B' + BW_h)](BB')^{-1}W_hA^{-1}$$

(2.27)

for $h = 1, \ldots, p$ and $g = 1, \ldots, q$.

Finally, the non-zero elements of $J_{12}$ are

$$J_{\beta_{\rho_h}} = \frac{1}{\sigma^2} X'(BB')^{-1}W_h(A^{-1}X\hat{\beta})$$

(2.28)

for $h = 1, \ldots, p$, and

$$J_{\sigma^2_{\lambda_g}} = \frac{1}{2\sigma^2} \text{tr}[W_gB' + BW_g'](BB')^{-1}$$

(2.29)

for $g = 1, \ldots, q$.

Under the null hypothesis, these expressions become

$$J_{11|H_0} = \left[ \begin{array}{cc}
\frac{1}{\sigma^2} X'X & 0 \\
0 & n \frac{1}{2\sigma^2}
\end{array} \right],$$

(2.30)

$$J_{\rho_{hj}|H_0} = \text{tr} W_jW_h + \text{tr} W_j'W_h + \frac{1}{\sigma^2} (W_jX\hat{\beta})'(W_hX\hat{\beta}),$$

(2.31)

$$J_{\lambda_i\lambda_i|H_0} = \text{tr} W_iW_g + \text{tr} W_i'W_g,$$

(2.32)

$$J_{\rho_{h\lambda}|H_0} = \text{tr}[W_h + W_g']W_h,$$

(2.33)

$$J_{\beta_{\rho_h}|H_0} = \frac{1}{\sigma^2} X'W_hX\hat{\beta}$$

(2.34)

$$J_{\sigma^2_{\lambda_g}|H_0} = \frac{1}{2\sigma^2} \text{tr}[W_g + W_g'] = 0,$$

(2.35)

with, as before, $h, j = 1, \ldots, p$, and $g, l = 1, \ldots, q$. 

With the same notational convention as in Anselin (1988a), consider the trace terms \( T_{hj} = \text{tr} W_h W_j + \text{tr} W_h' W_j \). Because of assumption (2.4), \( T_{hj} = 0 \) for \( h \neq j \) in (2.31), and, similarly, \( T_{gl} = 0 \) for \( g \neq l \) in (2.32) and \( T_{hg} = 0 \) for \( h \neq g \) in (2.33).\(^8\) Given these results, the submatrix \( J_{22} \) consists of four blocks: a \( p \times p \) matrix for the AR parameters, with as diagonal elements \( T_{hh} + (1/\hat{\sigma}^2)(W_h X \hat{\beta})' (W_h X \hat{\beta}) \), and as off-diagonal elements \( (1/\hat{\sigma}^2)(W_h X \hat{\beta})' (W_l X \hat{\beta}) \); a \( q \times q \) diagonal matrix for the MA parameters, consisting of the trace terms \( T_{gg} \), and \( p \times q \) \((q \times p)\) off-diagonal blocks with the trace term \( T_{hh} (T_{gg}) \) on the diagonal.\(^9\)

Unlike in many instances in mainstream econometrics, the needed inverse \( J(\theta_0)^{-1} \) of the part of the information matrix corresponding to the \( \rho_h, \lambda_q \) parameters has no closed form solution. However, following the same approach as in Anselin (1988a), it can be expressed as the inverse of a \((p + q) \times (p + q)\) matrix, using partitioned inversion with \( J_{22}^{-1} = (J_{22} - J_{21} J_{11}^{-1} J_{12})^{-1} \). Due to the form of \( J_{11} \) in (2.30) and \( J_{21} \) in (2.34)–(2.35), this simplifies to the inverse of a \((p + q) \times (p + q)\) matrix with the following structure:

\[
J_{22}^{-1} = \begin{bmatrix}
T_{11} + \frac{1}{\hat{\sigma}^2}(W_1 X \hat{\beta})' M(W_1 X \hat{\beta}) & \cdots & \frac{1}{\hat{\sigma}^2}(W_1 X \hat{\beta})' M(W_p X \hat{\beta}) & T_{11} & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{\hat{\sigma}^2}(W_p X \hat{\beta})' M(W_1 X \hat{\beta}) & \cdots & T_{pp} + \frac{1}{\hat{\sigma}^2}(W_p X \hat{\beta})' M(W_p X \hat{\beta}) & 0 & T_{pq} \\
0 & \cdots & T_{11} & 0 & T_{pp} \\
0 & \cdots & 0 & T_{pp} & T_{qq}
\end{bmatrix}^{-1},
\]

(2.36)

where \( M = I - X(X'X)^{-1}X' \) and thus \( M(W_h X \hat{\beta}) \) are the residuals of a regression of the spatially lagged predicted values \( W_h X \hat{\beta} \) on the original regressors \( X \).

On the basis of these results, a RS statistic against an alternative in the form of a spatial ARMA\((p, q)\) process can be constructed from the residuals of an OLS regression, using (2.22) and (2.23) as the elements of the score vector and (2.36) as the corresponding information matrix inverse. This statistic is asymptotically distributed as \( \chi^2(p + q) \). This result generalizes the derivation for a \((1, 1)\) process in Anselin (1988a).\(^{10}\)

---

\(^8\)This assumes that the same spatial weights are used in (2.2) and (2.3). While not absolutely required, this is the situation typically encountered in practice. If different weights structures are used, \( T_{gh} \) may be non-zero for some \( g, h \).

\(^9\)Typically, \( p = q \), but this is not required in general.

\(^{10}\)The statement on the asymptotic distribution is based on the assumption that the ‘standard’ results for RS tests are applicable in this context as well. See also footnote 6.
2.3. Properties and special cases

Using the general framework outlined in the previous section, a number of special cases can be derived and their distinguishing characteristics noted:

(a) The RS test against spatial processes does not simplify to a NR² form based on an auxiliary regression. This contrasts with the many results in the time series domain (see also Anselin and Bera, 1998, for an extensive discussion).

(b) A RS statistic against a spatial AR(p) process can be based on (2.22) and the upper $p \times p$ block of (2.36). Due to the presence of the off-diagonal elements in the information matrix, this statistic does not equal the sum of $p$ separate statistics for a test against a single order process with weights matrix $W_h$. The RS test against a single order spatial AR process (in the spatial weights $W$) follows as

$$RS_p = \left( \frac{1}{\sigma^2} \hat{e}' W \hat{e} \right)^2 / \left( T + \frac{1}{\sigma^2} (WX \hat{\beta})' M(WX \hat{\beta}) \right),$$

with $T = \text{tr} WW + \text{tr} W'W$ and with an asymptotic $\chi^2(1)$ distribution (Anselin, 1988b).

(c) A RS statistic against a spatial MA(q) error process can be based on (2.23) and the lower $q \times q$ block of (2.36). Since the latter is diagonal, this statistic equals the sum of $q$ separate statistics, similar to the familiar Box and Pierce (1970) statistic in time series. The RS test against a single order spatial error MA process follows as

$$RS_{\lambda} = \left( \frac{1}{\sigma^2} \hat{e}' W_g \hat{e} \right)^2 / T$$

with $T$ as above and with an asymptotic $\chi^2(1)$ distribution (Burridge, 1980; Anselin, 1988a).

