THREE ESSAYS ON SPATIAL ECONOMETRICS WITH AN EMPHASIS ON TESTING

BY

YU-HSIEN KAO

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Anil K. Bera, Chair
Professor Daniel McMillen
Professor Geoffrey Hewings
Professor Xiaofeng Shao

Abstract

Spatial Modeling has been one of the important parts in Applied Econometrics as well as Econometrics Theory in the past thirty years, not only because of the nature that the geographic locations and interactions play a crucial role in forming behavior, but also because of the challenging problems inherited from spatial dependence in Econometric models. Misspecifications of spatial dependence in regression models lead to misleading inferences and policy implications. In this dissertation I focus on issues of model specification tests which arise from the spatial structures of the data, and it contributes to the Spatial Econometric literature in two ways: *first*, the important consequences of misspecified spatial dependence in estimation, hypothesis testing, and calculation of impact effects, and *second*, the methodologies for non-standard tests in spatial regression models. I provide both econometric methods and empirical examples to demonstrate the usefulness of the proposed testing procedures.

In chapter 1 I study the behavior of standard and adjusted Rao score (RS) tests for spatial dependence in presence of negative spatial dependence. I found that the power of the standard test can be very low when there is negative spatial dependence. I also compared the features of negative autocorrelation between the time series and spatial contexts. In time series case, both the pattern of variance-covariance matrices and the power curves are symmetric for positive and negative serial correlations. This symmetry, however, is not observed in the spatial context. I applied my findings to the U.S. state government expenditure data, and found negative spatial lag dependence in U.S. state government expenditure, suggesting competitions among the state governments [Saavedra (2000); Boarnet, Marlon and Glazer (2002)]. Consistent with my theoretical derivation, the standard RS test is misleading, and under the negative spatial dependence, the values and interpretation of impact effects are also different.

When incorporating spatial dependence, the most common specification is a spatial autoregressive (AR) process, either in the dependent variable or disturbances. However, as argued in Anselin (2003), in many cases a spatial moving average (MA) is more appropriate if the mechanism of interest is a localized spatial spillover. In chapter 2 I consider the problem of testing no spatial dependence against a spatial autoregressive and moving average (ARMA) process, which allows for a *global* direct spatial effect in the dependent variable as well as an unobserved or indirect *local* spatial effect. I suggest a test procedure and the simulation results show that the proposed test has desired size and good power performance.

In chapter 3, I further study the problems of testing no spatial dependence against a spatial ARMA process in the disturbances, in the presence of spatial lag dependence. The problems of conducting such a test are twofold. First, under the null hypothesis of no spatial dependence in the disturbances, one underlying nuisance parameter is not identified. Besides, the possible presence of spatial lag dependence may affect the performance of the test. To deal with this twin-problem of nuisance parameters simultaneously, I apply the Davies (1977, 1987) procedure to the adjusted RS statistic [Anselin, Bera, Florax, and Yoon (1996)]. I conducted extensive Monte Carlo experiments to study the finite sample performance of my proposed test, and found my test has very good size and power properties in small samples and performs very well compared to other conventional RS tests. Finally I applied the test to a number of real data sets, such as the Columbus crime data [Anselin (1988); Anselin et all (1996); Sen, Bera, and Kao (2012)], Boston housing market data [Harrison and Rubinfeld (1978); Pace and Gilley (1997)], and Netherland investment data [Florax (1992); Anselin et all (1996)]. The empirical results clearly demonstrate the effectiveness of my test and the shortcomings of currently available tests.

To Manny and Ian.

Acknowledgments

My journey started from the eagerness to pursue the truth. I am grateful that God helped me to pursue this goal by leading me to the graduate school. During my studies in Economics, I learned and was always amazed at God's creation of such a sophisticated society of human beings and also grateful that He allows me to know His great power through developing knowledge to study human behavior.

I am grateful to my advisor, Professor Anil K. Bera, for his continuous encouragement and advice in all the projects. He is the role model of outstanding scholarship. Observing his enthusiasm, patience, and immense knowledge is the most important motivation for me to pursue a fruitful academic life, in all aspects.

I would also like to thank many people during the journey. I am grateful to my dissertation committee: Professor Geoffrey Hewings, Professor Daniel McMillen, and Professor Xiaofeng Shao for their insightful advice. In addition, I am grateful to Professor Hewings and his REAL family for their feedbacks on my work and the valuable experience of project cooperations.

Many thanks goes to my department, my family and friends for their continuous support. I am grateful to have many friends -Hui Wen, Jiaying Gu, Rachel Shafer among many others - to help me go through the tough time in the first year of my Ph.D.; Hsiang-Yi Yang provided the greatest help to overcome the difficulties of studying abroad; Steve and Ruth Wegman are always the biggest support during the change of our family, especially the last year of my Ph.D. when this dissertation is finalized.

Finally, I wish to thank my husband, Manny, and my lovely son, Ian. Their smiles are always my comforts and encouragements. I dedicate this dissertation to them.

Table of Contents

List of Tables

List of Figures

Chapter 1

Spatial Regression: The Curious Case of Negative Spatial Dependence

1.1 Introduction

Positive spatial dependence is predominant in the spatial data. Therefore, it is not surprising that most of the methodological papers are concerned with the positive spatial dependence (either in terms of spatial lag, spatial error, or both) when evaluating estimation, testing and forecasting procedures [For example, see Anselin, Bera, Florax, and Yoon (1996) and Baltagi, Song, and Koh (2003) for testing spatial dependence.] However, prevalence of *negative* spatial dependence is not uncommon as evidenced in many applied papers; just to mention a few: the studies of welfare competition or federal grants competition among local governments [Saavedra (2000) and Boarnet and Glazer (2002)]; the studies of regional employment [Filiztekin (2009) and Pavlyuk (2011)]; the cross-border lottery shopping [Garrett and Marsh (2002)]; foreign direct investment in OECD countries [Garretsen and Peeters (2009)] and locations of Turkish manufacturing industry [Basdas (2009)]. Griffith and Arbia (2010) investigated empirical situations in which negative spatial dependence may occur and situations in which it may be masked by positive spatial autocorrelation. In particular, they presented examples of negatively spatial autocorrelated phenomena based on the geographic competition for land surface, for territory, and for market area.

The literature of theoretic explanation for the presence of negative spatial dependence is still growing slowly. Some attempts have been made in economic theory recently, especially in trade and growth theory. For example, Frank and Botolf (2007) suggest that Myrdal's *backwash* effect [Myrdal (1957)] can be used to explain the empirical finding of negative spatial autocorrelation in the German regional information and computer technology (ICT) distribution. The backwash effect discussed in Myrdal (1957) implies that growth in one region is harmful for growth in neighbor regions since it may attract resources and skilled labor from neighbor regions and reduce their growth potential. Blonigen, Davies, Waddell, and Naughton (2004) discussed theoretical models for different kinds of Foreign Direct Investment (FDI) and their theory predicts negative spatial dependence for pure vertical FDI and export platform FDI because the production set-up from home country to host country is directly at the cost of other host countries. It appears that negative spatial autocorrelation is likely to occur when *competition* between regions (or agents) outweigh *cooperative* factors.

In contrast to the time-series analysis, positive and negative spatial dependence can have quite different implications. Consider a simple first-order autoregressive model,

$$y_t = \rho y_{t-1} + \epsilon_t, \ |\rho| < 1, \ t = 1, 2, ..., T.$$

where $\epsilon_t \sim \text{IID}(0, \sigma_\epsilon^2)$, and $y_t \sim (0, \sigma_\epsilon^2)$. The variance-covariance matrix of y has the diagonal elements equal to $\sigma_\epsilon^2/(1-\rho^2)$ and off-diagonal ones $V_{ij} = \frac{\sigma^2}{1-\rho^2}\rho^{|i-j|}$. Therefore, the only difference between positive or negative autocorrelation is just the *sign* of the elements in the matrix. Thus, theoretically there is not much difference between positive or negative autocorrelation in terms of properties of the model. Moreover, previous Monte Carlo studies such as Kramer and Zeisel (1990), King (1985), L'Esperance and Taylor (1975), and Park (1975) suggest that the empirical power functions of various tests for serial autocorrelation, for example Durbin-Watson test and BLUS test, appear to be symmetric around zero and the symmetry becomes more apparent when the sample size (T) grows. In particular, Park (1975) reported the empirical power functions of Durbin-Watson and Durbin's h tests in the presence of lagged dependent variable where coefficient (β_1) was fixed at 0.5. His results, shown in Table 1.1, indicate the close-to-symmetry feature of the power for positive and negative values. Rayner (1994) repeated the Monte Carlo experiments of Park (1975) and found similar results with different values of β_1, as can be seen from Figure 1.1.

Table 1.1: Power of the Tests for Serial Autocorrelation, Table 2 of Park (1975)

ρ	-0.95	-0.80	-0.60	-0.40	-0.20	0.00	0.20	0.40	0.60	0.80	0.95
DW	0.72	0.51	0.23	0.30	0.10	0.00	0.00	0.23	0.30	0.57	0.90
h	0.89	0.84	0.76	0.34	0.15	0.12	0.21	0.48	0.62	0.82	0.95

Now consider the first-order spatial autoregressive model:

$$Y = \rho W Y + \epsilon,$$

where Y is an $(N \times 1)$ vector of observations, W is an $(N \times N)$ spatial weights matrix and $\epsilon \sim (0, I\sigma_\epsilon^2)$. The variance-covariance matrix of Y can be written as,

$$Var(Y) = (I - \rho W)^{-1}((I - \rho W)')^{-1}\sigma_\epsilon^2.$$

Because of the feature of W, the structures of the above matrix for positive or negative values of ρ can be very different. For example, when $n = 6$, $\sigma_\epsilon^2 = 1$, and W based on a 3×2 regular grid with queen criterion, the variance-covariance matrix of Y for $\rho = 0.5$ and $\rho = -0.5$ are respectively.

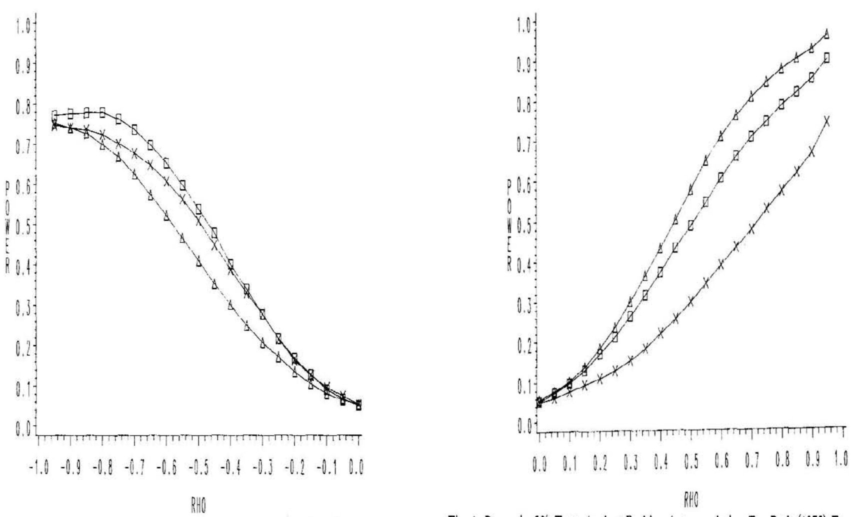


Fig. 2. Power in 5% Tests Against Negative Autocorrelation For Park (1975) Experiment #10. Triangle = d, X = h, Square = t.

Fig. 1. Power in 5% Tests Against Positive Autocorrelation For Park (1975) Experiment #10. Triangle = d, X = h, Square = t.

Figure 1.1: Power of the Tests for Serial Autocorrelation, Figures 1 and 2 of Rayner (1994)

$$\mathsf{Var}(Y) = \begin{bmatrix} 1.406 & 0.602 & 0.319 & 0.671 & 0.602 & 0.319 \\ 0.602 & 1.431 & 0.602 & 0.602 & 0.605 & 0.602 \\ 0.319 & 0.602 & 1.406 & 0.319 & 0.602 & 0.671 \\ 0.671 & 0.602 & 0.319 & 1.406 & 0.602 & 0.319 \\ 0.602 & 0.605 & 0.602 & 0.602 & 1.431 & 0.602 \\ 0.319 & 0.602 & 0.671 & 0.319 & 0.602 & 1.406 \end{bmatrix},$$

$$\mathsf{Var}(Y) = \begin{bmatrix} 1.174 & -0.230 & 0.087 & -0.266 & -0.230 & 0.087 \\ -0.230 & 1.162 & -0.230 & -0.230 & -0.072 & -0.230 \\ 0.087 & -0.230 & 1.174 & 0.087 & -0.230 & -0.266 \\ -0.266 & -0.230 & 0.087 & 1.174 & -0.230 & 0.087 \\ -0.230 & -0.072 & -0.230 & -0.230 & 1.162 & -0.230 \\ 0.087 & -0.230 & -0.266 & 0.087 & -0.230 & 1.174 \end{bmatrix}.$$

Comparing the two matrices, both in terms of *signs* and *magnitudes* of the elements, it is hard to detect any particular pattern. This is one of the motivating factors to study other properties with negative value of spatial autocorrelation.

Previous studies on testing spatial models usually concentrated only on positive spatial dependence. Anselin and Rey (1991), however, considered both positive and negative values in their Monte Carlo studies to compare the properties of Moran's I and Rao's score (RS) tests, separately for spatial error and spatial lagged dependence, and their results are shown in Figure 1.2. Though not symmetric around zero, we notice that the power of both tests increase as the true value of spatial autocorrelation coefficient move away from zero. As expected, asymmetry of the tests is more prominent for smaller sample sizes. Anselin, Bera, Florax and Yoon (1996) considered the joint presence of lag and error dependence; however, negative parameter values were excluded in their simulation study.

This chapter is concerned with the case of negative spatial dependence and its consequence on specification tests and calculation of impact effects. I will investigate how negative spatial dependence has bearings upon econometric analysis and in particular, first I will extend the theory and Monte Carlo results in the literature by including negative coefficients. I will also extend the theoretical derivation and simulations to compare various tests for spatial autocorrelation in Anselin et al. (1996). Then I will specifically show how I need to alter the standard methodologies for model specification and evaluation in the presence of negative spatial dependence.

1.2 A General Approach to testing in the presence of a nuisance parameter

Consider a general statistical model represented by the log-likelihood function $L(\gamma, \psi, \phi)$, where γ is a parameter vector, and for simplicity ψ and ϕ are taken as scalars to conform with the spatial autoregressive model. Suppose an investigator sets $\phi = 0$ and tests $H_0 : \psi = \psi_0$ using the log-likelihood function $L_1(\gamma, \psi) = L(\gamma, \psi, \phi_0)$, where ψ_0 and ϕ_0 are known values. The RS test

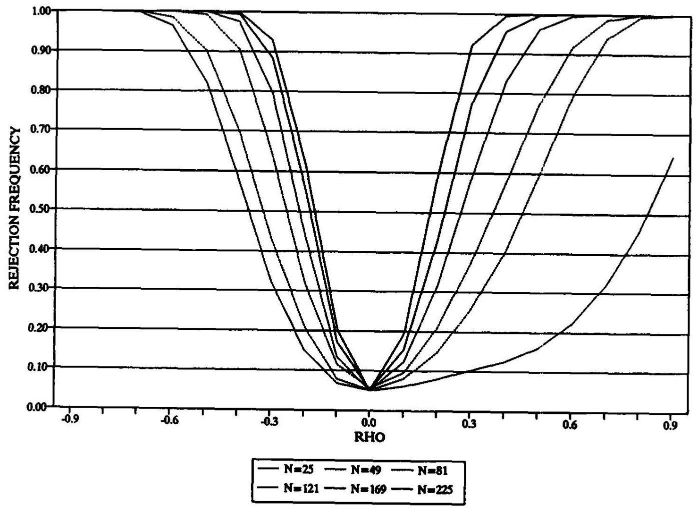


FIG. 6. Power of Moran's *I* (Normal Lag Model, W1)

(a) Power of Moran's I

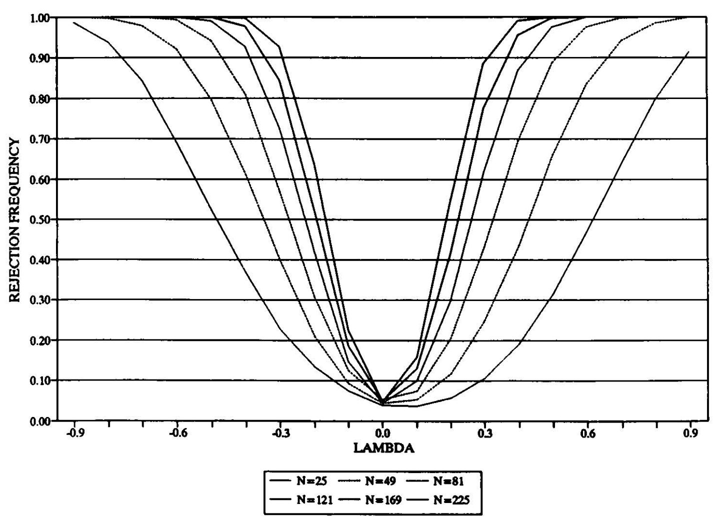


FIG. 5. Power of LM Error (Normal Error Model, W1)

(b) Power of Rao score Test

Figure 1.2: Empirical Power Functions in Anselin and Rey (1991)

statistic for testing H_0 under $L_1(\gamma, \psi)$ will be denoted by RS_ψ. Let us also denote $\theta = (\gamma', \psi', \phi')'$ and $\tilde{\theta} = (\tilde{\gamma}', \psi_0', \phi_0')$, where $\tilde{\gamma}$ is the maximum likelihood estimator (MLE) of γ when $\psi = \psi_0$ and $\phi = \phi_0$. The score vector and the information matrix are defined, respectively, as

$$d(\theta) = \frac{\partial L(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial L(\theta)}{\partial \gamma} \\ \frac{\partial L(\theta)}{\partial \psi} \\ \frac{\partial L(\theta)}{\partial \phi} \end{bmatrix}$$

and

$$J(\theta) = -E\left[\frac{1}{N}\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}\right] = \begin{bmatrix} J_\gamma & J_{\gamma\psi} & J_{\gamma\phi} \\ J_{\psi\gamma} & J_\psi & J_{\psi\phi} \\ J_{\phi\gamma} & J_{\phi\psi} & J_\phi \end{bmatrix}.$$

If $L_1(\gamma, \psi)$ were the true model, it is well known that under $H_0 : \psi = \psi_0$,

$$RS_\psi = \frac{1}{N} d_\psi(\tilde{\theta})' J^{-1}_{\psi.\gamma}(\tilde{\theta}) d_\psi(\tilde{\theta}) \xrightarrow{D} \chi^2_1(0),$$

where $J_{\psi.\gamma} = J_\psi(\theta) - J_{\psi\gamma}J_\gamma^{-1}J_{\gamma\psi}$. I use $\xrightarrow{D}$ to denote convergence in distribution. Under $H_1 : \psi = \psi_0 + \xi/\sqrt{N}$,

$$RS_\psi \xrightarrow{D} \chi^2_1(\lambda_1), \tag{1.1}$$

where the noncentral parameter $\lambda_1 = \xi' J_{\psi.\gamma}\xi$. Under the set-up, asymptotically the test will have the correct size and will be locally optimal. Now suppose that the true log-likelihood function is $L_2(\gamma, \phi) = L(\gamma, \psi_0, \phi)$, so the alternative $L_1(\gamma, \psi)$ becomes completely misspecified. Using a sequence of local values $\phi = \phi_0 + \delta/\sqrt{N}$, Davidson and MacKinnon (1987) and Saikkonen (1989) obtained the asymptotic distribution of RS_ψ under $L_2(\gamma, \phi)$ as

$$RS_\psi \xrightarrow{D} \chi^2_1(\lambda_2), \tag{1.2}$$

where $\lambda_2 = \delta' J_{\phi\psi.\gamma} J^{-1}_{\psi.\gamma} J_{\psi\phi.\gamma}\delta$, with $J_{\psi\phi.\gamma} = J_{\psi\phi} - J_{\psi\gamma}J_\gamma^{-1}J_{\gamma\phi}$.