(d) The RS test against a spatial AR error process is identical to the test against a spatial MA error process. These alternatives are thus locally equivalent and the RS test cannot be used to distinguish between the two (see also Anselin and Bera, 1998). However, the lag alternative is not equivalent to the error dependence and therefore the $RS_p$ and $RS_{\lambda}$ can be used to distinguish between these two important cases (see Anselin et al., 1996).

(e) Due to the presence of the off-diagonal block in (2.34), the RS test against a spatial ARMA(p, p) process does not decompose into the sum of the AR and MA components (Anselin, 1988b).³

(f) Score tests against spatial AR or MA processes are asymptotic tests. Several simulation experiments have demonstrated their performance in finite samples. For alternatives with error dependence, Moran’s $I$ test is slightly superior to $RS_{\lambda}$ in

³ Interestingly, this test does decompose into the sum of a variant that is robust to local specification of one form and the RS test against the other form. For example, $RS_{\lambda} = RS_{\lambda}^* + RS_p = RS_p^* + RS_{\lambda}$, where $RS_{\lambda}^*$ and $RS_p^*$ are the robust forms derived in Anselin et al. (1996) (see also Anselin and Bera, 1998).
small samples, but asymptotically the two are equivalent. For alternatives with spatial lag dependence, Moran’s $I$ has considerable power as well, but $RS_p$ is superior in this case (extensive experiments are given in Anselin and Rey (1991) and Anselin and Florax (1995)).

3. A Rao score test against spatial error components

3.1. General principles

The spatial error components specification was suggested by Kelejian and Robinson (1995) as an alternative to a spatial AR process that avoided some of its singularity problems. Formally, the model is

$$y = X\beta + \varepsilon, \quad (3.1)$$

$$\varepsilon = W\psi + \zeta, \quad (3.2)$$

where $\psi$ and $\zeta$ are i.i.d. error terms, $W$ is a $n \times n$ spatial weights matrix (which does not have to be row-standardized) and the other notation is as in Section 2. The random error components $W\psi$ and $\zeta$ correspond, respectively, to a spatial spillover (a weighted average of neighboring errors as determined by the non-zero elements of the rows in $W$) and a local effect, which are assumed to be uncorrelated (Kelejian and Robinson, 1995, p. 88). The weights matrix does not have to correspond to a notion of contiguity, but could express any meaningful ‘group’ effect. For example, this specification is useful when two types of effects are assumed to drive the error components, one associated with a larger group ($W\psi$), such as a region or network of interacting agents, the other considered to be an innovation (i.e., location-specific). This specification differs from the more standard random effects model in that the ‘group’ expressed in each row of $W$ can be different for each observation, providing a more flexible approach. A crucial assumption underlying this model is that the spillover effects are uncorrelated with the innovations.\(^{12}\)

The assumptions underlying this model are:

$$E[\psi] = 0, \quad E[\zeta] = 0,$$

$$E[\psi \psi'] = \sigma_\psi^2 I, \quad E[\zeta \zeta'] = \sigma_\zeta^2 I$$

and

$$E[\psi_i \zeta_j] = 0, \quad \forall i, j.$$

The covariance matrix for the regression error term $\varepsilon$ in (3.1) thus takes the form

$$E[\varepsilon \varepsilon'] = \sigma_\zeta^2 I + \sigma_\psi^2 WW' \quad (3.3)$$

\(^{12}\) See Kelejian and Robinson (1993, 1997) for recent empirical applications using this specification.
with $\sigma_\xi^2 > 0$, $\sigma_\psi^2 \geq 0$, or, alternatively

$$E[\varepsilon \varepsilon'] = \sigma^2 \Omega(\gamma) = \sigma^2(I + \gamma WW'),$$

(3.4)

where $\theta = [\sigma_\xi^2, \gamma]'$ is the parameter vector, with $\sigma^2 = \sigma_\xi^2 > 0$ and $\gamma = \sigma_\psi^2/\sigma_\xi^2 \geq 0$, the ratio of the spatial spillover variance relative to the local variance.\(^{13}\) Since $WW'$ is symmetric and non-negative definite and $I$ is positive definite, the matrix $\Omega$ is always positive definite. Kelejian and Robinson derive an estimator for model (3.1)–(3.3) based on the General Method of Moments (GMM), in which a consistent estimator for the variance components is obtained from an auxiliary regression using OLS.\(^{14}\) A specification test for spatial autocorrelation is formulated as an asymptotic significance test on the parameter $\sigma_\psi^2$ in (3.3) \citep{Kelejian and Robinson, 1993, p. 304}.

Estimation of the parameters in this model can also be based on maximum likelihood, with the additional assumption of normal distributions for $\psi$ and $\xi$. The model is a special case of a non-spherical error variance–covariance matrix, but unlike the standard case \citep[e.g.,][]{Magnus, 1978, pp. 283–284; Serfling, 1980, pp. 143–149; Godfrey, 1988, pp. 6–8}, special care must be taken to establish sufficient conditions that impose distance decay and bounds on the heterogeneity.\(^{15}\)

In the standard maximum likelihood context, one of the basic regularity conditions requires that the true parameter values for $\theta$ are interior points of a finite dimensional, closed and bounded parameter space $\Theta$. For a null hypothesis of the form $H_0: \gamma = 0$ in model (3.4), the true parameter value is on the boundary of the parameter space and hence does not satisfy this condition. As a consequence, and assuming that the same conditions are required in the spatial model, hypothesis tests for $\gamma$ based on a standard implementation of LR or W tests will not have the expected asymptotic $\chi^2(1)$ distribution under the null hypothesis. This same problem is encountered in a number of econometric specifications involving error components and has received considerable attention in the literature \citep[see also Godfrey, 1988, pp. 92–98; Bera et al., 1998 for recent reviews]{Chernoff, 1954; Moran, 1971; Chant, 1974}. A major result is that in contrast to the LR and W tests, the score test is unaffected by this condition.

### 3.2. Score test

The model with error covariance matrix (3.4) is a special case of a non-spherical error in a linear regression. Its log-likelihood, score and information matrix can be derived in a straightforward manner using the general principles outlined in Magnus \citeyear{Magnus, 1978} and Breusch \citeyear{Breusch, 1980}, among others. In contrast to the spatial ARMA model

\(^{13}\) Note the similarity between (3.3) and the error covariance for the model with a spatial MA error process, (2.10), for which $E[\varepsilon \varepsilon'] = \sigma^2[I + \lambda(W + W') + \lambda^2 WW'].$ Both specifications result in a small number of non-zero elements in the covariance matrix (corresponding to the first and/or second order neighbors).

\(^{14}\) No non-negativity constraint is imposed in this auxiliary regression.

\(^{15}\) See also the discussion in footnote 6. Note that Kelejian and Robinson \citeyear{Kelejian and Robinson, 1993} establish consistency of the GMM estimator for $\beta$ explicitly without resorting to normality.
considered in the previous section, for this case the information matrix is block diagonal between the elements corresponding to $\beta$ and those for the error parameter vector $\theta = [\sigma^2, \gamma]'$. Therefore, to obtain the information matrix for the RS statistic, only the submatrix corresponding to $\theta$ needs to be considered. The following expressions follow for the log-likelihood $L$, the score vector $d = \partial L / \partial \theta$ and the relevant part of the information matrix, $J_\theta = -E[\partial^2 L / \partial \theta \partial \theta']$, using $\varepsilon = y - X\beta$, and $tr$ as the matrix trace operator,

$$L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2} \ln|\Omega| - \frac{1}{2} \frac{\varepsilon' \Omega^{-1} \varepsilon}{\sigma^2}, \quad (3.5)$$

$$d = \left[ \begin{array}{c} -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} \varepsilon' \Omega^{-1} \varepsilon \\ -\frac{1}{2} \tr \Omega^{-1} WW' + \frac{1}{2\sigma^2} \varepsilon' \Omega^{-1} WW' \Omega^{-1} \varepsilon \end{array} \right], \quad (3.6)$$

and

$$J_\theta = \left[ \begin{array}{c} \frac{n}{2\sigma^2} \tr \Omega^{-1} WW' \\ \frac{1}{2\sigma^2} \tr \Omega^{-1} WW' \frac{1}{2} \tr WW' \Omega^{-1} WW' \end{array} \right]. \quad (3.7)$$

A Rao score test for the null hypothesis $H_0: \gamma = 0$ against the alternative $H_1: \gamma > 0$ is obtained as $RS_\gamma = d'(\theta_0) J(\theta_0)^{-1} d(\theta_0)$, with the score and information matrix evaluated under the null, or, with $\Omega^{-1} = I$, where $I$ is an $n \times n$ identity matrix. The corresponding expressions are

$$d(\theta_0) = -\frac{1}{2} \tr WW' + \frac{1}{2\sigma^2} \varepsilon' WW'e, \quad (3.8)$$

$$J(\theta_0) = \left[ \begin{array}{c} \frac{n}{2\sigma^2} \tr WW' \\ \frac{1}{2\sigma^2} \tr WW' \frac{1}{2} \tr WW' \Omega^{-1} WW' \end{array} \right], \quad (3.9)$$

where $e$ is a vector of OLS residuals and $\sigma^2 = e'e/n$. To simplify notation, set $\tr WW' = T_1$ and $\tr WW' WW' = T_2$. Using partitioned inversion, the inverse of the 2, 2 element of (3.5) then follows as

$$J''(\theta_0) = 2 \left[ T_2 - \frac{(T_1)^2}{n} \right]^{-1}. \quad (3.10)$$

Combining all terms, the test statistic\(^{16}\) is obtained as

$$RS_\gamma = \left[ \frac{e' WW'e}{\sigma^2} - T_1 \right]^2 / \left\{ 2 \left[ T_2 - \frac{(T_1)^2}{n} \right] \right\}. \quad (3.11)$$

As in the cases considered in Section 2, this test requires only the results from OLS estimation, and, under $H_0$, $RS_\gamma \overset{D}{\rightarrow} \chi^2(1)^{17}$. In practice, the test can also be implemented by comparing the positive square root of $RS_\gamma$, or $z_\gamma = \sqrt{RS_\gamma}$ to the one-sided significance levels of a standard normal variate.

\(^{16}\) An interesting interpretation of this statistic, suggested by Anil Bera, is to consider it as a moment test, where the interest focuses on the moment condition $E[e' WW'e/e'e] = T_1$. Under the null hypothesis, with $E[e'e] = \sigma^2 I$, this condition is satisfied.

\(^{17}\) Again, this assumes that the standard asymptotic results for RS tests are applicable. See also footnote 6.
3.3. Small sample performance

While an extensive investigation is beyond the scope of the current paper, some limited insight into the performance of the test is obtained from a small set of Monte Carlo simulation experiments.\(^\text{18}\) The data are generated over a regular square lattice, using a standard normal distribution for the error term \(\varepsilon\) and a fixed \(n \times 2\) matrix \(X\), with the first column consisting of ones and the second column generated as a uniformly distributed random variate between 0 and 10.\(^\text{19}\) Under the alternative hypothesis, the proper covariance structure is generated for the error term by applying a Choleski decomposition to the matrix \(\Omega = I + \gamma WW'\) using a first order rook contiguity for \(W\) and \(\gamma = 0.00, 0.01, 0.05, 0.10, 0.25, 0.50, 0.75, 1.0, 2.0, 3.0, 4.0\). Five sample sizes are considered, \(n = 25 (5 \times 5)\), 49 (7 \times 7), 81 (9 \times 9), 121 (11 \times 11) and 400 (20 \times 20).

The results are reported in Table 1, with the empirical Type I errors based on the one-sided \(z_\gamma\) test. All results are for 10,000 replications, using nominal significance levels of 0.05 and 0.01. The size of the test is acceptable (given in the first row of the table), yielding a slight under-rejection relative to the nominal level for 0.05, but rejection levels remain within the two standard deviation range for 0.01 for all but the smallest data set.\(^\text{20}\) This is in line with the properties found for RS tests against spatial AR and MA error processes where under the null the test under-rejects for all but the largest sample sizes (for extensive evidence, see Anselin and Rey, 1991; Anselin and Florax, 1995). The power of the test illustrates its asymptotic nature, since the null hypothesis is only rejected in the majority of cases for values of \(\gamma\) larger than 2.0 (for \(n = 49\)), 1.0 (for \(n = 81\)), 2.0 (for \(n = 121\)) and 0.5 (for \(n = 400\)). This is never accomplished in the smallest data set, even for \(\gamma = 4.0\).

---

\(^{18}\) A more comprehensive analysis and comparison to other tests is treated in Anselin and Moreno (2000).

\(^{19}\) Under the null hypothesis, this yields an average \(R^2\) of around 0.94 for the OLS regressions.

\(^{20}\) The 10,000 replications yield two standard deviation intervals around the respective nominal rejection frequencies of 0.0466–0.0544 and 0.0082–0.0118.
The relative low power of this test in the smaller samples is not totally surprising, since under alternative (3.4) the number of non-zero elements for the typical row (i.e., for interior locations) in the covariance matrix (determined by $WW'$) is limited to only eight for the rook criterion of contiguity (compared to 12 for a spatial MA process and many more for a spatial AR process). In other words, model (3.3) specifies the weakest form of spatial correlation among the alternatives considered in this paper. As a result, moderately large data samples or high values for the parameter $\gamma$ seem to be needed in order for deviations from the null hypothesis to be detected with some regularity by the RS test. On the other hand, lower values of the $\gamma$ parameter will have minor consequences for the properties of the OLS estimator, so that in practice the new score test may be a useful diagnostic in data sets that exceed a 100 or so observations. However, before endorsing this test statistic without reservations, a number of issues need to be further investigated, such as the extent to which its properties are maintained when the error term is non-normal, and the comparison of its power to that of alternatives such as LR and W tests that take into account the censored nature of the distribution of the estimators, as well as procedures based on GMM estimation.