Turning to the case of undermisspecification, let the true model be represented by the log-likelihood $L(\gamma, \psi, \phi)$. The alternative $L_1(\gamma, \psi)$ is underspecified with respect to nuisance parameter ϕ, leading to the problem of undertesting. Consider the local departure $\phi = \phi_0 + \delta/\sqrt{N}$ together with $\psi = \psi_0 + \xi/\sqrt{N}$. For this case Bera and Yoon (1991) derived the asymptotic distribution of RS_ψ,

$$RS_\psi \xrightarrow{D} \chi^2_1(\lambda_3), \tag{1.3}$$

where

$$\lambda_3 = (\delta' J_{\phi\psi.\gamma} + \xi' J_{\psi.\gamma}) J^{-1}_{\psi.\gamma} (J_{\psi\phi.\gamma}\delta + J_{\psi.\gamma}\xi)$$
$$= \lambda_1 + \lambda_2 + 2\xi' J_{\psi\phi.\gamma}\delta.$$

Using the result, I can compare the asymptotic local power of the underspecified test with that of the optimal test. It turns out that the contaminated non central parameter λ may increase or decrease the power depending on the configuration of the term $\xi' J_{\psi\phi.\gamma}\delta$.

Utilizing (1.2), adjusting the mean and variance of RS_ψ, Bera and Yoon (1993) suggested a modification so that resulting test is valid in the local presence of ϕ. The modified statistic is given by

$$\begin{aligned} RS^*_\psi = \frac{1}{N}&[d_\psi(\tilde{\theta}) - J_{\psi\phi.\gamma}(\tilde{\theta}) J^{-1}_{\phi.\gamma}(\tilde{\theta}) d_\phi(\tilde{\theta})]' \\ &\times [J_{\psi.\gamma}(\tilde{\theta}) - J_{\psi\phi.\gamma}(\tilde{\theta}) J^{-1}_{\phi.\gamma}(\tilde{\theta}) J_{\phi\psi.\gamma}(\tilde{\theta})]^{-1} \\ &\times [d_\psi(\tilde{\theta}) - J_{\psi\phi.\gamma}(\tilde{\theta}) J^{-1}_{\phi.\gamma}(\tilde{\theta}) d_\phi(\tilde{\theta})]. \end{aligned} \tag{1.4}$$

Under $\psi = \psi_0$ and $\phi = \phi_0 + \delta/\sqrt{N}$, RS^*_ψ has a *central* χ^2_1 distribution, and under the local alternative $\psi = \psi_0 + \xi/\sqrt{N}$,

$$RS^*_\psi \xrightarrow{D} \chi^2_1(\lambda_4), \tag{1.5}$$

where $\lambda_4 = \xi'(J_{\psi.\gamma} - J_{\psi\phi.\gamma} J^{-1}_{\phi.\gamma} J_{\phi\psi.\gamma})\xi$.

Similarly, I can also obtain RS^*_ϕ to test $H_0 : \phi = \phi_0$ in the presence of local misspecification and derive the noncentral parameters of RS_ϕ and RS^*_ϕ. If $L_2(\gamma, \phi)$ is the true log-likelihood function, under the null hypothesis RS_ϕ asymptotically follows central χ^2_1 distribution, and under local alternative $\phi = \phi_0 + \delta/\sqrt{N}$,

$$RS_\phi \xrightarrow{D} \chi^2_1(\lambda_5), \tag{1.6}$$

where $\lambda_5 = \delta' J_{\phi.\gamma}\delta$. In the case of complete misspecification, I have

$$RS_\phi \xrightarrow{D} \chi^2_1(\lambda_6), \tag{1.7}$$

where $\lambda_6 = \xi' J_{\psi\phi.\gamma} J^{-1}_{\phi.\gamma} J_{\phi\psi.\gamma}\xi$. And in the case of undermisspecification,

$$RS_\phi \xrightarrow{D} \chi^2_1(\lambda_7), \tag{1.8}$$

where $\lambda_7 = \lambda_5 + \lambda_6 + 2\delta' J_{\phi\psi.\gamma}\xi$.

On the other hand, the adjusted RS test statistic for testing $H_0 : \phi = \phi_0$ will follow a central χ^2_1 distribution under the null hypothesis even in the presence of locally misspecification of ψ. And under the local alternative $\phi = \phi_0 + \delta/\sqrt{N}$,

$$RS^*_\phi \xrightarrow{D} \chi^2_1(\lambda_8). \quad (1.9)$$

where $\lambda_8 = \delta'(J_{\phi.\gamma} - J_{\phi\psi.\gamma}J^{-1}_{\psi.\gamma}J_{\psi\phi.\gamma})\delta$.

1.3 Tests for SARMA Model

To make the study comparable to previous literature on spatial analyses, I consider a general model, the mixed regressive spatial autoregressive moving average (SARMA) model, as specified in Anselin et al. (1996):

$$\begin{aligned} y &= X\gamma + \phi W y + u, \\ u &= \psi W \epsilon + \epsilon, \\ \epsilon &\sim N(0, \sigma^2 I), \end{aligned} \quad (1.10)$$

where y is an $(n \times 1)$ vector of observations of dependent variable, X is an $(n \times k)$ matrix of observations of exogenous variables, and γ is a $(k \times 1)$ vector of parameters. ϕ and ψ are scalar spatial parameters, and W is a $(n \times n)$ spatial weights matrix.

We are interested in testing $H_0 : \psi = 0$ in the presence of the nuisance parameter ϕ. Let $\theta = (\gamma', \psi, \phi)'$, following the result of Anselin (1988a), we have the following equations:

$$\begin{aligned} \frac{\partial L}{\partial \gamma} &= d_\gamma = \frac{1}{\sigma^2} X' u, \\ \frac{\partial L}{\partial \psi} &= d_\psi = \frac{1}{\sigma^2} u' W u, \\ \frac{\partial L}{\partial \phi} &= d_\phi = \frac{1}{\sigma^2} u' W y, \end{aligned} \quad (1.11)$$

and

$$J = \begin{bmatrix} X'X & 0 & X'(WX\gamma) \\ 0 & \omega\sigma^2 & \omega\sigma^2 \\ (WX\gamma)'X & \omega\sigma^2 & (WX\gamma)'(WX\gamma) + \omega\sigma^2 \end{bmatrix}, \quad (1.12)$$

where $\omega = \mathrm{tr}[(W' + W)W]$. Using (1.11) and (1.12), it is easy to show:

$$\begin{aligned} J_{\psi\phi.\gamma} = J_{\psi.\gamma} = J_{\phi\psi.\gamma} &= \frac{\omega}{n}, \\ J_{\phi.\gamma} &= \frac{1}{n\sigma^2}[(WX\gamma)' M (WX\gamma) + \omega\sigma^2] \\ &= \frac{\omega}{n} + \frac{1}{n\sigma^2}(WX\gamma)' M (WX\gamma), \end{aligned} \quad (1.13)$$

where $M = I - X(X'X)^{-1}X'$. The adjusted RS statistic can be constructed as,

$$RS^*_\psi = \frac{[\tilde{u}' W \tilde{u}/\tilde{\sigma}^2 - \omega(n\tilde{J}_{\phi.\gamma})^{-1}\tilde{u}' W y/\tilde{\sigma}^2]^2}{\omega[1 - \omega(n\tilde{J}_{\phi.\gamma})^{-1}]}, \quad (1.14)$$

where $\tilde{u} = y - X\tilde{\gamma}$ are the OLS residuals, and $\tilde{\sigma}^2 = \tilde{u}'\tilde{u}/n$, and from (1.13) it follows that

$$(n\tilde{J}_{\phi.\gamma})^{-1} = \tilde{\sigma}^2[(WX\tilde{\gamma})'M(WX\tilde{\gamma}) + \omega\tilde{\sigma}^2]^{-1}.$$

The conventional one-directional test RS_ψ given in Burridge (1980) is obtained by setting $\phi = 0$ to yield

$$RS_\psi = \frac{[\tilde{u}'W\tilde{u}/\tilde{\sigma}^2]^2}{\omega}. \tag{1.15}$$

To see the behavior of RS_ψ and RS^*_ψ let us consider the case of local misspecification, i.e. $\phi = \phi_0 + \delta/\sqrt{n}$. Under the null $\psi = 0$ and alternative $\psi = \psi_0 + \xi/\sqrt{n}$, the noncentral parameters of RS_ψ are respectively

$$\lambda_2 = \frac{\omega\delta^2}{n}. \tag{1.16}$$

and

$$\begin{aligned}\lambda_3 &= \lambda_1 + \lambda_2 + 2\xi' J_{\psi\phi.\gamma}\delta \\ &= \xi'(\frac{T}{N})\xi + \delta'(\frac{T}{N})\delta + 2\xi'(\frac{T}{N})\delta \\ &= \frac{\omega}{n}(\xi^2 + \delta^2 + 2\xi\delta).\end{aligned} \tag{1.17}$$

Therefore, the noncentral parameters of RS_ψ under both the null and alternative are affected by δ, i.e., the local misspecification of ϕ. Comparing to the case where both of the spatial autocorrelation parameters are positive, the noncentral parameter is lower when they have opposite signs, and it can be as low as 0 when $\xi = -\delta$.

On the other hand, the noncentral parameter of RS^*_ψ under $\psi = \psi_0 + \xi/\sqrt{n}$ is not affected by the presence of local misspecification of ϕ, and is given by

$$\begin{aligned}\lambda_4 &= \xi'[\frac{\omega}{n} - (\frac{\omega}{n})^2 J^{-1}_{\phi.\gamma}]\xi \\ &= \frac{\omega\xi^2}{n}(1 - \frac{\omega\sigma^2}{\omega\sigma^2 + (WX\gamma)'M(WX\gamma)}).\end{aligned} \tag{1.18}$$

which depends, on ξ but free of δ, the local misspecification in ϕ.

In the presence of local misspecification, i.e. $\psi = \psi_0 + \xi/\sqrt{n}$, We can also study the performance of RS^*_ϕ and RS_ϕ, which are given by,

$$RS^*_\phi = \frac{[\tilde{u}'Wy/\tilde{\sigma}^2 - \tilde{u}'W\tilde{u}/\tilde{\sigma}^2]^2}{n\tilde{J}_{\phi.\gamma} - \omega} \tag{1.19}$$

and

$$\mathrm{RS}_{\phi} = \frac{[\tilde{u}'Wy/\tilde{\sigma}^2]^2}{n\tilde{J}_{\phi.\gamma}}. \tag{1.20}$$

Under the alternative $\phi = \phi_0 + \delta/\sqrt{N}$, the noncentral parameter of RS^*_{ϕ} and RS_{ϕ} are respectively,

$$\lambda_8 = \delta'(\frac{1}{n\sigma^2})(WX\gamma)'M(WX\gamma)\delta \tag{1.21}$$

and

$$\begin{aligned}\lambda_7 &= \delta' J_{\phi.\gamma}\delta + \xi' J_{\psi\phi.\gamma} J^{-1}_{\phi.\gamma} J_{\phi\psi.\gamma}\xi + 2\delta J_{\phi\psi.\gamma}\xi \\ &= \delta'(\frac{1}{n\sigma^2})[(WX\gamma)'M(WX\gamma) + \omega\sigma^2]\delta \\ &+ \xi'(\frac{\omega^2}{n^2})[(WX\gamma)'M(WX\gamma) + \omega\sigma^2]^{-1}\xi + 2\delta'(\frac{\omega}{n})\xi\end{aligned} \tag{1.22}$$

Again we observe that the noncentral parameter of RS_{ϕ} is affected by the combination of positive or negative values of δ and ξ, while that of RS^*_{ϕ} is free of this problem.

1.4 Empirical Applications

To gain more insights on how negative spatial dependence would affect model specification tests and estimation in practice, we examine the various test statistics and estimated parameters in the existing literature. Table 1.2 shows results from three common applications with positive spatial dependence. All of them have the common features that (i) the unadjusted one-directional tests are strongly significant and the joint tests are moderately significant, while the adjusted statistics are lower than the unadjusted ones and show less significance, (ii) estimated spatial parameters are positive and significant. These examples show that it is likely that both the unadjusted one-directional tests are spurious because of only one source of spatial dependence.

On the other hand, the cases with negative spatial dependence are more complicated. Table 1.3 summarizes some empirical results with negative spatial coefficients. As indicated in the previous section, the unadjusted test statistics can be higher or lower than the adjusted ones, depending on the combinations of the signs of two sources of spatial dependence. For example, Garret and Marsh (2002) estimated the revenue impact of cross-border lottery sales for 105 counties in Kansas and found negative spatial autocorrelation for both spatial lag and spatial error coefficients. However, it should be noted that in their study they estimated the two coefficients separately, not jointly. According to the reported values of Rao-score test statistics and the theoretical prediction in previous sections, I expect that both of the coefficients are negative.

Though there are many empirical studies that found negative spatial dependence, most of the studies only estimate one spatial autocorrelation coefficient, either spatial lag or spatial error, and reported one-directional Rao-score test result. Others consider both two kinds of spatial dependence, but estimate the coefficients separately, as we see in Garret and Marsh (2002), and

Table 1.2: Summary of Empirical Studies - Positive Dependence

	$RS_{\psi\phi}$	RS_{ψ}	RS^*_{ψ}	RS_{ϕ}	RS^*_{ϕ}	$\hat{\phi}$	$\hat{\psi}$
Anselin (1988)	9.44	5.72	0.08	9.36	3.72	0.431	0.562
	(<0.01)	(0.02)	(0.78)	(<0.01)	(0.05)	(<0.01)	(<0.01)
Florax (1992)	7.97	2.43	0.14	7.83	5.54	0.349	0.459
	(0.02)	(0.12)	(0.70)	(0.02)	(<0.01)	(<0.01)	(<0.01)
Anselin et al. (1996)	5.07	4.35	3.65	1.42	0.72	0.188	0.465
	(0.08)	(0.04)	(0.06)	(0.23)	(0.40)	(0.11)	(<0.01)

*p-values in parentheses.

Table 1.3: Summary of Empirical Studies - Negative Dependence

	$RS_{\psi\phi}$	RS_{ψ}	RS^*_{ψ}	RS_{ϕ}	RS^*_{ϕ}	$\hat{\phi}$	$\hat{\psi}$
Garret & Marsh (2002)	3.91	0.57	0.12	3.79	3.34	-0.064^*	-0.009^*
	(0.14)	(0.45)	(0.73)	(0.05)	(0.07)		
Pavlyuk(2011)	7.37	0.03	5.00	2.37	7.34	-1.91^*	
	(0.03)	(0.86)	(0.03)	(0.12)	(<0.01)		
Basdas (2009)	6.26	3.77	5.43	0.83	2.49		-0.48^*
	(0.04)	(0.05)	(0.02)	(0.36)	(0.11)		

*Estimate coefficients separately.

Basdas (2009). To further see the empirical applications of the interactions between the two kinds of spatial dependence with negative values, it is necessary to reinvestigate the data that finds negative spatial autocorrelations. Therefore, I illustrate the case of negative spatial dependence using the data of government direct expenditure for the 48 U.S. continental states, based on the empirical analyses of Case, Rosen, and Hines (1993) and Boarnet and Glazor (2001). In their studies, the question of interest is the multiplying effect of federal grants on state and local government expenditure, after controlling for the spatial dependence. The studies use a panel data from 1970-1985; however, to focus on the illustration of negative spatial dependence, I will only look at a cross sectional data set. Therefore, the data I examine contains the state and local government expenditure, grants received from federal government, and personal income per capita in 2010.

Table 1.4 presents the estimated spatial regression results which includes both spatial dependence in lagged dependent variable and error. From the results we see that the estimated spatial autocorrelation for lagged dependent variable ($\hat{\phi}$) is negative; while the estimated spatial autocorrelation for error ($\hat{\psi}$) is positive. The t statistics suggest that both spatial dependence are significant, and if we ignore the spatial feature of the data and run OLS regression, there would be an under estimation of the coefficient of federal grants, and the log-likelihood of the model would decrease.

Furthermore, Table 1.5 shows the Rao-score statistics for model specification tests of spatial dependence. The test statistics show how the negative values of spatial autocorrelation would lead to a contradictory results of unadjusted one-directional test with the joint test. From the test statistics, we can reject the joint null hypothesis: $H_0 : \phi = \psi = 0$, but we cannot reject the one-directional test of $H_0 : \phi = 0$ or $H_0 : \psi = 0$ based on unadjusted test statistics. On the other hand, the adjusted test statistics for both spatial dependence show that the coefficients are significantly different from zero, consistent with the joint test result and the t statistics in estimation result in Table 1.5. The results are similar with different specification of spatial weight matrices.

As addressed in Case et al. (1993) and Boarnet and Glazor (2001), the research interest lies in

Table 1.4: Estimation Results

Explanatory Variables	Model OLS	Rook W	Queen W
ϕ		−0.349	−0.337
		(−2.511)	(−2.426)
ψ		0.548	0.525
		(3.4172)	(3.163)
Intercept	-347.769	1029.200	956.450
	(-0.533)	(0.836)	(0.784)
State Income Per Capita	0.109	0.129	0.129
	(7.500)	(9.145)	(9.096)
Grants	1.611	1.761	1.747
	(9.634)	(12.377)	(12.222)
Sample Size	48	48	48
Log-Likelihood	-370.877	-364.979	-365.289

Notes: 1. t statistics in parenthesis.
2. Data Sources: Bureau of Census, Government Finance Series.

Table 1.5: Testing Results

	$RS_{\psi\phi}$	RS_{ψ}	RS^*_{ψ}	RS_{ϕ}	RS^*_{ϕ}
Rook W	13.781	3.080	11.099	2.683	10.701
Queen W	13.521	3.039	10.823	2.697	10.481

measuring the multiplying effect of grants from federal on state and local government expenditure, which requires a proper interpretation of the estimated coefficient. In the standard linear regression, there is a straightforward interpretation of estimated coefficients; while in the spatial regression, a change in the explanatory variable of one region may not only affect its own region, but also the neighboring regions, and in turn have another impact on the original region. This, termed as *feedback loops* in Pace and LeSage (2008, p.35), is from the nature that observation i is a neighbor of observation j, and observation j is also a neighbor of observation i. The estimated value of the coefficient includes both the effect on its own region (direct) and the feedback loops (indirect). Following Pace and LeSage (2008, p. 38-39), I can calculate summary of measures of impact effects:

$$\bar{M}(r)_{direct} = n^{-1}tr((S_r(W)))$$
$$\bar{M}(r)_{total} = n^{-1}l_n' S_r(W) l_n$$
$$\bar{M}(r)_{indirect} = \bar{M}(r)_{total} - \bar{M}(r)_{direct},$$

where $S_r(W) = (I - \phi W)^{-1}\beta_r$, β_r is the coefficient of $r-th$ explanatory variable x_r, and l_n is an $n \times 1$ vector of ones. Furthermore, the impact effect can be partitioned by order of neighbors since

$$S_r(W) \approx (I + \phi W + \phi^2 W^2 + \phi^3 W^3 + ... + \phi^q W^q)\beta_r$$

The calculated cumulative and partitioned direct, indirect and total effects based on the estimation results of government expenditure data is shown in Table 1.6. The calculation is based on rook specification of spatial weight matrix. From 1.6 we see that when there is negative spatial autocorrelation in lagged dependent variable, the direct effect would be offset by the negative impact of feedback loops. Therefore, the total effect is not as large as directly measuring the estimated coefficient.

Table 1.6: Spatial partitioning of direct, indirect and total impacts of Grants

Cumulative effects			
	Grants		
Direct effect	1.8087		
Indirect effect	−0.5040		
Total effect	1.3047		
Spatially partitioned effects			
W order	Total	Direct	Indirect
W^0	1.7605	1.7605	0.0000
W^1	−0.6150	0.0000	−0.6150
W^2	0.2148	0.0507	0.1642
W^3	−0.0751	−0.0051	−0.0699
W^4	0.0262	0.0031	0.0231
W^5	−0.0092	−0.0006	−0.0085
Cumulative	1.3023	1.8085	−0.5062

1.5 Monte Carlo Simulations

In this section I present the results of a Monte Carlo study to investigate the finite sample behavior of the tests. I focus specifically on the power of the tests and the comparison of adjusted test relative to unadjusted one. All the tests are based on estimation by OLS. To facilitate comparison with existing results I follow a structure similar to the one adopted by Anselin et al. (1996). The model under the null hypothesis of no spatial dependence is the classic regression model:

$$y = X\gamma + u$$

while under the SARMA alternative the model is specified in equation (10). The N observations on the dependent variables are generated from a vector of standard normal random variables u. The explanatory variables X, an $N \times 3$ matrix is obtained from a vector of a constant term combined with two variates drawn from a $\mathtt{uniform}(0, 10)$ distribution. The coefficients of explanatory variables (γ) is set to be a vector of ones. The matrix of explanatory variables is held fixed in the replications. For each combination of parameter values, 5,000 replication were carried out. The graphs are based on the theoretical size of 0.05, and the proportion of rejections (i.e. the proportion of times the computed test statistic exceeded its asymptotic value) is reported.

In order to make comparison with Anselin et al. (1996), the configurations used to generate spatial dependence are formally expressed in three spatial weight matrices. These correspond to sample size of 40 and 81. The weight matrices of size 40 is built from actual irregularly shaped regionalizations of the Netherland (see Florax, 1992 for more details.) The weight matrices for $N = 81$ correspond to a regular square 9×9 grid, with contiguity defined by the rook and queen criterion respectively.

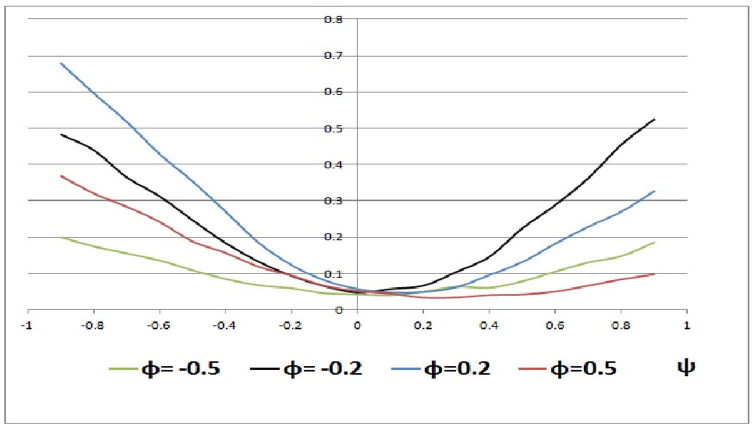
Figures 1.3 to 1.5 are the power functions of the RS^*_ψ and RS_ψ tests. As theoretical derivation predicts, the power functions of RS^*_ψ are U-shaped (symmetric around zero) under different values of ϕ, while the power functions of RS_ψ do not have symmetric feature and are sensitive to different values of ϕ. Also note that the power is extremely low when the two spatial autocorrelation coefficients have different signs but similar magnitude. Similar results can be shown under the specification of spatial weight matrices based on regular grids, which are presented in Figures 1.4 and 1.5. The adjusted test behaves well in the sense of the symmetry of the empirical power function except for the case that spatial weight matrices are built based on queen criterion and there is moderate spatial dependency in the dependent variable (i.e. the nuisance parameter, $|\phi| = 0.5$). One possible explanation is that queen criterion impose too much spatial relationship since it counts all of the 8 directions as one's neighbors, and the moderately spatial dependency strengthen the relationship further. However, since I am considering local departure of the parameters, I can still conclude that in the presence of negative spatial dependence, the adjusted Rao-score test performs better than unadjusted one in the sense of the power of the test.

As for the empirical power functions of RS^*_ϕ and RS_ϕ tests, both of them have a nicely U-shape around zero. The results are all similar under different design of spatial weight matrices, which can be seen in Figure 1.6. Still the symmetry is more apparent in RS^*_ϕ tests than RS_ϕ, but the discrepancies are not as large as the one-directional test of the hypothesis $H_0 : \psi = 0$.

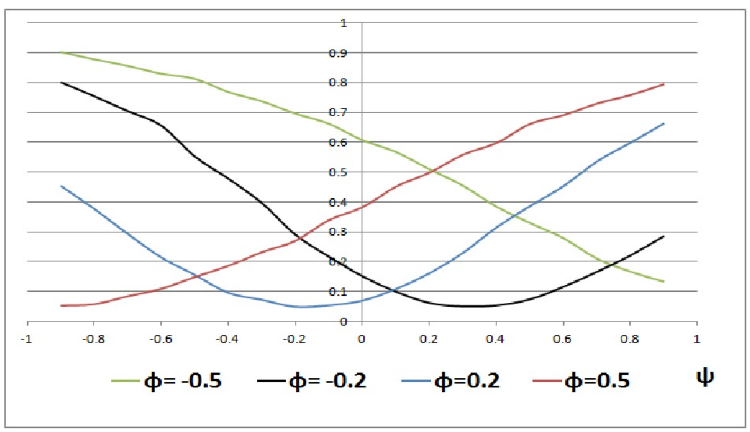
1.6 Conclusion

This chapter extends the theoretical derivation and Monte Carlo studies of model specification tests in spatial regression by examing the effect of negative spatial dependence. Previous studies focus on positive spatial autocorrelations, and therefore only address the over-sized problem of Rao-score tests. Our study suggests that under negative spatial dependence, the power of the conventional Rao-score tests can be very low, and hence, cautious is required when negative dependence is expected and the one-directional Rao-score test conclude no spatial dependence for the errors. By deriving the noncentral parameters of the asymptotic distributions of the test statistics, we are able to explain the low power of the unadjusted Rao-score test in some specific cases, and it can be shown that the power is especially low when one of the source of spatial dependence is positive while the other is negative, and they have similar magnitude. Monte Carlo results are consistent with theoretical prediction even when the sample size is finite.

There are some extensions to my study. First, since all the test statistics and the noncentral parameters include the spatial weight matrices (W), it would be important to look at how the formation of W would affect the results, especially when the Monte Carlo studies suggest that higher or lower spatial relations induced by different designs of W may make a difference, the theoretical comparison among different spatial weight matrices worths exploration. Besides, the asymmetric

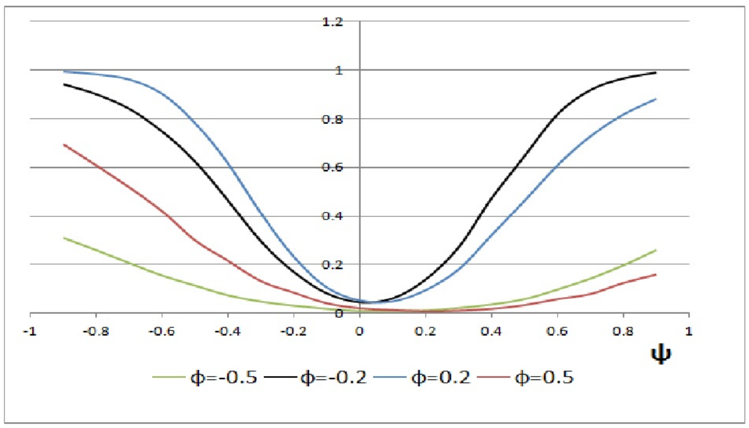


(a) Power of RS^*_ψ, $\mathrm{N} = 40$

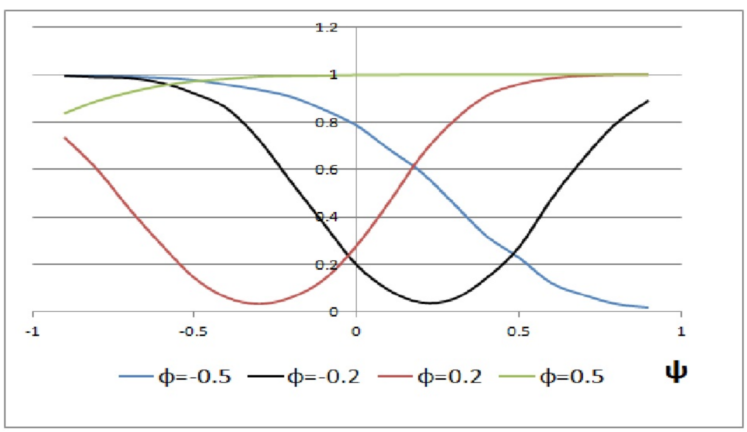


(b) Power of RS_ψ, $\mathrm{N} = 40$

Figure 1.3: Power of RS^*_ψ and RS_ψ, $\mathrm{N} = 40$

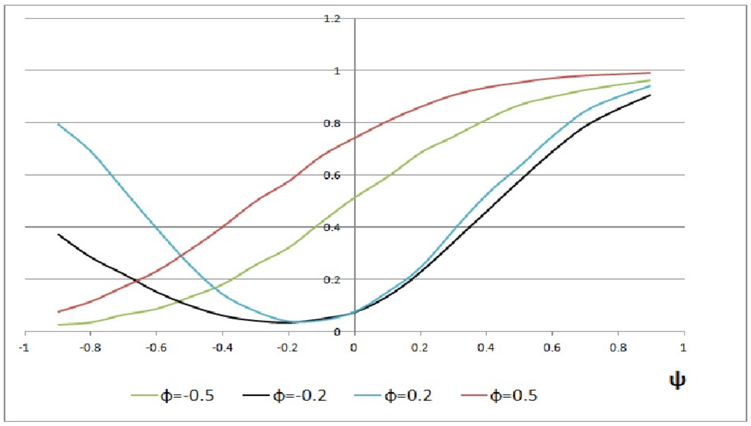


(a) Power of RS^*_ψ, $\mathsf{N} = 81$, W with rook design

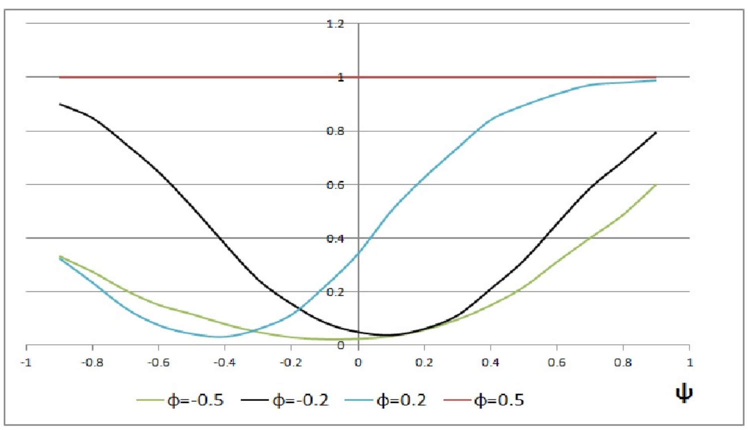


(b) Power of RS_ψ, $\mathsf{N} = 81$, W with rook design

Figure 1.4: Power of RS^*_ψ and RS_ψ, $\mathsf{N} = 81$, W with rook design

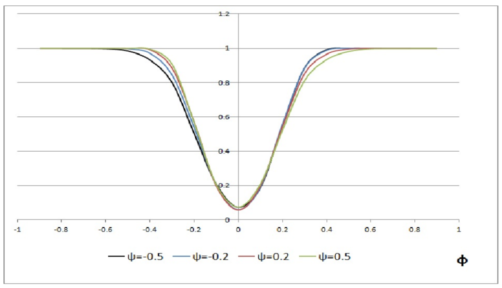


(a) Power of $\mathsf{RS}^*_\psi, \mathrm{N} = 81$, W with queen design

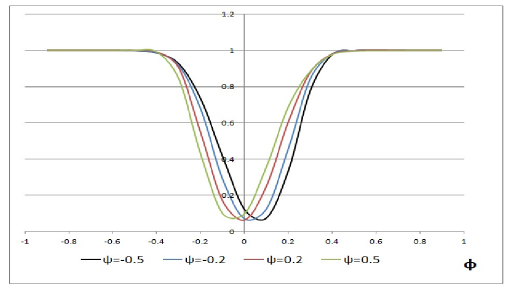


(b) Power of $\mathsf{RS}_\psi, \mathrm{N} = 81$, W with queen design

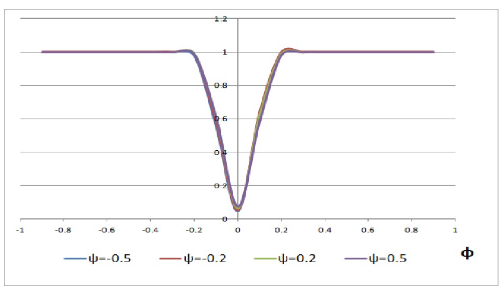
Figure 1.5: Power of RS^*_ψ and RS_ψ, $\mathrm{N} = 81$, W with queen design



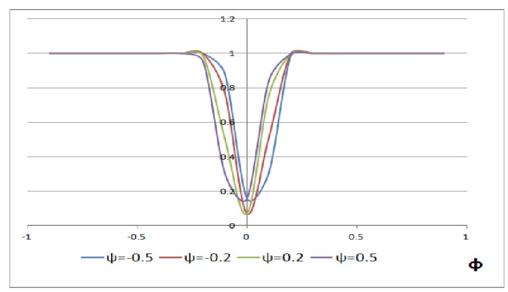
(a) Power of RS^*_ϕ, $\mathrm{N} = 40$



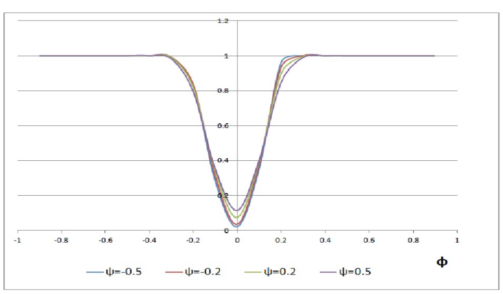
(b) Power of RS_ϕ, $\mathrm{N} = 40$



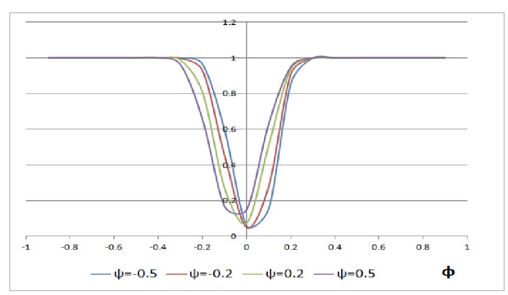
(c) Power of RS^*_ϕ, $\mathrm{N} = 81$, W with rook design



(d) Power of RS_ϕ, $\mathrm{N} = 81$, W with rook design



(e) Power of RS^*_ϕ, $\mathrm{N} = 81$, W with queen design



(f) Power of RS_ϕ, $\mathrm{N} = 81$, W with queen design

Figure 1.6: Power of RS^*_ϕ and RS_ϕ

structures of variance-covariance matrices provided in the example suggest that the information obtained from a dataset can be quite different for positive or negative spatial dependence. Therefore, it would be interesting to compare the precision of estimation based on information matrices and the variances of the estimators for the two cases. Finally, it would pose valuable applications to further study how negative spatial dependence affects the calculation of impact effects and the evaluation of model forecasting.