4. A Rao score test against direct representation of spatial covariance

4.1. General principles

Under the direct representation form of spatial autocorrelation, the elements of the error covariance matrix are specified as an inverse function of distance. Formally,

$$\text{Cov}[\varepsilon_i \varepsilon_j] = \sigma^2 f(d_{ij}, \varphi), \quad (4.1)$$

where $\varepsilon_i$ and $\varepsilon_j$ are regression disturbance terms, $\sigma^2$ is the error variance, $d_{ij}$ is the distance separating observations (locations) $i$ and $j$, and $f$ is a distance decay function such that $\hat{f}/\hat{d} < 0$ and $|f(d_{ij}, \varphi)| \leq 1$, with $\varphi \in \Phi$ as a $p \times 1$ vector of parameters on an open subset $\Phi$ of $\mathbb{R}^p$. This form is closely related to the variogram model used in geostatistics, although with stricter assumptions regarding stationarity and isotropy.\(^{21}\)

Using (4.1) for individual elements, the full error covariance matrix follows as

$$E[\varepsilon^\prime \varepsilon] = \sigma^2 \Omega(d_{ij}, \varphi), \quad (4.2)$$

\(^{21}\)The specification of spatial covariance functions is not arbitrary, and a number of conditions must be satisfied in order for the model to be ‘valid’ (details are given in Cressie, 1993, pp. 61–63, 67–68, 84–86). For a stationary process, the covariance function must be positive definite and it is required that $f \to 0$ as $d_{ij} \to \infty$, which is ensured by the conditions spelled out here. Furthermore, most models only consider positive spatial autocorrelation in this context. An exception is the so-called wave variogram, which allows both positive and negative correlation due to the periodicity of the process. This model has not seen application outside the physical sciences and is not considered here.
where, because of the scaling factor $\sigma^2$, the matrix $\Omega(d_{ij}, \varphi)$ must be a positive definite spatial correlation matrix, with $\omega_{ii} = 1$ and $|\omega_{ij}| \leq 1$, $\forall i, j$.\(^{22}\)

Estimation of regression models that incorporate an error covariance of this form has been explored in the statistical literature by Mardia and Marshall (1984), Warnes and Ripley (1987), Mardia and Watkins (1989), and Mardia (1990), among others. In spatial econometrics, models of this type have been used primarily in the analysis of urban housing markets, e.g., in Dubin (1988, 1992) and Olmo (1995). While this specification has a certain intuition, in the sense that it incorporates an explicit notion of spatial clustering as a function of the distance separating two observations (i.e., positive spatial correlation), it is also fraught with a number of estimation and identification problems. Mardia and Marshall (1984, pp. 138–139) spell out the assumptions under which maximum likelihood yields a consistent and asymptotically normal estimator. For the purposes of the discussion here, it suffices to note that the error covariance matrix should be twice differentiable and continuous in the parameters $\varphi$. In addition, the elements of the covariance matrix, and their first and second partial derivatives should be absolutely summable.\(^{23}\) The importance of these conditions is that they impose constraints on the parameter space $\Phi$ as well as on the type of distance decay functions $f$.

Several specifications suggested in the literature do not meet these estimability and identifiability conditions. For example, as illustrated in Mardia (1990, p. 212), the popular geostatistical spherical model with range parameter is not twice differentiable. In spatial econometrics, Dubin (1988, 1992) proposed the use of a negative exponential correlogram. In her specification, the error covariance matrix is of the form (Dubin, 1988, p. 467)

$$E[\varepsilon \varepsilon'] = \sigma^2 K$$

with

$$K_{ij} = \exp(-d_{ij}/b^2),$$

where $d_{ij}$ is the distance between observations $i$ and $j$, and $b^2$ is a parameter. In Dubin (1992, Eq. (2))\(^{24}\) a slight variant is introduced that contains an additional parameter $b_1$, such that

$$K_{ij} = b_1 \exp(-d_{ij}/b^2).$$

Clearly, $\sigma^2$ in (4.3) and $b_1$ in (4.5) are not separately identifiable. Dubin (1992, p. 445) suggests the use of LR tests for spatial autocorrelation, either as a test

\(^{22}\) $\omega_{ii} = 1$ is ensured by selecting a functional form for $f$ such that $f(d_{ij}, \varphi) = 1$ for $d_{ij} = 0$. Also, in contrast to what holds for the spatial process models discussed in Section 2, the error terms $\varepsilon_i$ are not heteroskedastic in this specification, unless this heteroskedasticity is introduced explicitly. In order to focus the discussion on the aspects of spatial autocorrelation, this is excluded here.

\(^{23}\) These conditions are similar in spirit to the conditions formulated in Magnus (1978) and Mandy and Martins-Filho (1994) for general non-spherical error models. However, in the spatial case, the standard regularity conditions do not apply directly and further conditions are needed as well (see also footnote 6).

\(^{24}\) Specification (4.5) was originally suggested in Dubin (1988, footnote 8) but not implemented in that paper.
on the significance of the parameter $b_2$ in (4.4) (Dubin, 1988, p. 473) or on the joint significance of $b_1$ and $b_2$ in (4.5) (Dubin, 1992, p. 445). However, upon closer examination, it turns out that this approach is problematic. Apart from the identification problem in (4.5), a null hypothesis of the form $H_0: b_2 = 0$ does not correspond to an interior point of the parameter space and hence does not satisfy the regularity conditions for ML, similar to the situation encountered in Section 3. At first sight, this could be fixed by means of a straightforward reparametrization to $K_{ij} = \exp(-b_3d_{ij})$, but under the null of $b_3 = 0$, $K_{ij} = 1$, $\forall i, j$, yielding a singular error covariance matrix. Hence, this null hypothesis does not correspond to the absence of spatial autocorrelation.25

Mardia and Marshall (1984, p. 141) suggest the use of an error covariance specification that avoids some of these identification problems:

$$E[\hat{e}'e'] = \sigma^2_1 I + \sigma^2_2 P_x$$  \hspace{1cm} (4.6)

with $\sigma^2_1 \geq 0$, $\sigma^2_2 \geq 0$, and $P_x$ as a parameterized spatial correlation matrix.26 For ease of interpretation, this can also be written as

$$E[\hat{e}'e'] = \sigma^2 \Omega(\gamma, \phi) = \sigma^2 [I + \gamma \Psi(d_{ij}, \phi)],$$  \hspace{1cm} (4.7)

where $\theta = [\sigma^2, \phi, \gamma]'$ is the parameter vector, with $\sigma^2 = \sigma^2_1 > 0$, $\gamma = \sigma^2_2/\sigma^2_1 \geq 0$ and the elements of the matrix $\Psi(d_{ij}, \phi)$ are such that $|\Psi_{ij}| \leq 1$, $\forall i \neq j$ and $\Psi_{ii} = 0$.27 A test against spatial autocorrelation can be formulated as a test on the null hypothesis $H_0: \gamma = 0$. As in Section 3, this is the situation where the parameter value is on the boundary of the parameter space.28 However, in addition to this non-regular condition, the parameters $\phi$ are not identified under the null, resulting in a singular information matrix. The problem where the nuisance parameters are only identified under the alternative hypothesis was first treated by Davies (1977, 1987), who suggested an approximate solution. This is considered more closely in the next sections.