Chapter 2

Testing Spatial Dependence When a Nuisance Parameter is not Identified Under the Null Hypothesis

2.1 Introduction

When incorporating spatial dependence, most commonly used specifications in the literature include: (i) spatial autoregressive (SAR), (ii) spatial error model (SEM), and the combination of (i) and (ii), which was termed as spatial autocorrelation (SAC) model by Pace and LeSage (2009, p.32). Anselin (2003) interpreted the autoregressive (AR) parameters as *global* spillovers of shocks that agglomerate from higher order neighbors. For example, consider the SAC model with both AR processes in the lag dependent variable and disturbances,

$$\begin{aligned} y &= \rho W y + u \\ u &= \lambda W u + \epsilon, \end{aligned} \tag{2.1}$$

where W is the spatial weight matrix, and $\epsilon \sim \text{IIDN}(0, I\sigma^2)$. The variance-covariance matrix is given by $E(uu') = \sigma^2[(I-\lambda W)'(I-\lambda W)]^{-1} = \sigma^2[I+\lambda(W+W')+\lambda^2 WW'+\lambda^3 W(W+W')W'+...]$; therefore, the spillover effect can be relevant from higher order neighbors.

The AR specification, however, may be inappropriate or too simplistic when the research interest lies in mechanisms of *local* spillovers. Therefore, an alternative specification that allows for a localized neighboring effect may be required when there is evidence that the spillovers are not transmitted globally. Haining (1978), Anselin (1988, p.33), Hepple (2003), and Fingleton (2008a, 2008b) consider a spatial moving average (MA) process for the disturbances. Pace and LeSage (2009, p. 32-33) discuss in detail the different interpretations of spatial AR and MA processes. Also see Andersson and Gråsjö (2009) for a survey of empirical models with local or global spillover effects and interpretations of spatially interaction patterns.

Analogous to the Box-Jenkins approach in time series analysis, Anselin (1988, pp.33-34) and Anselin and Bera (1998) suggest a spatial regression specification with spatial AR lag and spatial MA in disturbances (spatial ARMA model). The specification allows for a global direct spatial effect in the dependent variable as well as an unobserved or indirect local spatial effect. In contrast to (2.1), we consider the following spatial ARMA process,

$$\begin{aligned} y &= \rho W y + u \\ u &= \epsilon - \lambda W \epsilon. \end{aligned} \tag{2.2}$$

The variance-covariance matrix now is $E(uu') = \sigma^2[I - \lambda(W + W') + \lambda^2 WW']$, i.e., neighbors up to second order are only relevant, and the range of a shock is more limited.

Compared to the vast literature on ARMA process in time series analyses, there are only a few papers that deal with the spatial ARMA model. Fingleton (2008a, 2008b) propose a generalized method of moment (GMM) estimator based on two-stage-least-squared estimation for regression models with the spatial ARMA specification. Dogan and Taspinar (2013) introduce an one-step GMM estimator for the spatial ARMA model and compare the properties of their suggested estimators with those in Fingleton (2008a, 2008b), both analytically and through Monte Carlo studies. Behrens, Ertur and Koch (2012) use a spatial ARMA specification to study bilateral trade flows on a quantity-based structural gravity equation system and apply the estimation methodology to US-Canada trade dataset. All these papers focus on estimation methods and properties of various estimators. Testing problems in the context of spatial ARMA model remain unexplored, in contrast to many studies on hypothesis testing of SAR or SEM, such as Kelejian and Prucha (2001), Baltagi, Song, Jung, and Koh (2007), and Jin and Lee (2013). Andersson and Gråsjö (2009) identified one of the problems in spatial modeling as "there are no truly well-formed spatial model." One way to resolve this model uncertainty is to consider various economically viable models and pass them through a battery of specification tests; see also Florax, Folmer and Rey (2003), Mur and Angulo (2009), and Elhorst (2014).

This chapter considers the problem of the model specification test of no spatial dependence against the alternative of a spatial ARMA process. Standard Rao's score (RS) test statistics for testing spatial dependence are locally optimal when the alternative hypothesis is either a spatial AR or MA process. However, theoretically it is not possible to derive an RS test under the ARMA alternative due to the problem that under the null hypothesis, the nuisance parameter is unidentified, and hence the information matrix becomes singular. Davies (1977) studied the problem when the parameters are not identified under the null, and Davies (1987) proposed a test procedure based on the supremum of any conventional test statistic.

Similar issues in testing when the parameters are not identified under the null have been widely studied in the time series literature. Poskitt and Tremayne (1980), Hallin, Ingenbleek and Puri (1985), Bera and Ra (1993), and Hallin and Paindaveine (2002) considered the problem of testing white noise versus an ARMA process. Andrews (1993) and Garcia and Perron (1996) dealt with testing the structural break at an unknown change-point. Balke and Fomby (1997) and Hansen (1999) studied threshold cointegration. Such issues have not been explored in the spatial context. In this chapter I consider the problem in testing no spatial dependence against the alternative of a spatial ARMA process. Following Davies (1987) I suggest a test procedure based on the supremum of RS test statistic. Power and size of the suggested test are compared with the standard RS test through simulation experiments.

2.2 The Testing Problem

Consider the testing problem of whether there is spatial dependence where y follows an spatial ARMA process, as in (2.2), i.e.,

$$y = \rho W y + \epsilon - \lambda W \epsilon. \tag{2.3}$$

When either $\rho = 0$ or $\lambda = 0$, the model reduces to a spatial MA process or a spatial AR process, respectively. When $\rho = \lambda$, y in (2.3) is simply a white noise. Therefore, to test no spatial dependence against the spatial ARMA process, we could simply test $H_0 : \rho = \lambda$ against $H_a : \rho \neq \lambda$. The log-likelihood of the spatial ARMA model is,

$$l(\theta) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 + \log|B^{-1}| + \log|A| - \frac{1}{2\sigma^2} y' A B^{-1'} B^{-1} A y, \tag{2.4}$$

where $\theta = (\rho, \lambda, \sigma^2)'$, $A = I - \rho W$ and $B = I - \lambda W$. Thus under H_0, $A = B$.

The information matrix defined as $I(\theta) = E[-\frac{\partial^2 l}{\partial\theta\partial\theta'}]$ for this model can be derived as,

$$I(\theta) = \begin{bmatrix} tr[W'B^{-1'}B^{-1}W + (A^{-1}W)^2] & -tr[A'B^{-1'}W'B^{-1'}B^{-1}W + A'B^{-1'}B^{-1}WB^{-1}W] & \sigma^{-2}tr(A'B^{-1'}B^{-1}W) \\ -tr[A'B^{-1'}W'B^{-1'}B^{-1}W + A'B^{-1'}B^{-1}WB^{-1}W] & tr[W'B^{-1'}B^{-1}W + (B^{-1}W)^2] & -\sigma^{-2}tr(B^{-1}W) \\ \sigma^{-2}tr(A'B^{-1'}B^{-1}W) & -\sigma^{-2}tr(B^{-1}W) & \frac{n}{2\sigma^4} \end{bmatrix}, \tag{2.5}$$

and under H_0 (i.e., B=A), it reduces to

$$I(\theta) = \begin{bmatrix} tr[W'A^{-1'}A^{-1}W + (A^{-1}W)^2] & -tr[W'A^{-1'}A^{-1}W + (A^{-1}W)^2] & \sigma^{-2}tr(A^{-1}W) \\ -tr[W'A^{-1'}A^{-1}W + (A^{-1}W)^2] & tr[W'A^{-1'}A^{-1}W + (A^{-1}W)^2] & -\sigma^{-2}tr(A^{-1}W) \\ \sigma^{-2}tr(A^{-1}W) & -\sigma^{-2}tr(A^{-1}W) & \frac{n}{2\sigma^4} \end{bmatrix}. \tag{2.6}$$

It is easy to see that $I(\theta)$ is singular. Therefore, it is not possible to derive a standard RS test statistic based on the inverse of the information matrix. This kind of non-standard test for other problems has been studied by previous literature, Davies (1977) first introduced this problem and Davies (1977, 1987) proposed a test procedure based on the supremum of conventional test statistics. Watson and Engle (1985) test constant versus time-varying coefficients based on Davis' procedure, and further discuss the distribution of the test statistic. King and Shively (1991) propose a reparametirization technique to approach the problem. Andrews and Ploberger (1994) suggest a test based on weighted average power criterion, and discuss the condition where their test reduces to a standard likelihood based test. Hansen (1996) studied the large-sample behavior of these tests through a wide range of Monte Carlo simulations.

If we consider the counterpart of (2.2) in the time series case, an ARMA(1,1) process

$$y_t = \rho y_{t-1} + \epsilon_t - \lambda\epsilon_{t-1},\ t = 0, 1, 2, ..., T, \tag{2.7}$$

one would test $H_0 : \rho = \lambda$ for testing a white noise. For this case, Bera and Ra (1993) showed that theoretically the RS test is not feasible because under H_0 the nuisance parameter is not identified, and the information matrix

$$I(\theta) = T \begin{bmatrix} [(T-1)(1-\rho^2)]^{-1} & -[(T-1)(1-\rho\lambda)]^{-1} & \mathbf{0} \\ -[(T-1)(1-\rho\lambda)]^{-1} & [(T-1)(1-\lambda^2)]^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{2\sigma^4} \end{bmatrix} \tag{2.8}$$

is singular under $H_0 : \rho = \lambda$. Besides, it is block-diagonal with respect to σ^2. However, $I(\theta)$ in (2.6) is not block-diagonal even under the null, which is also different from the block-diagonal information matrix while testing $\rho = \lambda = 0$ in (2.1), as in Anselin, Bera, Florax, and Yoon (1996).

2.3 RS Test for Fixed λ

To implement the Davies procedure, I derive the RS test statistic assuming λ is fixed (given). I can rewrite the log-likelihood in (2.3) as a function of $\underline{\theta} = (\rho, \sigma^2)'$ given λ,

$$l(\underline{\theta}|\lambda) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 + \log|B^{-1}| + \log|I - \rho W| - \frac{1}{2\sigma^2}y'(I-\rho W)B^{-1\prime}B^{-1}(I-\rho W)y. \tag{2.9}$$

I can also derive the score functions for a given value of λ, and under the null the score functions are,

$$d_\rho(\lambda) = \frac{\partial l(\tilde{\theta}|\lambda)}{\partial\rho} = -tr[(I-\lambda W)^{-1}W] + \frac{1}{\sigma^2}y'(I-\lambda W)^{-1}Wy,$$

$$d_{\sigma^2}(\lambda) = \frac{\partial l(\tilde{\theta}|\lambda)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}y'(I-\lambda W)B^{-1\prime}B^{-1}(I-\lambda W)y.$$

The information matrix given fixed λ is,

$$\begin{aligned} I(\underline{\theta}|\lambda) &= \begin{bmatrix} J_\rho(\lambda) & J_{\rho\sigma^2}(\lambda) \\ J_{\sigma^2\rho}(\lambda) & J_{\sigma^2}(\lambda) \end{bmatrix} \\ &= \begin{bmatrix} tr[W'B^{-1\prime}B^{-1}W + (B^{-1}W)^2] & \sigma^{-2}tr(B^{-1}W) \\ \sigma^{-2}tr(B^{-1}W) & \frac{n}{2\sigma^4} \end{bmatrix}. \end{aligned}$$

The RS statistic for testing H_0 for fixed λ can be expressed as,

$$RS(\lambda) = \frac{\{\frac{1}{\hat{\sigma}^2}y'B^{-1}Wy - tr[B^{-1}W]\}^2}{tr\{W'B^{-1\prime}B^{-1}W + [B^{-1}W]^2\} - \frac{2}{n}\{tr[B^{-1}W]\}}. \tag{2.10}$$

When $\lambda = 0$, $B = I - \lambda W = I$ and hence $RS(\lambda)$ becomes

$$RS(0) = \frac{y'Wy/\hat{\sigma^2}}{tr[(W' + W)W]}, \tag{2.11}$$

which is the test statistic for testing no spatial dependence against an spatial AR alternative, as in Moran (1948) and Burridge (1980).

Again for comparison for the time series ARMA(1,1) case as in model (2.7), Bera and Ra (1993) derived the RS statistic as,

$$RS^T(\lambda) = T(1 - \lambda^2)\frac{[\sum_{t=1}^{T} y_t \sum_{s=1}^{t-1} \lambda^{s-1} y_{t-s}]^2}{(\sum_{t=1}^{T} y_t^2)^2}, \tag{2.12}$$

which under $\lambda = 0$, becomes

$$RS^T(0) = T\frac{(\sum_{t=2}^{T} y_t y_{t-1})^2}{(\sum_{t=1}^{T} y_t^2)^2}. \tag{2.13}$$

By comparing expressions in (2.10) and (2.11) with those in (2.12) and (2.13), we notice similar features for the spatial and time series cases. The test statistic in (2.10) is based on $y'(I - \lambda W)^{-1}Wy = y'(I + \lambda W + \lambda^2 WW + ...)Wy = y'Wy + y'(\lambda W + \lambda^2 WW + ...)Wy$, which consists of higher order interactions through the higher powers of the weight matrix. When $\lambda = 0$, those higher order interactions vanish. Similarly in the time series case, $RS^T(\lambda)$ in (2.12) contains interactions of y_t with "all" the lags $\{y_{t-1}, y_{t-2}, ..., y_2, y_1\}$, but when $\lambda = 0$, only one interaction term $\sum_{t=2}^{T} y_t y_{t-1}$ remains, as in (2.13).

To approach the testing problem, Davies (1987) proposed using a test based on the supremum of statistics and provided an upper bound of the significance probability. I can express $RS(\lambda)$ as

$$RS(\lambda) = S^2(\lambda), \tag{2.14}$$

so that $S(\lambda) = RS^{\frac{1}{2}}(\lambda)$. Under the regularity conditions given in Kelejian and Prucha (2001), $S(\lambda)$ has an asymptotic standard normal distribution under the null hypothesis. The two essential conditions can be stated as follows:

(i) W is a row-standardized weight matrix whose diagonal elements are zero. Besides, W is uniformly bounded in row and column sums in absolute value and $(I - \rho W)^{-1}$ is also uniformly bounded.

(ii) The disturbances ϵ_i, $i = 1, 2, ..., n$ are *iid* with zero mean, variance σ^2 and $E|\epsilon_i|^{4+\delta} < \infty$ for some $\delta > 0$.

These conditions are standard assumptions in the spatial regression. For a given value of λ, the test statistic $RS(\lambda)$ can be viewed as a Moran's I type (up to some scale) statistic. Therefore, I can apply the asymptotic result of Kelejian and Prucha (2001). Define

$$
\begin{aligned}
Y(\lambda) =& \frac{\partial S(\lambda)}{\partial \lambda} \\
=& \frac{\{-\frac{1}{\hat{\sigma}^2} y' B^{-1} W B^{-1} W y + tr[B^{-1} W B^{-1} W^2]\}}{H(\lambda)} - \{\frac{1}{\hat{\sigma}^2} y' B^{-1} W y - tr[B^{-1} W]\} \cdot J(\lambda),
\end{aligned}
$$

where

$$
H(\lambda) = \{tr\{W' B^{-1'} B^{-1} W - [B^{-1} W^2]\} - \frac{2}{n} \{tr[B^{-1} W]\}^2\}^{1/2}
$$

and

$$
\begin{aligned}
J(\lambda) =& \frac{1}{2} \{tr\{W' B^{-1'} W B^{-1'} B^{-1} W + W' B^{-1'} B^{-1} W B^{-1} W B^{-1} W B^{-1} W \\
& - 2[B^{-1} W] B^{-1} W B^{-1} - \frac{4}{n} tr[B^{-1} W] B^{-1} W B^{-1} W\}.
\end{aligned}
$$

Let $\eta(\lambda) \sim N(0, \phi(\lambda))$, where $\phi(\lambda) = Var(Y(\lambda)) - \{Cov[S(\lambda), Y(\lambda)]\}^2$. Davies (1987) showed that an upper bound for the p-values can be computed by

$$
Pr(sup_{\lambda \in \Lambda} RS(\lambda) > u) \leqslant Pr(\chi_1^2 > u) + \int_\Lambda \psi(\lambda) d\lambda, \tag{2.15}
$$

where Λ is an appropriate range for λ and

$$
\psi = E[\eta(\lambda)^2]^{1/2} \frac{exp(-u/2)}{(2\pi)^{1/2}}.
$$

Given that λ is the parameter of the MA process, it can be any real number without affecting stationarity. In the spatial literature, it is customary to take $\Lambda = (-1, 1)$.

Given the complex nature of $Y(\lambda)$, it is difficult to compute the integral in (2.15) explicitly in our case. Davies (1987) suggested to use the following approximation by replacing $E[\eta(\lambda)^2]^{1/2}$ by its convenient "sample" counterpart:

$$
Pr(sup_{\lambda \in \Lambda} RS(\lambda) > u) \leqslant Pr(\chi_1^2 > u) + V \frac{exp(-u/2)}{(2\pi)^{1/2}}, \tag{2.16}
$$

where

$$
\begin{aligned}
V =& \int_\Lambda |\frac{\partial RS^{1/2}(\lambda)}{\partial \lambda}| d\lambda \\
=& |RS^{1/2}(\lambda_1) - RS^{1/2}(\lambda_L)| + |RS^{1/2}(\lambda_2) - RS^{1/2}(\lambda_1)| \\
& + |RS^{1/2}(\lambda_3) - RS^{1/2}(\lambda_2)| + ... + |RS^{1/2}(\lambda_U) - RS^{1/2}(\lambda_m)|,
\end{aligned} \tag{2.17}
$$

where λ_L and λ_U are the lower and upper bounds for λ, and λ_i $(i = 1, 2, ..., m)$ denotes the i-th turning point of $\mathsf{RS}(\lambda)$. The main result is Davies' Theorem A.2. Davies (1987) emphasized that (2.17) is only an approximation, but one would expect it to be much better than $\Pr(\chi_1^2 > u)$. As in his normal case, in the chi-squared case that we have, the approximation V scans across a range of widely different hypothesis (by varying λ) and the value of $\mathsf{RS}^{1/2}(\lambda)$ might tend to be independent for $\lambda = \lambda_1, \lambda_2, ..., \lambda_m$. As Davies (1987) argued, in that case, we would expect the law of large numbers to apply so that V would be a good approximation for $E[\eta(\lambda)^2]^{1/2}$. Simulation results reported in Davies (1987, Section 5) supports his conjecture.

Here I should mention another approach to take account of the unidentified λ. Andrews and Ploberger (1994, 1996) derived a class of optimal tests that maximize a weighted average power criterion. In our spatial context, Andrews and Ploberger (AP) test statistic can be written as

$$\mathsf{AP} = (1+c)^{-1/2} \int exp\{\frac{1}{2}\frac{c}{1+c}\mathsf{RS}(\lambda)\}dG(\lambda),$$

where $G(\lambda)$ is a weight function which can be taken as a "prior" distribution of λ and the constant c is selected according to the close or distant alternatives one would like to direct the power. Choices of $G(\lambda)$ and c make the application of the test somewhat difficult to implement in practice, especially in our spatial context. Andrews and Ploberger (1994) applied their procedure to test unidentified points of structural change using their tabulated asymptotic critical values by taking $G(\lambda)$ to be an uniform distribution, with two extreme cases as $c \to 0$ and $c \to \infty$.

2.4 Monte Carlo Simulations

In this section I investigate the finite sample performances of the standard RS and Davies tests through a Monte Carlo experiment.

The model I consider is

$$y = \rho W y + \epsilon - \lambda W \epsilon, \tag{2.18}$$

where ϵ are generated from a vector of standard normal distribution. The weight matrix W is built corresponding to a regular square 7×7 grid, with contiguity defined by the rook criterion. Each experiment is replicated 500 times, and therefore the maximum standard errors of point estimates reported in the following tables are $\sqrt{(0.5)(0.5)/500} \simeq 0.0223$. The results are based on the theoretical size of 0.05, and the proportions of rejections (i.e. the proportion of times the computed p-values are less than 0.05) of Davies and standard RS tests for different combinations of ρ and λ are reported in Table 2.1.

In Table 2.1, the first entry in the parentheses corresponds to Davies test and the second one is for the standard RS test. The p-values of Davies test are computed using the approximation method in (2.16)-(2.17) and $[\lambda_L, \lambda_U] = [-0.95, 0.95]$ with step length of 0.01. Since our null hypothesis is $\rho = \lambda$, the diagonal elements (highlighted in green color) in the table are the estimated sizes of the tests, and off-diagonal elements represent the power. From the table, we observe that the estimated sizes are close to the nominal 5% level, although the Davies test has somewhat higher sizes. Both tests show monotonicity in the power, i.e, as we move away from the null hypothesis, the power (the off-diagonal entries) increases.

When either $\rho = 0$ or $\lambda = 0$, the standard RS is optimal. Since for $\lambda = 0$, the alternative is a pure spatial AR process, and when $\rho = 0$, the model under the alternative hypothesis involves only a spatial MA component. For both these cases, the form of the RS statistic is identical. Looking at the entries in the first row ($\rho = 0$) and the first column ($\lambda = 0$), as expected, we indeed observe that the RS test performs better than that of Davies, though only marginally.

The superiority of Davies test is revealed when ρ and λ are close but different. These entries are highlighted in bold. For example, when $\rho = 0.9$ and $\lambda = 0.8$, Davies test has power 0.428, while the RS test has power 0.332. When $\rho = 0.2$ and $\lambda = 0.1$, the powers of Davies and RS tests are 0.108 and 0.068, respectively. In particular, for lower close values of ρ and λ, the powers of the RS test are close to the nominal size of 0.05.

In Table 2.2 I represent the rejection probabilities of Davies and RS tests, where the p-values are computed using a finer approximation with step length of 0.005 and $\lambda \in [-0.99, 0.99]$. Now the sizes of Davies test are closer to the nominal 5% level, while the earlier superiority in power still remains. Therefore, my suggestion would be to use as finer approximation as possible.

Tables 2.3-2.6 further illustrate the patterns of power with smaller intervals of different combinations of ρ and λ. In these tables, the p-values of Davies test are computed with step length of 0.005 and $\lambda \in [-0.99, 0.99]$. We observe that as we move away from the null hypothesis, the discrepancy between the two tests first increases and then decreases. For example, in Table 2.3, when $\lambda = 0.2$, and $\rho = 0.25, 0.3, 0.4$, the power combinations of the Davies and RS tests are $(0.076, 0.048)$, $(0.120, 0.088)$, $(0.234, 0.222)$, respectively. Another interesting feature is that the better power performance of Davies test is not symmetric around the null hypothesis when λ and ρ have relatively high values, while we notice some symmetry when they have small or median values. For instance, when $\lambda = 0.2$, the pattern of the power performance of Davies test is similar when ρ is smaller or larger than 0.2, and we see similar feature in Table 2.4 when $\lambda = 0.5$. However, when $\lambda = 0.7$, the performance of Davies test is better when ρ is larger than 0.7 compared to the cases where ρ is smaller than 0.7, as we observe in Table 2.5. Similar pattern is seen in Table 2.6 for $\lambda = 0.8$. Thus from my simulation results I can conclude that the Davies approach yield a test that has better power properties very close to the null hypothesis of $H_0 : \rho = \lambda$ against the spatial ARMA alternative.

2.5 Conclusion

This chapter considers the problem of testing no spatial dependence against a spatial ARMA process. The conventional RS test is not feasible since the parameters are not identified under the null hypothesis, and hence the information matrix in this case is singular. This is one of the so-called *non-standard* test cases. My proposed procedure is based on the supremum of the standard RS test statistic, following Davies (1977, 1987). The simulation results show that the proposed test has desired size and higher power for the spatial ARMA alternative, especially when the two parameters corresponding to AR and MA processes are different but close to each other.

As suggested by one of the referees, I now mention some shortcomings of my approach to testing for particular types of spatial dependence. There is a debate about whether to base the spatial model selection on a specific-to-general testing [as in Florax, Folmer, and Rey (2003)], a general-to-specific approach [as suggested by Mur and Angulo (2009)], or a mix of the two [as explored in Elhorst (2014)]. LeSage and Pace (2009) object to such test procedures for particular forms of spatial dependence. LeSage and Pace (2009) also make the case for the spatial Durbin model because of concerns about the robustness of diagnostic tests to misspecifications of the spatial dependence.

Table 2.1: Rejection Probabilities

ρ/λ	0.0	0.1	0.2	0.3	0.4
0.0	(0.072, 0.042)	**(0.092, 0.084)**	(0.154, 0.176)	(0.270, 0.360)	(0.412, 0.526)
0.1	**(0.086, 0.064)**	(0.074, 0.044)	**(0.086, 0.080)**	**(0.154, 0.148)**	(0.234, 0.328)
0.2	(0.184, 0.202)	**(0.108, 0.068)**	(0.088, 0.048)	**(0.102, 0.074)**	(0.142, 0.166)
0.3	(0.286, 0.344)	(0.196, 0.238)	**(0.108, 0.088)**	(0.070, 0.048)	**(0.084, 0.074)**
0.4	(0.478, 0.604)	(0.364, 0.432)	**(0.248, 0.222)**	**(0.104, 0.102)**	(0.060, 0.048)
0.5	(0.712, 0.808)	(0.518, 0.592)	(0.410, 0.434)	(0.254, 0.282)	**(0.122, 0.094)**
0.6	(0.874, 0.932)	(0.754, 0.828)	(0.624, 0.692)	(0.454, 0.514)	**(0.276, 0.272)**
0.7	(0.964, 0.984)	(0.908, 0.956)	(0.856, 0.904)	(0.716, 0.780)	(0.570, 0.592)
0.8	(0.990, 1.000)	(0.984, 0.996)	(0.956, 0.976)	(0.924, 0.962)	(0.844, 0.880)
0.9	(1.000, 1.000)	(1.000, 1.000)	(0.998, 1.000)	(0.982, 0.998)	(0.974, 0.986)
ρ/λ	0.5	0.6	0.7	0.8	0.9
0.0	(0.508, 0.680)	(0.694, 0.866)	(0.792, 0.916)	(0.856, 0.958)	(0.936, 0.980)
0.1	(0.358, 0.504)	(0.468, 0.680)	(0.684, 0.854)	(0.794, 0.910)	(0.886, 0.962)
0.2	(0.264, 0.330)	(0.368, 0.526)	(0.484, 0.680)	(0.662, 0.858)	(0.778, 0.924)
0.3	(0.112, 0.176)	(0.230, 0.338)	(0.378, 0.574)	(0.482, 0.712)	(0.714, 0.890)
0.4	(0.074, 0.080)	(0.136, 0.202)	(0.264, 0.384)	(0.388, 0.618)	(0.546, 0.768)
0.5	(0.070, 0.056)	(0.080, 0.086)	(0.138, 0.186)	(0.216, 0.384)	(0.378, 0.612)
0.6	**(0.148, 0.118)**	(0.080, 0.042)	(0.082, 0.084)	(0.146, 0.220)	(0.264, 0.450)
0.7	**(0.350, 0.328)**	**(0.160, 0.114)**	(0.074, 0.036)	(0.086, 0.096)	(0.184, 0.266)
0.8	(0.710, 0.754)	**(0.506, 0.480)**	**(0.234, 0.176)**	(0.082, 0.044)	(0.090, 0.124)
0.9	(0.958, 0.968)	(0.912, 0.914)	**(0.772, 0.734)**	**(0.428, 0.332)**	(0.084, 0.034)

1. The first and second entries in each parenthesis correspond to Davies and standard RS tests, respectively.

2. The p-values of Davies test are computed for the parameter space, $\lambda \in [-0.99, 0.99]$ with step length 0.005.

Table 2.2: Rejection Probabilities with Finer Approximations

ρ/λ	0	0.1	0.2	0.3	0.4
0.0	(0.064, 0.042)	**(0.114, 0.084)**	(0.134, 0.176)	(0.214, 0.360)	(0.322, 0.526)
0.1	**(0.090, 0.064)**	(0.062, 0.044)	**(0.100, 0.080)**	**(0.150, 0.148)**	(0.212, 0.328)
0.2	(0.124, 0.202)	**(0.116, 0.068)**	(0.068, 0.048)	**(0.104, 0.074)**	(0.146, 0.166)
0.3	(0.276, 0.344)	(0.184, 0.238)	**(0.120, 0.088)**	(0.078, 0.048)	(0.074, 0.074)
0.4	(0.432, 0.604)	(0.338, 0.432)	**(0.234, 0.222)**	**(0.104, 0.102)**	(0.062, 0.048)
0.5	(0.676, 0.808)	(0.490, 0.592)	(0.364, 0.434)	(0.218, 0.282)	**(0.130, 0.094)**
0.6	(0.854, 0.932)	(0.752, 0.828)	(0.614, 0.692)	(0.398, 0.514)	**(0.308, 0.272)**
0.7	(0.946, 0.984)	(0.924, 0.956)	(0.800, 0.904)	(0.714, 0.780)	(0.556, 0.592)
0.8	(0.988, 1.000)	(0.984, 0.996)	(0.946, 0.976)	(0.916, 0.962)	(0.800, 0.880)
0.9	(1.000, 1.000)	(1.000, 1.000)	(0.998, 1.000)	(0.990, 0.998)	(0.974, 0.986)

ρ/λ	0.5	0.6	0.7	0.8	0.9
0.0	(0.464, 0.680)	(0.606, 0.866)	(0.726, 0.916)	(0.844, 0.958)	(0.924, 0.980)
0.1	(0.342, 0.504)	(0.486, 0.680)	(0.626, 0.854)	(0.742, 0.910)	(0.854, 0.962)
0.2	(0.214, 0.330)	(0.324, 0.526)	(0.444, 0.680)	(0.616, 0.858)	(0.748, 0.924)
0.3	(0.112, 0.176)	(0.228, 0.338)	(0.328, 0.574)	(0.452, 0.712)	(0.634, 0.890)
0.4	**(0.102, 0.080)**	(0.124, 0.202)	(0.188, 0.384)	(0.354, 0.618)	(0.522, 0.768)
0.5	(0.052, 0.056)	(0.060, 0.086)	(0.118, 0.186)	(0.218, 0.384)	(0.372, 0.612)
0.6	**(0.146, 0.118)**	(0.064, 0.042)	**(0.086, 0.084)**	(0.138, 0.220)	(0.246, 0.450)
0.7	**(0.352, 0.328)**	**(0.152, 0.114)**	(0.064, 0.036)	(0.082, 0.096)	(0.136, 0.266)
0.8	(0.658, 0.754)	(0.472, 0.480)	**(0.248, 0.176)**	(0.066, 0.044)	(0.090, 0.124)
0.9	(0.952, 0.968)	(0.856, 0.914)	(0.726, 0.734)	**(0.388, 0.332)**	(0.054, 0.034)

1. The first and second entries in each parenthesis correspond to Davies and standard RS tests, respectively.
2. The p-values of Davies test are computed for the parameter space, $\lambda \in [-0.99, 0.99]$ with step length 0.005.

Table 2.3: Power of the Tests, $\lambda = 0.2$

		$\lambda = 0.2$	
		Davies	RS
	0.0	0.134	0.176
	0.1	**0.100**	**0.080**
	0.15	**0.074**	**0.048**
	0.2	0.068	0.048
	0.25	**0.076**	**0.048**
	0.3	**0.120**	**0.088**
ρ	0.4	**0.234**	**0.222**
	0.5	0.364	0.434
	0.6	0.614	0.692
	0.7	0.800	0.904
	0.8	0.946	0.976
	0.9	0.998	1.000

Table 2.4: Power of the Tests, $\lambda = 0.5$

		$\lambda = 0.5$	
		Davies	RS
	0.0	0.464	0.680
	0.1	0.342	0.504
	0.2	0.214	0.330
	0.3	0.102	0.176
	0.4	**0.102**	**0.080**
	0.45	**0.068**	**0.048**
ρ	0.5	0.052	0.056
	0.55	**0.076**	**0.056**
	0.6	**0.146**	**0.118**
	0.65	**0.252**	**0.214**
	0.7	**0.352**	**0.328**
	0.8	0.658	0.754
	0.9	0.952	0.968

Table 2.5: Power of the Tests, $\lambda = 0.7$

		$\lambda = 0.7$	
		Davies	RS
	0.0	0.726	0.916
	0.1	0.626	0.854
	0.2	0.444	0.680
	0.3	0.328	0.574
	0.4	0.188	0.384
	0.5	0.118	0.186
ρ	0.6	**0.086**	**0.084**
	0.65	**0.066**	**0.052**
	0.7	0.064	0.036
	0.75	**0.110**	**0.070**
	0.8	**0.248**	**0.176**
	0.85	**0.432**	**0.394**
	0.9	0.726	0.734

Table 2.6: Power of the Tests, $\lambda = 0.8$

		$\lambda = 0.8$	
		Davies	RS
	0.0	0.844	0.958
	0.1	0.742	0.910
	0.2	0.616	0.858
	0.3	0.452	0.712
	0.4	0.354	0.618
	0.5	0.218	0.384
ρ	0.6	0.138	0.220
	0.7	0.082	0.096
	0.75	**0.072**	**0.058**
	0.8	0.066	0.044
	0.85	**0.136**	**0.102**
	0.9	**0.388**	**0.332**
	0.95	**0.754**	**0.704**

Chapter 3

Testing Spatial Regression Models Under Nonregular Conditions

3.1 Introduction

Problems of spatial dependence clouds the regression analysis, much more so than the presence of time-series dependence. In particular, modeling the spatial dependence structure in the disturbance term of regression has not received much attention in the literature. Even the limited suggested approaches are somewhat ad hoc and shrouded with ambiguity. Spatial autoregressive (SAR) model for the systemic part (sometimes along with a SAR process for the disturbance term) is the most common specifications in practice, while a few studies adopt a spatial moving average (SMA) model. As interpreted in Anselin (2003), the autoregressive (AR) process stands for the global spillover effect while the moving average (MA) process is more appropriate for capturing the local neighborhood effects. Ignoring the MA part in the error model can lead to serious misspecification. Recently, some attempts, for instances, Yao and Brockwell (2006) and Lam and Souza (2013), have been made to incorporate both into the error terms by considering a spatial autoregressive and moving average (SARMA) error model. However, estimation of such a seemingly simple spatial model is not a trivial task and most of the research questions remain unanswered with respect to this model, especially in terms of specification testing.

In this chapter I address the testing problems when the alternative model has spatially autocorrelated errors with a SARMA process. When trying to distinguish the two hypothesis: no spatial dependence versus dependence with a SARMA structure, problems arises and the standard testing procedures are not valid. There are two nonregularities in this testing problem. First, under the null hypothesis of no spatial autocorrelation in the disturbances, one underlying nuisance parameter is not identified. In such a case the information matrix is singular and all the likelihood based tests break down. Besides, the possible presence of spatial lag dependence in the systematic part may affect the performance of the test.

To deal with this twin-problems of nuisance parameters simultaneously, we construct the test that overcomes the singularity problem and the test still preserves the computational advantage of the Rao score (RS) test statistic in that I can conduct hypothesis testing based only on the ordinary least squared (OLS) residuals, avoiding the burden of estimating the full model. Besides, the suggested test is robust to the problem of possible presence of locally misspecified spatial lag dependence and keep the size as desired. Our proposed test follows the procedure of Davies (1977, 1987) based on the supremum of RS test statistic, and I adjust the RS statistic to take into account the presence of spatial lag dependence, suggested by Anselin, Bera, Florax, and Yoon (1996).