Finally, it is important to note that the parameter space $\Phi$ not only depends on the functional form for the distance decay function, but also on the metric and scale used for distance itself.

4.2. Score test

To focus the discussion, consider a negative exponential distance decay function for the spatial correlation in (4.7), such that the off-diagonal elements $\Psi_{ij} = e^{-\phi d_{ij}}$, with

---

25 For the latter to follow, the error covariance matrix should be diagonal under the null, or, $K_{ij} = 0$, $\forall i, j$ and $i \neq j$.

26 The constraints on the parameters are imposed to ensure a positive definite spatial correlation function. Taking $\sigma^2_2 \geq 0$ limits the conditions on positive definiteness to restrictions on the functional forms and parameters in $P_x$, without loss of generality. While a model such as (4.6) can be estimated in unconstrained form, this may lead to unacceptable results (negative variance estimates) and does not guarantee a ‘valid’ spatial covariance following the terminology of Cressie (1993, pp. 84–86).

27 The latter condition is imposed without loss of generality to simplify the treatment of the error variance and to avoid problems with defining a distance decay for distance equal zero.

28 In a slightly different context dealing with a LR statistic, this problem was also noted by Mardia (1990, p. 245).
\( \varphi \geq 0 \). By convention, the diagonal elements of the matrix \( \Psi \) are set to zero. Under the null hypothesis, \( H_0: \gamma = 0 \), and \( E[\varepsilon, \varepsilon'] = \sigma^2 I \). This specification therefore expresses a form of spatial autocorrelation that follows a smooth (negative exponential) decay with distance and satisfies the requirements (under the proper constraints on the parameters) for a valid spatial correlogram. As in (4.4), a parameter constraint \( \varphi = 0 \) alone does not yield a diagonal error covariance matrix and hence does not form a proper basis for a test for spatial correlation. Ignoring for a moment the non-identifiability, it follows that (as in Section 3) to obtain a RS statistic, only the score and information matrix elements corresponding to the error parameters \( \Theta = [\sigma^2, \varphi, \gamma]' \) need to be considered.

The log-likelihood for this problem is (3.4), with \( \Omega = I + \gamma \Psi(\varphi) \). The corresponding expression for the score vector \( d = \partial L / \partial \Theta \) is

\[
d = \begin{bmatrix}
- \frac{n}{2 \sigma^2} + \frac{1}{2 \sigma^2} \varepsilon' \Omega^{-1} \varepsilon \\
\frac{1}{2} \gamma \text{tr} \Omega^{-1} \Psi \varphi - \frac{1}{2 \sigma^2} \gamma' \Omega^{-1} \Psi \varphi \Omega^{-1} \varepsilon \\
- \frac{1}{2} \text{tr} \Omega^{-1} \Psi + \frac{1}{2 \sigma^2} \varepsilon' \Omega^{-1} \Psi \Omega^{-1} \varepsilon
\end{bmatrix},
\]

(4.8)

where \( \Psi_\varphi \) is a matrix with off-diagonal elements \( \varepsilon^{-2 \varphi d_i d_j} \) (and zeros on the diagonal), \( \varepsilon = \gamma - X\beta \), and \( \text{tr} \) is the trace operator. Under the null hypothesis, \( \gamma = 0 \), and thus the element corresponding to \( \partial L / \partial \varphi \) in (4.8) is always zero, precluding its estimation.

Furthermore, consider the relevant part of the information matrix, \( J_{00} = -E[\partial^2 L / \partial \Theta \partial \Theta'] \):

\[
J_{00} = \begin{bmatrix}
\frac{n}{2 \sigma^2} & -\frac{1}{2 \sigma^2} \gamma \text{tr} \Omega^{-1} \Psi \varphi & \frac{1}{2 \sigma^2} \text{tr} \Omega^{-1} \Psi \\
-\frac{1}{2 \sigma^2} \gamma \text{tr} \Omega^{-1} \Psi \varphi & \frac{1}{2} \gamma^2 \text{tr} \Omega^{-1} \Psi \varphi \Omega^{-1} \Psi \varphi & -\frac{1}{2} \gamma \text{tr} \Omega^{-1} \Psi \varphi \Omega^{-1} \Psi \\
\frac{1}{2 \sigma^2} \text{tr} \Omega^{-1} \Psi & -\frac{1}{2} \gamma \text{tr} \Omega^{-1} \Psi \varphi \Omega^{-1} \Psi & \frac{1}{2} \text{tr} \Omega^{-1} \Psi \Omega^{-1} \Psi
\end{bmatrix}.
\]

(4.9)

Under the null hypothesis, with \( \gamma = 0 \), \( \Omega^{-1} = I \), and keeping in mind that \( \text{tr} \Psi = 0 \), (4.9) would reduce to

\[
J(\theta_0) = \begin{bmatrix}
\frac{n}{2 \sigma^2} & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & \frac{1}{2} \text{tr} \Psi \Psi
\end{bmatrix}
\]

(4.10)

which is clearly singular, again highlighting the identification problem for \( \varphi \).

Assuming a value for \( \varphi \), say \( \varphi_0 \) (and thus avoiding the need for estimation), a score test for spatial correlation conditional upon this value can be constructed in the usual fashion. Using the results for the score in (4.8) and the diagonal \( 2 \times 2 \) information matrix that results after eliminating the elements corresponding to \( \varphi \) from (4.10), the RS statistic follows as

\[
\text{RS}(\varphi_0) = \left( \frac{\varepsilon' \Psi \varepsilon}{\sigma^2} \right)^2 / (2 \text{tr} \Psi \Psi)
\]

(4.11)
with \( e \) as a vector of OLS residuals, \( \hat{\sigma}^2 = e'e/n \), and, under \( H_0 \), \( \text{RS}(\phi_0) \xrightarrow{D} \chi^2(1) \).\(^{29}\) The elements of the matrix \( \Psi \) are evaluated for \( \phi = \phi_0 \). The limiting distribution is obtained for any choice of \( \phi \), but clearly the small sample properties of the test will be affected (Godfrey, 1988, p. 90). An alternative approach, which has received considerable attention in the mainstream econometric literature, can be based on the approximation outlined in Davies (1977, 1987).