The chapter adds a missing piece in spatial Econometrics as I compare time-series literature and spatial analysis. Spatial Modeling is analogous to time series analysis from the perspective of incorporating autocorrelation. There are many studies in time series literature discussing modeling serial correlation using autoregressive (AR), moving average (MA), or ARMA representation. From the ideas of time series analysis, the spatial AR and MA process are also considered in spatial literature. However, unlike the large amount of studies on autoregressive and moving average (ARMA) process in time series literature, the study of spatial ARMA model is still limited. Therefore, the discussion in this chapter can provide a foundation for further model specification search in spatial Econometrics.

Although the problems arise from the attempt to test spatial dependence, it can be extended to non-spatial context. In this chapter I also address the similar problems in more general set up and provide the testing procedure. Therefore, my methodology is not limited to spatial analysis, but can be implement in general hypothesis testing.

The SARMA model to be tested in this chapter can have a broad application in economic analysis. Ever since the attention to incorporate spatial dependence arose from econometric perspective, it was quickly applied to a wide range of empirical economic studies, not only in real estates, regional, and urban economics, where the location and spatial interaction play a crucial role, but also in public economics, agricultural and environmental economics, industrial organization, and social interaction and networking. For example, Case, Rosen and Hines (1993) analyzed the U.S. state expenditure patterns with competition among local governments. Kim, Phipps, and Anselin (2003) measures the benefits of air quality for the Seoul metropolitan area while including spatial interaction in the housing market. There are still a variety of applications, such as potential spillovers from public infrastructure investments (Holtz-Eakin, 1994), cross-border lottery sales revenue (Garret and Marsh, 2002), externalities across regions in long-run growth (Rey and Montouri, 1999; Egger and Pfaffermayr, 2006; Fingleton and Lopez-Bazo, 2006), just to mention a few. More studies that adopts spatial models are undergoing and the methodology in this chapter provides a way to formally test the model that incorporates different sources of spatial dependence.

The plan of the rest of the chapter is as follows. Section 3.2 reviews the literature of spatial modeling. Section 3.3 formulates the testing problem. Section 3.4 provides the general approach to the testing problem. In section 3.5 I derive the test procedure following the approach in section 3.4. Section 3.6 illustrate the empirical application of the model specification tests through a variety of data sets. Section 3.7 presents the simulations studies of the tests. Section 3.8 concludes the chapter.

3.2 Review of Spatial Models

To explicitly specify spatial effects, most studies include spatial lag dependence in a regression model. This is similar to the inclusion of the lagged dependent variable in time-series context. In spatial econometrics, this is referred to as a spatial autoregressive (SAR) model (see Anselin, 1988, p.35), which is formally written as

$$y = \rho W y + X\beta + \epsilon, \tag{3.1}$$

where y is an $N \times 1$ vector of observations of dependent variable, W is the spatial weight matrix, X is an $N \times K$ observation matrix of explanatory variables, $\epsilon \sim \text{IIDN}(0, I\sigma^2)$, and ρ is the parameter for spatial lag dependence. The presence of ρ can be interpreted as a *direct* contagion or spatial interaction, i.e., the extent of spatial spillovers, copy-catting or diffusion.

Instead of *direct* spatial effects, there may be some *indirect* or *unobserved* dependence among economic agents. Therefore, there are attempts to incorporate spatial dependence in the unobserved disturbances. The most commonly specification includes a spatial autoregressive process in the disturbances, leading to the spatial error model (SEM):

$$\begin{aligned} y &= X\beta + u \\ u &= \tau W u + \epsilon, \end{aligned} \tag{3.2}$$

where τ is the spatial autoregressive coefficient for the error term. Following the discussion in Anselin (2003), the parameter ρ in (3.1) can be interpreted as a *global* spatial effect. However, the AR specification may be inappropriate when the interest lies in mechanisms of *local* spillovers. An alternative to the AR specification (3.2) suggested by Cliff and Ord (1981) and Haining (1978, 1990) is to specify a spatial moving average (SMA) error process that can take account of local spatial effect:

$$u = \epsilon - \lambda W \epsilon, \tag{3.3}$$

where λ is the spatial moving average coefficient. Hepple (2003) constructed the Bayesian and maximum likelihood estimator for this specification. Fingleton (2008a) proposed a GMM estimator, an extension of Kelejian and Prucha (1999) to estimate the SMA model. Following the idea of Arnold and Wied (2010) in SAR model, Baltagi and Liu (2011) further suggest an improved GMM estimator. By comparing the variance-covariance matrices of the error structure, the difference of the two specification lies in the range of spatial effects of the unobserved shocks.

In addition to specify either AR or MA process in the disturbances, Yao and Brockwell (2006) used the spatial error structure with spatial autoregressive and moving average (ARMA) error and derived the maximum likelihood estimators. Lam and Souza (2013) propose regularization method of spatial panel data using adaptive LASSO also with spatial ARMA error specification. The study of this kind of specification is still evolving and the related empirical studies are also limited. The most commonly used model in the empirical studies is to combine the spatial lag and spatial error dependence, i.e., the SAR-SEM specification:

$$\begin{aligned} y &= \rho W y + X\beta + u \\ u &= \tau W u + \epsilon. \end{aligned} \tag{3.4}$$

This is in contrast to a wide applications of ARMA process in time series literature. Analogous to the Box-Jenkins approach in time series analysis, Anselin (1988, p.33-34) and Anselin and Bera (1998) suggest a spatial regression specification with spatial AR lag and spatial MA in disturbance (spatial ARMA model). The specification allows for a global direct spatial effect in dependent variable as well as an indirect local spatial effect. The model can be written as,

$$\begin{aligned} y &= \rho W y + X\beta + u \\ u &= \epsilon - \lambda W \epsilon. \end{aligned} \tag{3.5}$$

Fingleton (2008b) propose a generalized method of moment (GMM) estimator of (3.5) based on two-stage-least-square estimator. Dogan and Taspinar (2013) introduce an one-step GMM estimator for spatial ARMA model and compare the properties of their suggested estimators with those in Fingleton (2008a, 2008b), both analytically and through Monte Carlo studies. Behrens, Ertur and Koch (2012) use spatial ARMA specification to study bilateral trade flows on a quantity-based structural gravity equation system and apply the estimation methodology to US-Canada trade dataset.

While including spatial lag dependent variable to incorporate direct spatial effect is a settled fact, how to model spatial interaction in the error process varies. Therefore, I start with a more general specification,

$$\begin{aligned} y &= \rho W y + X\beta + u \\ u &= \tau W u + \epsilon - \lambda W \epsilon. \end{aligned} \tag{3.6}$$

It is easily seen that model (3.6) encompasses models (3.1), (3.2), (3.4), and (3.5). As discussed above, previous literature related to spatial MA process focus on estimation methods and properties of various estimators. Studies of testing issues are, however, limited. In this paper I consider testing spatial models when the alternative is specified as (3.6). In particular, I test the null hypothesis of no spatial dependence against a spatial ARMA process in the disturbances. This model is an extension to Anselin et al. (1996) that I consider a more general spatial error process that allows for both global and local effects. I will compare the differences of the results of my test and those in Anselin et al. (1996) through empirical illustrations and Monte Carlo simulations in later sections.

3.3 The Testing Problem

I can rewrite (3.6) as

$$Y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}(I - \tau W)^{-1}(I - \lambda W)\epsilon. \tag{3.7}$$

Since I want to test for no spatial dependence in u in (3.6) against a spatial ARMA process, allowing for the possible presence of lag dependence (i.e. $\rho \neq 0$), the null hypothesis can be stated as $H_0 : \tau = \lambda = \tau_0$ (say); i.e., one restriction on the two parameters. Thus, under H_0 the nuisance parameter τ_0 is not identified, and hence, as will be shown shortly, the information matrix is singular. Therefore, the conventional likelihood-based tests cannot be derived.

The log-likelihood function is given by

$$\begin{aligned} l(\theta) =& -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\epsilon'\epsilon + \log|I - \rho W| + \log|I - \tau W| + \log|(I - \lambda W)^{-1}| \\ =& \text{Constant} - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(AY - X\beta)'B'C^{-1\prime}C^{-1}B(AY - X\beta) \\ &+ \log|A| + \log|B| + \log|C^{-1}|, \end{aligned} \tag{3.8}$$

where $\theta = (\beta', \sigma^2, \rho, \tau, \lambda)'$ and $A = I - \rho W$, $B = I - \tau W$, $C = I - \lambda W$, and thus $B = C$ under H_0.

The information matrix defined as $I(\theta) = E[-\frac{\partial^2 l}{\partial\theta\partial\theta'}]$, under H_0 can be expressed as,

$$\begin{aligned} I(\theta) &= \frac{1}{n\sigma^2}\begin{bmatrix} J_\beta & J_{\beta\sigma^2} & J_{\beta\rho} & J_{\beta\tau} & J_{\beta\lambda} \\ J_{\sigma^2\beta} & J_{\sigma^2} & J_{\sigma^2\rho} & J_{\sigma^2\tau} & J_{\sigma^2\lambda} \\ J_{\rho\beta} & J_{\rho\sigma^2} & J_\rho & J_{\rho\tau} & J_{\rho\lambda} \\ J_{\tau\beta} & J_{\tau\sigma^2} & J_{\tau\rho} & J_\tau & J_{\tau\lambda} \\ J_{\lambda\beta} & J_{\lambda\sigma^2} & J_{\lambda\rho} & J_{\lambda\tau} & J_\lambda \end{bmatrix}, \\ &= \frac{1}{n\sigma^2}\left[\begin{array}{c|c} J_{11} & J_{12} \\ \hline J_{21} & J_{22} \end{array}\right] \text{ (say)}, \end{aligned} \tag{3.9}$$

where the partition matrices are

$$J_{11} = \begin{bmatrix} X'X & 0 \\ 0 & \frac{n}{2\sigma^2} \end{bmatrix},$$

$$J_{12} = J_{21}' = \begin{bmatrix} X'WA^{-1}X\beta & 0 & 0 \\ 0 & tr(C^{-1}W) & -tr(C^{-1}W) \end{bmatrix},$$

$$J_{22} = \begin{bmatrix} \sigma^2[tr(A^{-1}WA^{-1}W) + tr(A^{-1}W'WA^{-1})] + \beta'XA^{-1\prime}W'WA^{-1}X\beta & \sigma^2[tr(W'C^{-1\prime}W) + tr(C^{-1}WW)] & -\sigma^2[tr(W'C^{-1\prime}W) + tr(C^{-1}WW)] \\ \sigma^2[tr(W'C^{-1\prime}W) + tr(C^{-1}WW)] & \sigma^2[tr(W'C^{-1\prime}C^{-1}W) + tr(C^{-1}WC^{-1}W)] & -\sigma^2[tr(W'C^{-1\prime}C^{-1}W) + tr(C^{-1}WC^{-1}W)] \\ -\sigma^2[tr(W'C^{-1\prime}W) + tr(C^{-1}WW)] & -\sigma^2[tr(W'C^{-1\prime}C^{-1}W) + tr(C^{-1}WC^{-1}W)] & \sigma^2[tr(W'C^{-1\prime}C^{-1}W) + tr(C^{-1}WC^{-1}W)] \end{bmatrix}.$$

The first part of the partition matrices J_{11} is as in the standard linear model $y = X\beta + u$, $u \sim N(0, I\sigma^2)$. The second part $J_{12} = J_{21}'$ can be viewed as the "interaction" among (β, σ^2) and spatial dependence parameters (ρ, τ, λ). We note that the information matrix is not block-diagonal with respect to σ^2, i.e., $J_{\sigma^2\tau} = \text{tr}(C^{-1}W)$ and $J_{\sigma^2\lambda} = -\text{tr}(C^{-1}W)$ even under $H_0 : \tau = \lambda = \tau_0$ unless $\tau_0 = 0$, (which is the case in Anselin et al. (1996)). Finally, we can see that the spatial dependence parameters combination, J_{22} is singular since the third column is the negative of the second column. Therefore, the information matrix $I(\theta)$ is also singular.

Due to the singularity of the information matrix, it is not possible to derive a standard RS test. Moreover, if we want to carry out test based on OLS estimation, one problem would be the presence of ρ. If we force ρ to be 0 and then test $H_0 : \tau = \lambda$, Anselin et al. (1996) showed that will lead to a size distortion in testing no spatial dependence in disturbances due to misspecification, i.e., possible presence of spatial lag dependence. Therefore, the goal of this chapter is to propose a test for $H_0 : \tau = \lambda$ in local presence of ρ but without estimating it. I apply Davies (1977, 1987) procedure to take account of the singularity of $I(\theta)$, and the adjusted RS test statistic following Anselin et al. (1996) to adjust for the local presence of ρ.

Based on Bera and Yoon (1993) and Anselin et al. (1996), the adjusted RS test statistic utilizes the information matrix under $\tau = \lambda$ and $\rho = 0$. Putting $\rho = 0$ in (3.9), we have $A = I$ and

$$I(\theta) = \frac{1}{n\sigma^2} \begin{bmatrix} X'X & 0 & X'WX\beta & 0 & 0 \\ 0 & \frac{n}{2\sigma^2} & 0 & \text{tr}(C^{-1}W) & -\text{tr}(C^{-1}W) \\ X'WX\beta & 0 & J_\rho & J_{\rho\tau} & J_{\rho\lambda} \\ 0 & \text{tr}(C^{-1}W) & J_{\tau\rho} & J_\tau & J_{\tau\lambda} \\ 0 & -\text{tr}(C^{-1}W) & J_{\lambda\rho} & J_{\lambda\tau} & J_\lambda \end{bmatrix}, \tag{3.10}$$

where

$$\begin{aligned} J_\rho &= \sigma^2[\text{tr}(WW) + \text{tr}(W'W)] + \beta'XW'WX\beta \\ J_{\rho\tau} &= J_{\tau\rho}^{J} = \sigma^2[\text{tr}(W'C^{-1}W + \text{tr}(C^{-1}WW)] \\ J_{\rho\lambda} &= J_{\lambda\rho}^{J} = -\sigma^2[\text{tr}(W'C^{-1}W + \text{tr}(C^{-1}WW)] \\ J_\tau &= \sigma^2[\text{tr}(W'C^{-1\prime}C^{-1}W) + \text{tr}(C^{-1}WC^{-1}W)] \\ J_{\tau\lambda} &= J_{\lambda\tau}^{J} = -\sigma^2[\text{tr}(W'C^{-1\prime}C^{-1}W) + \text{tr}(C^{-1}WC^{-1}W)] \\ J_\lambda &= \sigma^2[\text{tr}(W'C^{-1\prime}C^{-1}W) + \text{tr}(C^{-1}WC^{-1}W)]. \end{aligned}$$

3.4 Testing When a Nuisance Parameter is not Identified Under the Null Hypothesis with Local Misspecification

Let $f(y; \beta, \sigma^2, \rho, \tau, \lambda)$ be the probability density function of the random variable y where β, σ^2, ρ, τ and λ are the parameters. Suppose that under the null hypothesis $H_0 : \tau = \lambda$, we have

$$f(y; \beta, \sigma^2, \rho, \tau, \lambda) \equiv f(y; \beta, \sigma^2, \rho). \tag{3.11}$$

Then the parameters τ and λ are not identified under the null hypothesis. In this case, the information matrix is singular under H_0, and hence the standard likelihood based tests, such as Wald, Likelihood ratio, and RS tests cannot be derived. Some studies, such as Silvey (1959) and Poskitt and Tremayne (1981) suggested using a generalized inverse of the singular information matrix. However, this still leads to the standard RS test and is equivalent to setting $\tau = \lambda = 0$ and applying the test.

The problem of unidentified nuisance parameter is quite common in econometric literature, especially in time series cases. Davies (1977) was the first to consider this problem in the general context and he proposed a test procedure based on test statistics with normal distribution at any fixed value of the unidentified parameter. Davies (1987) extended his approach to the case where the test statistics follow central χ^2 distribution and also suggest an approximation method on the upper bound of the significance level.

Suppose that for any given value of λ, the RS statistic denoted by $RS(\lambda)$ has the standard asymptotic χ^2_1 distribution under the null hypothesis, Davies (1977, 1987) suggested the test based on a critical region of the form

$$M = sup_{\lambda \in \Lambda} RS(\lambda) > V$$

where V is a properly chosen constant and the range of λ is $\lambda = [\lambda_L, \lambda_U]$. Unlike $RS(\lambda)$, M does not have an asymptotic χ^2_1 distribution under the null hypothesis. If one uses the χ^2_1 critical value, the size of the test would be higher than desired. Davies (1987) provided an upper bound for $Pr[sup_{\lambda \in \Lambda} RS(\lambda) > V]$ which may be analytically calculated from the autocorrelation function of $RS(\lambda)$. Suppose we can express $RS(\lambda)$ as $RS(\lambda) = S^2(\lambda)$ where $S(\lambda)$ follow standard normal distribution for a given value of λ. Let $Y = \partial S(\lambda)/\partial \lambda$ and define $\phi(\lambda) = Var(Y(\lambda)) - \{Cov[S(\lambda), Y(\lambda)]\}^2$. Then Davies' asymptotic upper bound of the p-value is given by

$$Pr[sup_{\lambda \in \Lambda} RS(\lambda) > m] \leqslant Pr(\chi^2_1 > m) + \int_{\Lambda} \xi(\lambda) d\lambda, \tag{3.12}$$

where

$$\xi(\lambda) = E(\|\eta(\lambda)\|) \frac{e^{-m/2}}{\sqrt{2\pi}},$$

with $\|\eta(\lambda)\| = [\eta(\lambda)'\eta(\lambda)]^{1/2}$, and $\eta(\lambda) \sim N(0, \phi(\lambda))$. The second term on the right hand side of the inequality, $\int_{\Lambda} \xi(\lambda) d\lambda$, is the correction factor and it represents the expected number of upcrossings of the level m by the process of $RS(\lambda)$ over the range $\lambda \in \Lambda$. In many cases the explicit expression of the integral may be difficult to calculate. Following Sharpe (1978), Davies (1987) proposed an approximation of the correction factor by

$$V\frac{e^{-m/2}}{\sqrt{2\pi}},$$

where V is a measure of "total variation" in $RS^{1/2}(\lambda)$,

$$\begin{aligned}V=&\int_{\Lambda}|\frac{\partial RS^{1/2}(\lambda)}{\partial\lambda}|d\lambda\\=&|RS^{1/2}(\lambda_1)-RS^{1/2}(\lambda_L)|+|RS^{1/2}(\lambda_2)-RS^{1/2}(\lambda_1)|\\&+|RS^{1/2}(\lambda_3)-RS^{1/2}(\lambda_2)|+...+|RS^{1/2}(\lambda_U)-RS^{1/2}(\lambda_m)|,\end{aligned}$$

and $\lambda_1,...,\lambda_m$ are the turning points of $RS^{1/2}(\lambda)$.

The upper bound of the significance level suggested by Davies (1987) is then given by

$$\alpha=Pr(\chi_1^2>m)+V\frac{e^{-m/2}}{\sqrt{2\pi}}. \tag{3.13}$$

The first term of the right hand side comes from, $RS(\lambda)\xrightarrow{D}\chi_1^2$, and the second term, specially, V take care of the variation in $RS(\lambda)$ over possible values of λ.

The above test procedure exploits the fact that for any given value of λ, $RS(\lambda)$ has asymptotic central χ_1^2 distribution. However, this may not be the case under the local misspecification of the nuisance parameter ρ. Denoting $\gamma=(\beta',\sigma^2)'$, the log-likelihood function of the general statistical model can be expressed as $l(\gamma,\rho,\tau,\lambda)$ and in the above discussion, the question of interest is to test $H_0:\tau=\lambda$. Suppose one sets $\rho=\rho_0$ and tests $H_0:\tau=\lambda=\tau_0$(say) using the likelihood function $l_1(\gamma,\rho_0,\tau_0)$ and apply Davies procedure since τ_0 is not identified under the null hypothesis. The RS statistic used in the Davies test under $l_1(\gamma,\rho_0,\tau_0)$ will be denoted by $RS_\tau(\lambda)$. Let us further denote $\theta=(\gamma,\rho,\tau)'$ and $l(\theta|\lambda)$ the log-likelihood of θ for fixed λ, and also $\tilde{\theta}(\lambda)=(\tilde{\gamma}'(\lambda),\rho_0,\tau_0)'$ where $\tilde{\gamma}(\lambda)$ is the maximum likelihood estimator (MLE) of γ when $\tau=\tau_0$, $\rho=\rho_0$ and λ is given. The score vector and the information matrix can be defined, respectively, as

$$d(\theta|\lambda)=\frac{\partial l(\theta|\lambda)}{\partial\theta}=\begin{bmatrix}\frac{\partial l(\theta|\lambda)}{\partial\gamma}\\\frac{\partial l(\theta|\lambda)}{\partial\rho}\\\frac{\partial l(\theta|\lambda)}{\partial\tau}\end{bmatrix}$$

and

$$J(\theta|\lambda)=-E\left[\frac{1}{n}\frac{\partial^2 l(\theta|\lambda)}{\partial\theta\partial\theta'}\right]=\begin{bmatrix}J_\gamma(\lambda)&J_{\gamma\rho}(\lambda)&J_{\gamma\tau}(\lambda)\\J_{\rho\gamma}(\lambda)&J_\rho(\lambda)&J_{\rho\tau}(\lambda)\\J_{\tau\gamma}(\lambda)&J_{\tau\rho}(\lambda)&J_\tau(\lambda)\end{bmatrix}.$$

If $l_1(\gamma,\rho_0,\tau_0)$ is the true model, under $H_0:\tau=\tau_0$ we have

$$RS_\tau(\lambda)=\frac{1}{n}d_\tau(\tilde{\theta}|\lambda)'J_{\tau\cdot\gamma}^{-1}(\tilde{\theta}|\lambda)d_\tau(\tilde{\theta}|\lambda)\xrightarrow{D}\chi_1^2(0),$$

where $J_{\tau.\gamma}(\lambda) = J_{\tau}(\lambda) - J_{\tau\gamma}(\lambda)J_{\gamma}^{-1}(\lambda)J_{\gamma\tau}(\lambda)$. Therefore, under the null hypothesis $RS_{\tau}(\lambda)$ has asymptotic central χ_1^2 distribution for any given value of λ. Now suppose the true log-likelihood function is $l_2(\gamma, \tau_0, \rho)$, following Davidson and MacKinnon (1987) and Saikkonen (1989) that use a sequence of local values $\rho = \rho_0 + \delta/\sqrt{n}$, the asymptotic distribution of $RS_{\tau}(\lambda)$ for fixed λ can be obtained as

$$RS_{\tau}(\lambda) \xrightarrow{D} \chi_1^2(\omega_1), \tag{3.14}$$

where $\omega_1 = \delta' J_{\rho\tau.\gamma}(\lambda)J_{\tau.\gamma}^{-1}(\lambda)J_{\tau\rho.\gamma}(\lambda)\delta$, with $J_{\tau\rho.\gamma}(\lambda) = J_{\tau\rho}(\lambda) - J_{\tau\gamma}(\lambda)J_{\gamma}^{-1}(\lambda)J_{\gamma\rho}(\lambda)$, is the non-central parameter. When applying Davies procedure based on $RS_{\tau}(\lambda)$, the result of asymptotic central χ_1^2 distribution under the null hypothesis no longer holds with locally misspecified parameter ρ. Therefore, I expect that the test will be over-sized, even after Davies' correction.

To deal with the two problems: 1) the nonidentification of the parameter of spatial error dependence under the null hypothesis, and 2) the misspecified priori model for neglecting ρ, the spatial lag dependence, I suggest two corrections. First, I make the *non-central* χ^2 distribution (3.14) *central* using the net score function,

$$\begin{aligned} d^*_{\tau.\gamma}(\tilde{\theta}|\lambda) &= d_{\tau.\gamma}(\tilde{\theta}|\lambda) - E[d_{\tau.\gamma}(\tilde{\theta}|\lambda)|d_{\rho.\gamma}(\tilde{\theta}|\lambda)] \\ &= d_{\tau}(\tilde{\theta}|\lambda) - J_{\tau\rho.\gamma}(\tilde{\theta}|\lambda)J_{\rho.\gamma}^{-1}(\tilde{\theta}|\lambda)d_{\rho}(\tilde{\theta}|\lambda), \end{aligned}$$

and use $RS^*_{\tau}(\lambda)$ instead of $RS_{\tau}(\lambda)$, where

$$\begin{aligned} RS^*_{\tau}(\lambda) =& \frac{1}{n} d^*_{\tau.\gamma}(\tilde{\theta}|\lambda)'[J_{\tau.\gamma}(\tilde{\theta}|\lambda) \\ &- J_{\tau\rho.\gamma}(\tilde{\theta}|\lambda)J_{\rho.\gamma}^{-1}(\tilde{\theta}|\lambda)J_{\rho\tau.\gamma}(\tilde{\theta}|\lambda)]^{-1} d^*_{\tau.\gamma}(\tilde{\theta}|\lambda), \end{aligned} \tag{3.15}$$

is the adjusted RS statistic that adjusts for the local presence of ρ following Bera and Yoon (1993) principle. Under $\tau = \lambda$ and $\rho = \rho_0 + \delta/\sqrt{N}$, RS^*_{τ} has a *central* χ_1^2 distribution for any given value of λ.

The second correction I propose deals with the nonidentification problem. I apply Davies procedure to the adjusted RS statistic. I utilize the fact that $RS^*_{\tau}(\lambda)$ has a central χ^2 distribution for a given value of λ and apply Davies procedure to test $H_0 : \tau = \lambda$. The upper bound of the p-value is given by

$$Pr[sup_{\lambda \in \Lambda} RS^*_{\tau}(\lambda) > V] \leqslant Pr(\chi_1^2 > V) + \int_{\Lambda} \xi(\lambda) d\lambda \tag{3.16}$$

3.5 Derivation of the Test Statistic

Turning to the problem of testing no spatial dependence against a spatial ARMA process in the disturbance, where there is spatial lag dependence. To implement Davies procedure with the presence of nuisance parameter ρ, I first derive adjusted RS test statistic assuming λ is given. Defining $\underline{\theta} = (\beta', \sigma^2, \rho, \tau)$, the log-likelihood can be rewritten as

$$\begin{aligned} l(\underline{\theta}|\lambda) = &\text{Constant} - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}[(I-\rho W)Y - X\beta]'(I-\tau W)'C^{-1\prime}\cdot \\ &C^{-1}(I-\tau W)[(I-\rho W)Y - X\beta] + \log|I-\rho W| + \log|I-\tau W| + \log|C^{-1}|, \end{aligned} \tag{3.17}$$

And for a given value of λ, the score functions when $\rho = 0$ and $\tau = \lambda$ are

$$\begin{aligned} d_\beta(\lambda) &= \frac{1}{\sigma^2}X'(Y - X\beta) \\ d_{\sigma^2}(\lambda) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta)'(Y - X\beta) \\ d_\rho(\lambda) &= \frac{1}{\sigma^2}(Y - X\beta)'WY \\ d_\tau(\lambda) &= \frac{1}{\sigma^2}(Y - X\beta)'C^{-1}W(Y - X\beta) \end{aligned}$$

The information matrix given λ when $\rho = 0$ and $\tau = \lambda$, denoted as $I(\underline{\theta}|\lambda)|_{H_0}$, can be derived as

$$I(\theta)|_{H_0} = \frac{1}{n\sigma^2}\begin{bmatrix} X'X & 0 & X'WX\beta & 0 \\ 0 & \frac{n}{2\sigma^2} & 0 & tr(C^{-1}W) \\ X'WX\beta & 0 & \sigma^2[tr(WW) + tr(W'W)] + \beta'XW'WX\beta & \sigma^2[tr(W'C^{-1}W + tr(C^{-1}WW)] \\ 0 & tr(C^{-1}W) & \sigma^2[tr(W'C^{-1}W + tr(C^{-1}WW)] & \sigma^2[tr(W'C^{-1\prime}C^{-1}W) + tr(C^{-1}WC^{-1}W)] \end{bmatrix}, \tag{3.18}$$

The standard RS test statistic to test H_0 for fixed λ, assuming $\rho = 0$, can be expressed as

$$RS(\lambda) = \frac{\{\frac{1}{\hat{\sigma^2}}\hat{u}'C^{-1}W\hat{u} - tr[C^{-1}W]\}^2}{tr\{W'C^{-1\prime}C^{-1}W + [C^{-1}W]^2\} - \frac{2}{n}[tr(C^{-1}W)]^2}. \tag{3.19}$$

When $\lambda = 0$, $C = I - \lambda W = I$, and $RS(\lambda)$ becomes

$$RS(0) = \frac{\hat{u}'W\hat{u}/\hat{\sigma^2}}{tr[(W' + W)W]},$$

which is essentially Moran's I statistic.

The adjusted RS statistic for fixed λ adjusting for the presence of ρ can be derived as

$$RS^*_\tau(\lambda) = \frac{\{\hat{u}'C^{-1}W\hat{u}/\hat{\sigma}^2 - tr(C^{-1}W) - [tr(W'C^{-1\prime}W + C^{-1}WW)](nJ_{\hat{\rho\cdot\gamma}})^{-1}\hat{u}'Wy/\hat{\sigma}^2\}^2}{tr(W'C^{-1\prime}C^{-1}W + C^{-1}WC^{-1}W) - \frac{2}{n}[tr(C^{-1}W)]^2 - [tr(W'C^{-1\prime}W + C^{-1}WW)]^2(nJ_{\hat{\rho\cdot\gamma}})^{-1}} \tag{3.20}$$

where

$$(nJ_{\hat{\rho\cdot\gamma}})^{-1} = \hat{\sigma}^2 \cdot \{\hat{\beta}X'W'MWX\beta + \hat{\sigma}^2[tr((W' + W)W)]\}^{-1}$$

and $M = I - X(X'X)^{-1}X'$.

Similarly when $\lambda = 0$, $RS^*(\lambda)$ becomes

$$RS^*_\tau(0) = \frac{[\tilde{u}'W\tilde{u}/\tilde{\sigma}^2 - [tr((W'+W)W)](nJ_{\hat{\rho\cdot\gamma}})^{-1}\tilde{u}'Wy/\tilde{\sigma}^2]^2}{[tr((W'+W)W)]\{1 - [tr((W'+W)W)](nJ_{\hat{\rho\cdot\gamma}})^{-1}\}},$$

which is the test statistic for testing no spatial dependence against the alternative of spatial autoregressive process in the disturbances adjusted for spatial lag dependence, as in Anselin et al. (1996).

From the previous discussion, I constructed my test based on the supremum of the test statistic following Davies (1977, 1987). I can define

$$S(\lambda) = RS^{*1/2}_\tau(\lambda), \tag{3.21}$$

Under fixed λ and some assumptions for the validity of the asymptotic properties, $S(\lambda)$ converges to a standard normal distribution under the null hypothesis. I state the assumptions in the following.

Assumption 1.

(i) W is a row-standardized weight matrix whose diagonal elements are zero.

(ii) W is uniformly bounded in row and column sums in absolute value and $(I - \rho W)^{-1}$ and $(I - \tau W)^{-1}$ are also uniformly bounded.

Assumption 2. The disturbances ϵ_i, $i = 1, 2, ..., n$ are *iid* with zero mean, variance σ^2 and $E|e_i|^{4+\delta} < \infty$ for some $\delta > 0$.

Assumption 3.

(i) The elements of X are nonstochastic and uniformly bounded in n.

(ii) $lim_{n\to\infty}\frac{X'X}{n}$ exists and is nonsingular.

Assumption 1 is a standard assumption in spatial econometrics and boundness conditions for spatial weight matrix W. *Assumption 2* provides regularity assumptions for ϵ_i. When exogenous variables X are included in the model, it is convenient to assume that they are uniformly bounded as in *Assumption 3.* Both *Assumption 2* and *Assumption 3* are standard assumptions in linear regression analysis.

Following Davies (1987), the upper bound for the p-values can be computed by

$$Pr(sup_{\lambda\in\Lambda}RS^*_\tau(\lambda) > m) \leqslant Pr(\chi^2_1 > m) + \int_\Lambda \psi(\lambda)d\lambda, \tag{3.22}$$

where

$$\psi = E[\eta(\lambda)^2]^{1/2}\frac{\exp(-m/2)}{(2\pi)^{1/2}}.$$

Given the complex structure of $RS^*_\tau(\lambda)$, it is difficult to compute the integral in (3.22) explicitly in this case. I adopt the approximation method discussed in section 3.4, and the upperbound of the p-value is

$$\Pr(\sup_{\lambda\in\Lambda} RS^*(\lambda) > u) \leqslant \Pr(\chi^2_1 > u) + V\frac{\exp(-u/2)}{(2\pi)^{1/2}}, \tag{3.23}$$

where

$$\begin{aligned} V =& \int_\Lambda |\frac{\partial RS^{*1/2}(\lambda)}{\partial\lambda}|d\lambda \\ =& |RS^{*1/2}(\lambda_1) - RS^{*1/2}(\lambda_L)| + |RS^{*1/2}(\lambda_2) - RS^{*1/2}(\lambda_1)| \\ &+ |RS^{*1/2}(\lambda_3) - RS^{*1/2}(\lambda_2)| + ... + |RS^{*1/2}(\lambda_U) - RS^{*1/2}(\lambda_m)|, \end{aligned} \tag{3.24}$$

with λ_L and λ_U the lower and upper bound for λ, and λ_i $(i = 1, 2, ..., m)$ denotes the i-th turning point of $RS^*(\lambda)$.

3.6 Empirical Illustration

To compare the properties of different tests and gain more insight on their applications to regional science and urban economics, we implement the standard RS, adjusted RS and Davies tests to a variety of data sets, and Table 3.1 presents the results. For each data set, the test statistics and the corresponding p-values are reported. The Davies test procedure does not have the exact test statistic. Therefore, I report the maximum value of $RS^*_\tau(\lambda)$ and the p-values computed by the approximation method. I also estimated the full model (SAR-ARMA) with all three parameters of spatial dependence and include the estimated coefficients and the p-values in the table. These data sets are chosen specifically to highlight different types of spatial effects. The first regression is the relationship between crime and housing value and income in 1980 for 49 neighborhoods in Columbus, OH. Previous literature on model specification search for this data set (Anselin, 1988a; Anselin et all, 1996; Sen, Bera, and Kao, 2012) suggest that there is significant spatial lag dependence but no spatial error dependence. From Table 3.1, we see that standard RS test still show significance of spatial error dependence. This may suggest the over-rejection feature in the presence of spatial lag dependence. On the other hand, both adjusted RS and Davies tests are consistent with the conclusion of previous studies. The p-values are higher than 5% and hence we do not reject the null hypothesis of no spatial dependence in the disturbances. From the estimation results, we see the coefficient of spatial lag dependence (ρ) is significant, while the coefficients for both spatial error dependence (τ and λ) are not significant. The results are consistent with the conclusions from my test and adjusted RS test, as well as the findings in previous literature.

Table 3.1: Empirical Illustration

	Columbus Crime $N = 49$	Boston Housing $N = 506$	Netherland Investment $N = 40$
RS_τ	5.72	105.51	2.43
	(0.017)	(9.462×10^{-25})	(0.120)
RS^*_τ	0.08	43.17	0.14
	(0.777)	(5.014×10^{-11})	(0.708)
Davies - $maxRS^*_\tau(\lambda)$	3.76	89.41	7.88
	(0.197)	(9.324×10^{-20})	(0.019)
$\hat{\rho}$	0.43^{*}	0.35^{***}	0.35^{***}
	(0.062)	(7.449×10^{-8})	(0.009)
$\hat{\tau}$	-0.04	0.79^{***}	0.82^{***}
	(0.444)	(6.410×10^{-10})	(2.742×10^{-4})
$\hat{\lambda}$	0.02	-0.04	0.55^{**}
	(0.159)	(0.139)	(0.024)

1. p-values are in the parentheses.
2. The p-value of Davies test are computed for the parameter space, $\lambda \in [-0.99, 0.99]$ with step length 0.005.

Table 3.2: Netherland Investment Model Comparison

	SAR	SEM	SMA	SAR-SEM	SAR-ARMA
$\hat{\rho}$	0.35^{***}			0.11	0.35^{***}
	(0.002)			(0.417)	(0.009)
$\hat{\tau}$		0.46^{***}		0.59^{***}	0.82^{***}
		(0.005)		(0.009)	(2.742×10^{-4})
$\hat{\lambda}$			-0.13^{***}		0.55^{**}
			(0.005)		(0.024)
log-likelihood	-155.5	-157.3	-187.2	-144.8	-137.82

1. p-values are in the parentheses.