### 4.3. Davies approximation

The central problem in this situation is the fact that the unknown parameter \( \phi \) enters the score test through \( \Psi \). Davies (1977, 1987) suggested a test statistic for this situation which is of the form

\[
M = \sup \{ \text{RS}(\phi); L \leq \phi \leq U \}
\]

or, \( M \) is the maximum of the score test \( \text{RS}(\phi) \) for all possible parameter values of \( \phi \) in the interval \([L, U]\).\(^{30}\) The statistic \( M \) does not have an asymptotic \( \chi^1(1) \) distribution under the null, and if it is considered as such, it will lead to a Type I error probability that is too high.\(^{31}\) However, as shown in Davies (1987, p. 36), in general, an upper bound for the significance level of \( M \) may be approximated by

\[
q = \text{Prob}(\chi^2(s) > M) + VM^{(1/2)(s-1)}e^{-(1/2)M^2}2^{-(1/2)s}/\Gamma(\frac{1}{2}s),
\]

where \( s \) corresponds to the degrees of freedom of the test. For the RS test against spatial correlation, \( s = 1 \), so that (4.13) simplifies to:

\[
q = \text{Prob}(\chi^2(1) > M) + V e^{(1/2)M}/\sqrt{2}\Gamma(1/2),
\]

where \( \sqrt{2}\Gamma(1/2) = 2.51 \). The term \( V \) is approximated by

\[
V \approx |\text{RS}^{1/2}(\phi_1) - \text{RS}^{1/2}(\phi_L)| + |\text{RS}^{1/2}(\phi_2) - \text{RS}^{1/2}(\phi_1)| + \ldots + |\text{RS}^{1/2}(\phi_U) - \text{RS}^{1/2}(\phi_m)|,
\]

where \( \phi_1, \phi_2, \ldots, \phi_m \) are the turning points of \( \text{RS}^{1/2}(\phi) \) (for further details, see Davies, 1987, pp. 35–36). This approximation can be carried out in a straightforward but computationally intensive manner by evaluating the \( \text{RS}(\phi) \) statistics for \( m + 2 \) values of \( \phi \) in a proper interval \([L, U]\), finding the supremum and computing the adjustment term \( V \).

\(^{29}\) Again, this is not strictly valid, but assumes that the general results by Davies hold in the spatial case as well.

\(^{30}\) Note that a similar approach can be formulated for the normal case as well. An extensive discussion of the Davies approach to testing when the nuisance parameter is only identified under the alternative is given in Godfrey (1988, pp. 87–91), Bera and Higgins (1992), Bera and Ra (1995), and Bera et al. (1998), among others. Alternative strategies are outlined in Gouriéroux et al. (1982), King and Shively (1993), Andrews and Ploberger (1994) and Hansen (1996). Their consideration in the spatial case merits further attention, but is outside the scope of the present paper.

\(^{31}\) For an illustration, see the results of the simulation experiments in Bera and Higgins (1992) and Bera and Ra (1995).
4.4. Small sample performance

A limited number of Monte Carlo simulation experiments are carried out to assess the performance of four alternative testing procedures: the $M$ statistic, i.e., the supremum in (4.12) evaluated as if it had an asymptotic $\chi^2(1)$ distribution under the null, the Davies approximation to the upper bound for the significance level, (4.14), and the $RS(\phi)$ statistic evaluated for values of $\phi$ of, respectively, 1.0 and 2.0. As in Section 3, the data are generated over a regular square lattice with coordinates ranging from $(1,1)$ to $(m,m)$, where $m$ is the size of the lattice (the number of observations, $n=x^2$). The observations are generated as in Section 3, for the same sample sizes, $X$ matrix and error distribution, except for the error covariance, which is specified as in (4.7). The spatial correlation follows a negative exponential function $\psi_{ij} = e^{-\phi d_{ij}}$, and the Euclidean distances between observations range from 1 to $(m-1)\sqrt{2}$. Without loss of generality, the error variance is set to 1.0. The lower and upper bound for the parameter $\phi$ were set to 0.0 and 3.0. All simulations were carried out for 10,000 replications under the null hypothesis and 1000 replications under the alternatives.

Under the alternative, the allowable parameter space for $\gamma, \phi$ is determined by the condition $0 \leq e^{-\phi d_{ij}} \leq 1$ and the requirement for $\Omega$ to be positive definite. This is not trivial, since as $\phi \rightarrow 0$, $e^{-\phi d_{ij}} \rightarrow 1$ and $\Omega$ becomes near-singular. Also, $e^{-\phi}$ is the spatial correlation between nearest neighbors, i.e., lattice vertices that are one unit of distance apart. Increasing values of $\gamma$ combine with decreasing (in absolute value) values of $\phi$ to yield higher degrees of spatial correlation. For the simulations under the alternative hypothesis, $\gamma$ was set to 0.10, 0.5 and 1.0, and $\phi$ to 0.10, 0.5, 1.0 and 2.0. This yields a range of spatial correlation between nearest neighbors going from 0.01 to 0.09 for $\gamma = 0.10$, from 0.07 to 0.48 for $\gamma = 0.5$ and from 0.14 to 0.91 for $\gamma = 1$. The power of the statistics can thus be evaluated in function of an intuitive and quantifiable notion of first order spatial autocorrelation.

The empirical Type I errors for the four tests are reported in Table 2, for nominal significance levels of 0.05 and 0.01. The evidence provided in the table is only preliminary, since a number of issues remain to be investigated, such as the sensitivity of the results to the range for $\phi$. However, it is clear that the naive $M$ test yields unacceptably high rejection levels, a finding which conforms to most other results in the literature (see, e.g., Bera and Higgins, 1992; Bera and Ra, 1995). The Davies-adjusted

---

Table 2
Empirical rejection levels under the null hypothesis RS tests against direct representation

<table>
<thead>
<tr>
<th>$n$</th>
<th>0.05</th>
<th>0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$D$</td>
</tr>
<tr>
<td>25</td>
<td>0.0751</td>
<td>0.0242</td>
</tr>
<tr>
<td>49</td>
<td>0.0962</td>
<td>0.0372</td>
</tr>
<tr>
<td>81</td>
<td>0.1053</td>
<td>0.0418</td>
</tr>
<tr>
<td>121</td>
<td>0.1121</td>
<td>0.0427</td>
</tr>
<tr>
<td>400</td>
<td>0.1389</td>
<td>0.0506</td>
</tr>
</tbody>
</table>
Table 3
Power of tests against direct representation

<table>
<thead>
<tr>
<th>n</th>
<th>γ</th>
<th>φ</th>
<th>0.05</th>
<th></th>
<th></th>
<th>0.01</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>D</td>
<td>RS(1)</td>
<td>RS(2)</td>
<td>D</td>
<td>RS(1)</td>
<td>RS(2)</td>
</tr>
<tr>
<td>25</td>
<td>0.1</td>
<td>2.0</td>
<td>0.018</td>
<td>0.028</td>
<td>0.035</td>
<td>0.005</td>
<td>0.004</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.029</td>
<td>0.046</td>
<td>0.053</td>
<td>0.004</td>
<td>0.010</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.039</td>
<td>0.056</td>
<td>0.059</td>
<td>0.012</td>
<td>0.020</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.022</td>
<td>0.043</td>
<td>0.042</td>
<td>0.008</td>
<td>0.010</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>2.0</td>
<td>0.036</td>
<td>0.049</td>
<td>0.059</td>
<td>0.012</td>
<td>0.018</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.069</td>
<td>0.101</td>
<td>0.089</td>
<td>0.024</td>
<td>0.039</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.113</td>
<td>0.176</td>
<td>0.119</td>
<td>0.058</td>
<td>0.100</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.079</td>
<td>0.130</td>
<td>0.108</td>
<td>0.037</td>
<td>0.057</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>2.0</td>
<td>0.048</td>
<td>0.070</td>
<td>0.072</td>
<td>0.017</td>
<td>0.027</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.143</td>
<td>0.217</td>
<td>0.144</td>
<td>0.069</td>
<td>0.124</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.247</td>
<td>0.352</td>
<td>0.226</td>
<td>0.157</td>
<td>0.230</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.318</td>
<td>0.409</td>
<td>0.297</td>
<td>0.237</td>
<td>0.303</td>
<td>0.193</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>2.0</td>
<td>0.052</td>
<td>0.066</td>
<td>0.074</td>
<td>0.023</td>
<td>0.027</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.124</td>
<td>0.147</td>
<td>0.106</td>
<td>0.063</td>
<td>0.074</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.248</td>
<td>0.274</td>
<td>0.140</td>
<td>0.163</td>
<td>0.178</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.238</td>
<td>0.272</td>
<td>0.146</td>
<td>0.143</td>
<td>0.169</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>2.0</td>
<td>0.065</td>
<td>0.079</td>
<td>0.073</td>
<td>0.028</td>
<td>0.040</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.199</td>
<td>0.229</td>
<td>0.144</td>
<td>0.128</td>
<td>0.141</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.467</td>
<td>0.468</td>
<td>0.258</td>
<td>0.377</td>
<td>0.354</td>
<td>0.161</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.704</td>
<td>0.612</td>
<td>0.397</td>
<td>0.587</td>
<td>0.528</td>
<td>0.272</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>2.0</td>
<td>0.048</td>
<td>0.057</td>
<td>0.068</td>
<td>0.019</td>
<td>0.019</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.048</td>
<td>0.068</td>
<td>0.066</td>
<td>0.027</td>
<td>0.028</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.076</td>
<td>0.081</td>
<td>0.065</td>
<td>0.036</td>
<td>0.037</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.079</td>
<td>0.093</td>
<td>0.072</td>
<td>0.041</td>
<td>0.042</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>2.0</td>
<td>0.075</td>
<td>0.079</td>
<td>0.071</td>
<td>0.038</td>
<td>0.038</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.150</td>
<td>0.160</td>
<td>0.092</td>
<td>0.093</td>
<td>0.092</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.334</td>
<td>0.283</td>
<td>0.150</td>
<td>0.235</td>
<td>0.187</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.405</td>
<td>0.347</td>
<td>0.171</td>
<td>0.299</td>
<td>0.256</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>2.0</td>
<td>0.094</td>
<td>0.112</td>
<td>0.080</td>
<td>0.051</td>
<td>0.058</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.296</td>
<td>0.276</td>
<td>0.154</td>
<td>0.205</td>
<td>0.176</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.587</td>
<td>0.487</td>
<td>0.259</td>
<td>0.507</td>
<td>0.384</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.871</td>
<td>0.695</td>
<td>0.412</td>
<td>0.808</td>
<td>0.609</td>
<td>0.310</td>
</tr>
<tr>
<td></td>
<td>121</td>
<td>0.1</td>
<td>2.0</td>
<td>0.044</td>
<td>0.046</td>
<td>0.052</td>
<td>0.018</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0</td>
<td>0.059</td>
<td>0.077</td>
<td>0.071</td>
<td>0.029</td>
<td>0.029</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.100</td>
<td>0.089</td>
<td>0.055</td>
<td>0.053</td>
<td>0.040</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.127</td>
<td>0.113</td>
<td>0.075</td>
<td>0.073</td>
<td>0.054</td>
<td>0.028</td>
</tr>
</tbody>
</table>
significance levels seem to provide a slight under-rejection for $p = 0.05$ to a slight over-rejection for $p = 0.01$. For the former, the empirical rejection frequency is within two standard deviations of the nominal level only for $n = 400$, but for the latter, the two standard deviation band is exceeded in all but the smallest sample size. This merits further attention. For a Type I error of 0.05, the ad hoc choice of a value for $\phi$ results in a systematic under-rejection for RS($\phi$). In all but the largest sample size. For 0.01, the empirical rejection levels are within the two standard deviation range for $n \geq 81$. This seems to indicate that the RS($\phi$) statistic indeed follows the expected null distribution, but only for moderate to large sample sizes.

The Davies-adjusted test and the two ad hoc RS($\phi$), statistics are compared in terms of power in Table 3. First, consider the ad hoc tests. There is some evidence that a choice for $\phi = 1$ when this is in fact the true parameter value yields a higher empirical rejection frequency than either of the other tests, but this only holds for the smallest sample sizes. By $n = 121$, this advantage has largely disappeared. Of course, it is unrealistic to expect that the researcher knows the true value of this parameter. Moreover, the superior power pertains to parameter combinations that yield very small levels of first order spatial correlation, so its advantage is largely illusory in practice. The $D$ statistic does not perform well in the smallest sample, even when the implied

---

Table 3 (continued)

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\gamma$</th>
<th>$\phi$</th>
<th>0.05</th>
<th> </th>
<th>0.01</th>
<th> </th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$D$</td>
<td>$RS(1)$</td>
<td>$RS(2)$</td>
<td>$D$</td>
</tr>
<tr>
<td>0.5</td>
<td>2.0</td>
<td>0.068</td>
<td>0.090</td>
<td>0.072</td>
<td>0.033</td>
<td>0.043</td>
</tr>
<tr>
<td>1.0</td>
<td>0.195</td>
<td>0.172</td>
<td>0.098</td>
<td>0.118</td>
<td>0.100</td>
<td>0.047</td>
</tr>
<tr>
<td>0.5</td>
<td>0.385</td>
<td>0.277</td>
<td>0.148</td>
<td>0.289</td>
<td>0.185</td>
<td>0.081</td>
</tr>
<tr>
<td>0.1</td>
<td>0.581</td>
<td>0.438</td>
<td>0.185</td>
<td>0.500</td>
<td>0.330</td>
<td>0.113</td>
</tr>
<tr>
<td>1.0</td>
<td>2.0</td>
<td>0.093</td>
<td>0.107</td>
<td>0.073</td>
<td>0.048</td>
<td>0.049</td>
</tr>
<tr>
<td>1.0</td>
<td>0.361</td>
<td>0.290</td>
<td>0.165</td>
<td>0.275</td>
<td>0.192</td>
<td>0.100</td>
</tr>
<tr>
<td>0.5</td>
<td>0.679</td>
<td>0.492</td>
<td>0.240</td>
<td>0.574</td>
<td>0.391</td>
<td>0.147</td>
</tr>
<tr>
<td>0.1</td>
<td>0.952</td>
<td>0.742</td>
<td>0.437</td>
<td>0.919</td>
<td>0.661</td>
<td>0.336</td>
</tr>
<tr>
<td>400</td>
<td>0.1</td>
<td>2.0</td>
<td>0.066</td>
<td>0.070</td>
<td>0.070</td>
<td>0.031</td>
</tr>
<tr>
<td>1.0</td>
<td>0.093</td>
<td>0.077</td>
<td>0.068</td>
<td>0.056</td>
<td>0.041</td>
<td>0.034</td>
</tr>
<tr>
<td>0.5</td>
<td>0.171</td>
<td>0.107</td>
<td>0.090</td>
<td>0.093</td>
<td>0.061</td>
<td>0.031</td>
</tr>
<tr>
<td>0.1</td>
<td>0.400</td>
<td>0.137</td>
<td>0.079</td>
<td>0.319</td>
<td>0.072</td>
<td>0.039</td>
</tr>
<tr>
<td>0.5</td>
<td>2.0</td>
<td>0.089</td>
<td>0.093</td>
<td>0.086</td>
<td>0.041</td>
<td>0.042</td>
</tr>
<tr>
<td>1.0</td>
<td>0.261</td>
<td>0.172</td>
<td>0.104</td>
<td>0.182</td>
<td>0.092</td>
<td>0.043</td>
</tr>
<tr>
<td>0.5</td>
<td>0.644</td>
<td>0.309</td>
<td>0.144</td>
<td>0.537</td>
<td>0.208</td>
<td>0.078</td>
</tr>
<tr>
<td>0.1</td>
<td>0.916</td>
<td>0.494</td>
<td>0.230</td>
<td>0.881</td>
<td>0.401</td>
<td>0.142</td>
</tr>
<tr>
<td>1.0</td>
<td>2.0</td>
<td>0.138</td>
<td>0.126</td>
<td>0.091</td>
<td>0.079</td>
<td>0.064</td>
</tr>
<tr>
<td>1.0</td>
<td>0.496</td>
<td>0.288</td>
<td>0.154</td>
<td>0.390</td>
<td>0.204</td>
<td>0.086</td>
</tr>
<tr>
<td>0.5</td>
<td>0.835</td>
<td>0.503</td>
<td>0.259</td>
<td>0.779</td>
<td>0.403</td>
<td>0.172</td>
</tr>
<tr>
<td>0.1</td>
<td>0.997</td>
<td>0.763</td>
<td>0.427</td>
<td>0.995</td>
<td>0.697</td>
<td>0.318</td>
</tr>
</tbody>
</table>

---

The $M$ statistic was not further considered since its performance under the null renders it unreliable.
spatial correlation is as high as 0.9. For example, for $p = 0.05$, with $n = 25$, $\gamma = 1.0$ and $\varphi = 0.1$, the null is only rejected in 32% of the cases. However, as the sample size increases, the power of the $D$ test increases as well and smaller degrees of first order spatial correlation are rejected in a majority of the cases. For $n = 49$ and 81, more than 50% rejection is achieved for $\gamma = 1.0$ and $\varphi < 0.5$ (correlation $> 0.6$), with $n = 121$ for $\gamma = 0.5$ and $\varphi = 0.1$ (correlation 0.48), and with $n = 400$, 40% rejection is already obtained for $\gamma = 0.1$ and $\varphi = 0.1$, with an implied first order spatial correlation of 0.10. For the highest degree of spatial correlation (0.9), the $D$ test yields rejection levels of 87% for $n = 81$, 95% for $n = 121$ and 99.7% for $n = 400$. This seems to indicate that when the problem is severe, the $D$ statistic properly detects it in even medium-size samples. Further evidence is needed to assess its performance in non-standard situations, such as non-normal error distributions.

5. Concluding remarks

The score test principle advanced in Rao’s (1948) classic paper has been shown to be particularly relevant in applications dealing with alternatives to the classic regression specification that incorporate models of spatial dependence. For the three broad alternatives considered in this paper, the RS test offers some significant advantages over other approaches.

When the alternative is in the form of a spatial autoregressive or moving average process, the RS statistic is superior to the other classic testing principles in two important respects. The first is computational, since the statistics can be constructed from the results of ordinary least squares regression. The RS tests share this property with Moran’s $I$ test and it turns out that both are asymptotically equivalent. More importantly, the RS statistics (and particularly the form robust to local misspecification) allow for the discrimination between alternatives in the form of spatial lag dependence ($Wy$) and spatial error dependence ($W\varepsilon$). This is crucial both from an inferential and from a substantive point of view, given the important differences between these models.

The RS statistic also offers some important possibilities to tackle alternatives for spatial dependence in the form of spatial error components and direct representation. The new statistics derived in this paper avoid the problems associated with the failure to meet some regularity conditions for maximum likelihood estimation. For spatial error components, the parameter value under the null hypothesis is not an interior point of the parameter space. While Wald and Likelihood Ratio statistics need to be adjusted to take into account the resulting censored nature of their distribution, this is not the case for the RS statistic. For direct representation models, the situation is slightly more complex in that not only the parameter under the null is not an interior point, but in addition the nuisance parameter is not identified under the null. This requires a Davies-type adjustment to the RS statistic.

Both new tests that are outlined in this paper perform reasonably well in a limited series of Monte Carlo simulation experiments. While they seem to be unreliable in
small samples, for data sets of more than a 100 observations they start to approach their asymptotic properties. Given the increasingly common situation in practice where ‘spatial’ data sets are large (with up to thousands of observations), this is encouraging evidence for the practitioner. However, more in depth study of the properties of these statistics in required, both in other contexts (non-normal error distribution, presence of other forms of misspecification) as well as compared to some of the recently suggested procedures that do not rely on the maximum likelihood paradigm.

In this paper, the discussion of Rao’s score principle has been limited to the classical linear regression model. Clearly, many other issues remain to be examined, such as extensions of these approaches to the space–time domain and to models that incorporate limited dependent variables. It is hoped that the review presented here will stimulate statisticians and econometricians to tackle these interesting and challenging problems.

Acknowledgements

Encouragement and helpful comments from Anil Bera, Harry Kelejian and anonymous referees are gratefully acknowledged.

References


Anselin, L., Moreno, R., 2000. Properties of tests for spatial error components. REAL Discussion paper 00-T-12, University of Illinois, Urbana-Champaign.


Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64, 247–254.

Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.


Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.


King, M., Shively, T.S., 1993. Locally optimal testing when a nuisance parameter is present only under the alternative. Rev. Econ. Statist. 75, 1–7.