The second model estimates the demand for clean air using housing market data from the Boston Standard Metropolitan Statistical Area in 1970, originated by Harrison and Rubinfeld (1978). In their study the dependent variable in the hedonic equation is the median value of owner-occupied houses in each of the 506 census tracts, and there are 14 covariates. Pace and Gilley (1997) introduced the spatial feature into the model with the location of each tract in latitude and longitude out of the 1970 census to the data set. They suggested that there is strong error dependence and the estimation errors of the coefficients of the explanatory variables fell largely with the estimation of error dependence using a two-dimensional grid search. From Table 3.1 we see all the tests agree with the finding of significant spatial dependence in the disturbances. The estimation results show the significance of both spatial lag and error dependence. However, the coefficient of the spatial MA process (λ) is not significant, which suggests it might be more appropriate to consider the SAR-SEM specification.

The third model is the neoclassical multiregional investment model from Florax (1992), estimated using data of 40 COROP regions in the Netherlands. The study addressed the topic of knowledge impacts of universities on the investment in manufacturing industry. I use the linearized version (Florax, 1992, p. 201) that relates investment in buildings by the manufacturing sector to output, investment in equipment and distance to the core region, contagious knowledge diffusion and hierarchical knowledge diffusion. Incorporating spatial dependence is especially relevant in the context of the firm's location choice, which is directly related to the investment in buildings. The results from Table 3.1 show that while the Davies test concludes there is significant spatial error dependence, the RS tests fail to reject the null hypothesis of no spatial dependence at 5% significance level. Therefore, instead of concluding there is no spatial autocorrelation in the error term, one may consider a spatial ARMA process that might better capture the feature of the data set. Consistent with the conclusion from my test, the estimation results of the full model show all three parameters of spatial dependence are highly significant.

I further estimated several competing models and report the results in Table 3.2. When only one of the parameter of spatial dependence is included in the model, the estimated coefficient is significant, showing a strong spatial dependence in the data. However, compared to models with both spatial lag and error dependence, the values of log-likelihood functions are lower if we only consider one source of spatial dependence. The estimation of SAR-SEM model shows a significant spatial error dependence but insignificant spatial lag dependence, while the estimation of my full model suggests the significance of all the parameters. Ignoring the possible spatial MA dependence may lead to inefficient estimators, and hence the insignificant result as we see in the SAR-SEM specification. Moreover, including all three parameters of spatial dependence increases the value of log-likelihood function. From the analysis I suggest to include both spatial lag and error dependence, with the specification of a spatial ARMA process in the disturbance.

The application shows the usefulness of my suggested test. While the RS tests fail to detect the spatial error dependence, my test still finds it significant. The estimation results confirm the conclusion from my test and suggest that the model incorporating spatial ARMA process in the error term is better from the comparison of values of log-likelihood. The unobserved shocks in this model may contain both global and local spatial effects.

3.7 Monte Carlo Simulations

To compare the performance for the various tests, I conduct Monte Carlo experiments for different combinations of the spatial dependence parameters. The model under the alternative is

$$\begin{aligned} Y &= \rho W Y + X\beta + u \\ u &= \tau W u + \epsilon - \lambda W \epsilon. \end{aligned} \tag{3.25}$$

I generate ϵ from a vector of standard normal distribution. The sample size $N = 49$, and the weight matrix W is built corresponding to a regular square 7×7 grid, with contiguity defined by the rook criterion. Each experiment is replicated 1,000 times, and therefore the maximum standard errors of point estimates reported in the following tables are $\sqrt{(0.5)(0.5)/1000} \simeq 0.0111$. The results are based on the theoretical size of 0.05, and the proportions of rejections (i.e. the proportion of times the computed p-values are less than 0.05) for standard RS, adjusted RS and Davies tests are calculated and reported in the following tables for different values of spatial autocorrelation parameters. For Davies test I report two p-value results. The first uses the approximation method in (3.23) and (3.24) with step length of 0.005 and $\lambda \in [-0.99, 0.99]$. The second is obtained from the bootstrap critical value with resampling 200 times. With the bootstrap method, I utilize the resampling distribution of the supremum of $RS^*_\tau(\lambda)$, and therefore the correction factor is not needed.

Table 3.3 presents the size of various tests under different values of ρ, the spatial lag dependence. When there is no misspecification ($\rho = 0$), all tests have empirical size close to the nominal size 5%. With locally misspecified spatial lag dependence, the standard RS rejects more often than desired, and as the value of ρ gets higher, the over-rejection is more serious. On the other hand, the size of both adjusted RS and Davies test are close to the chosen size under different values of ρ. we also note that the simulation results are very similar in Davies tests between the approximation method and the bootstrap method.

Table 3.3: Size of the tests

ρ	RS_τ	RS^*_τ	$Davies^a$	$Davies^b$
0.0	0.056	0.050	0.052	0.042
0.1	0.068	0.054	0.044	0.066
0.2	0.098	0.066	0.058	0.048

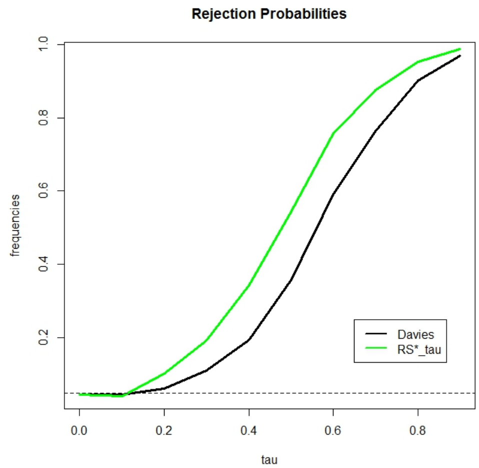
1. The 5% significance level is used.
2. $Davies^a$ is calculated using approximation with step length 0.005.
3. $Davies^b$ uses bootstrap critical value with resampling 200 times.

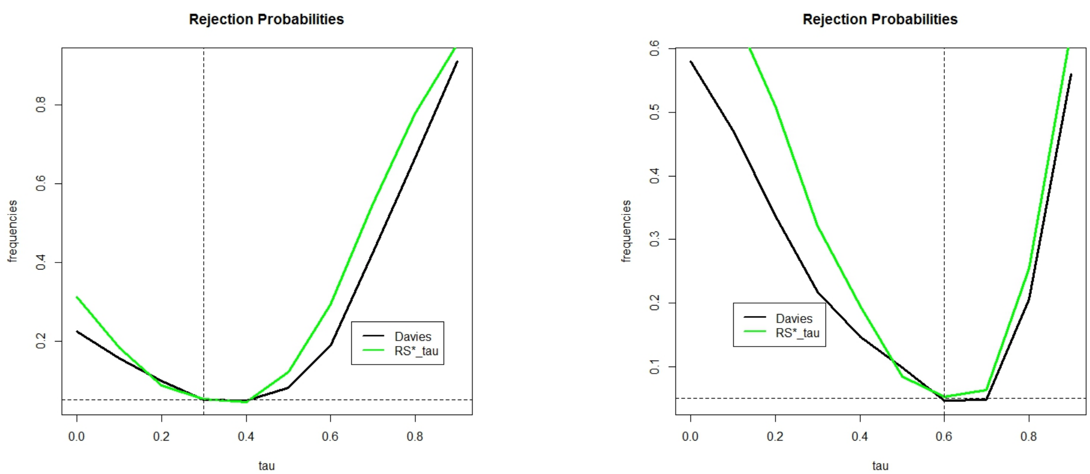
I focus my discussion on the power comparison between adjusted RS and Davies test. Figure 3.1 plots the empirical power functions of the two tests for different values of τ when both ρ and $\lambda = 0$. Theoretically RS test should be locally optimal when $\lambda = 0$. The simulation result, as expected, shows that the $RS^*_\tau(\lambda)$ has higher power for all values of τ.

Figures 3.2(a) and 3.2(b) are the empirical power curves for different combinations of τ and λ. The case under $H_0 : \tau = \lambda$ is drawn as the vertical dashed line in the figures. We observe that when both τ and λ are different from zero and when their values are close to each other, Davies test has higher power compared to adjusted RS test. These results are similar under different values of ρ, which I present in Figure 3.3.

To see more clearly the power gain of Davies test, I plot the power curves in smaller intervals around the null hypothesis. From Figure 3.4 we can see the higher power performance of Davies

Figure 3.1: Rejection Probabilities: $\rho = 0$, ˘ $= \mathbf{0}$





(a) $\rho = 0.1$, ˘ $= \mathbf{0.6}$

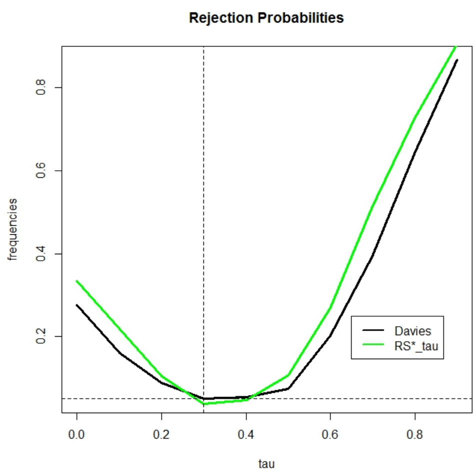
(b) $\rho = 0$, ˘ $= \mathbf{0.6}$

Figure 3.2: Rejection Probabilities, $\lambda = 0.6$

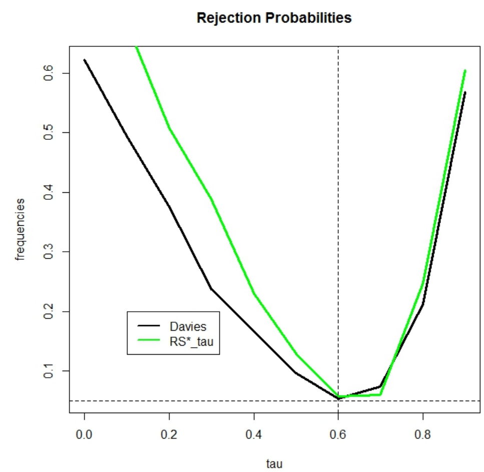
test under the alternative local to the null hypothesis. Similarly for other values of ρ and λ in Figure 3.5.

When the value of ρ becomes higher, the power gain of Davies test becomes more global, as we can see in Figures 3.6(a) and 3.6(b). Under the alternative of a high value of τ, which is far away from the null hypothesis, the power performance is still better than adjusted RS test. In addition, the superiority of Davies test in terms of power performance is asymmetric. When $\rho > \lambda$, Davies test has higher power, but not in the cases $\rho < \lambda$. Therefore, we observe better power performance of Davies test to the right of null hypothesis in the figures.

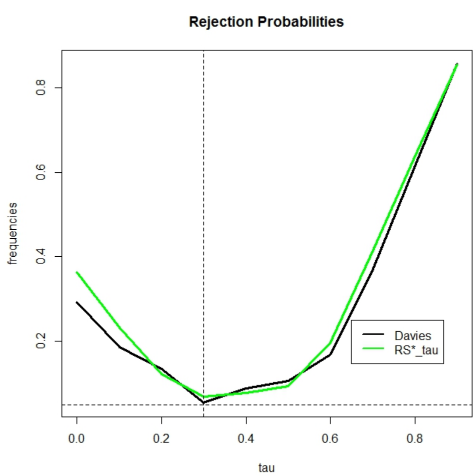
Table 3.4-3.6 reports the detailed numbers of rejection probabilities from the simulation results. Table 3.4 reports the rejection probabilities for different combinations of τ and λ where there is no spatial lag dependence, i.e., $\rho = 0$. In this case there is no local misspecification and I expect the



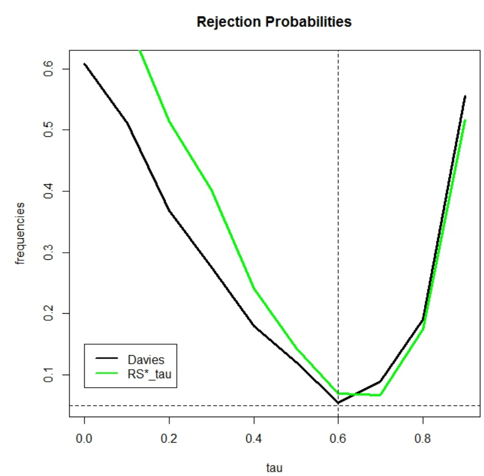
(a) $\rho = 0.1$, ˘ $= \mathbf{0.3}$



(b) $\rho = 0.1$, ˘ $= \mathbf{0.6}$



(c) $\rho = 0.2$, ˘ $= \mathbf{0.3}$



(d) $\rho = 0.2$, ˘ $= \mathbf{0.6}$

Figure 3.3: Rejection Probabilities

Figure 3.4: Rejection Probabilities: $\rho = 0.1$, ˘ $=$ **0.3**

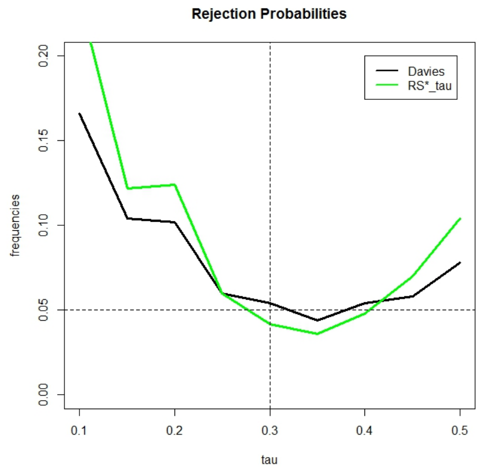
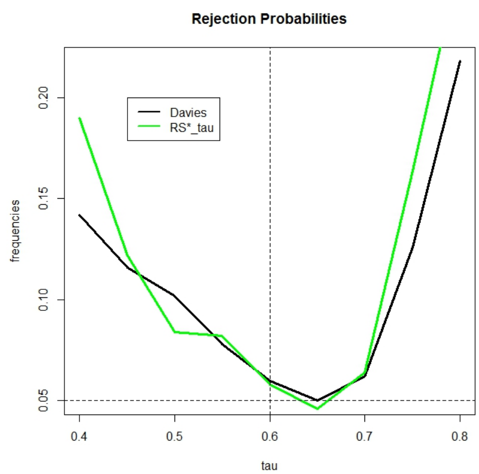
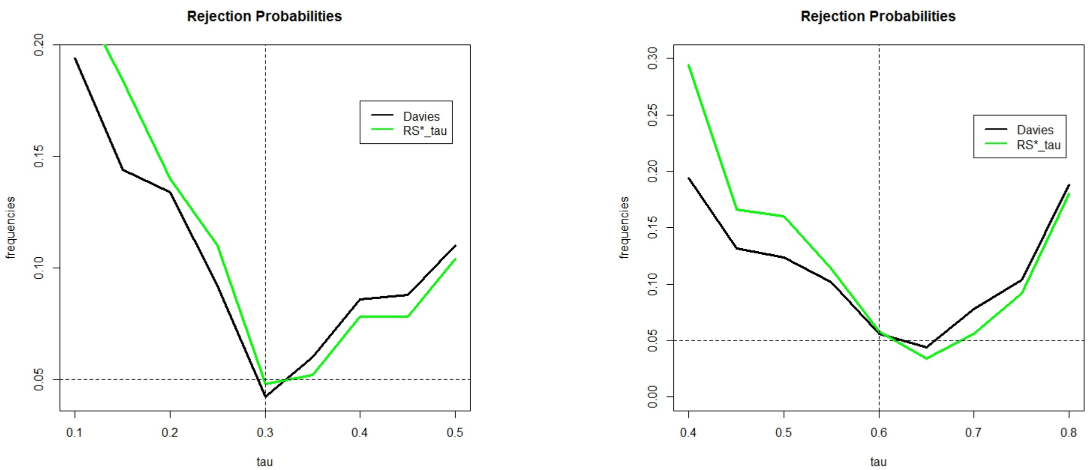


Figure 3.5: Rejection Probabilities: $\rho = 0$, ˘ $=$ **0.6**



standard RS test to have desired size and is locally most powerful. From Table 3.4 we see that under $H_0 : \tau = \lambda$, all three tests have rejection proportions close to the nominal size of 5%. Besides, all the tests show monotonicity in the power, i.e, as we move away from the null hypothesis, the power increases. The standard RS test has highest power for all the different values of τ and λ. Davies test has the lowest power compared to the standard and adjusted RS tests, but barely marginally especially with high values of τ or λ.

In Tables 3.5 and 3.6 I present the rejection probabilities when there is local misspecifications of ρ. Table 3.5 shows the rejection rates when $\rho = 0.1$ and Table 3.6 are the testing results when $\rho = 0.2$. With the presence of nuisance parameter ρ, the estimated size of standard RS test is higher than the nominal size of 5%. On the other hand, both adjusted RS and Davies tests have estimated sizes close to 5% for different values of ρ. In particular, while the rejection rates is a little higher than 5% when $\rho = 0.2$ for RS^*_τ, the size of Davies test is still very close to 5%. From



(a) $\rho = 0.2$, ˘ $= \mathbf{0.3}$

(b) $\rho = 0.2$, ˘ $= \mathbf{0.6}$

Figure 3.6: Rejection Probabilities, $\rho = 0.2$

Table 3.5, we see that when $\rho = 0.2$ and $\tau = \lambda = 0.8$, the rejection rates of Davies and adjusted RS tests are 5.5% and 7.6%, respectively.

We also observe from Tables 3.5 and 3.6 that the standard RS test always has the highest power compared to the other two tests when the values of τ and λ are far apart. However, this may be due to the over rejection feature with ignorance of the presence of nuisance parameter ρ. When $\tau < \lambda$, the standard RS test does not perform very well and the power is the lowest. From Tables 3.5 and 3.6, the feature becomes more apparent with higher value of ρ. For example, when $\rho = 0.2$, $\tau = 0.4$ and $\lambda = 0.6$, the rejection probabilities for Davies, standard RS, and adjusted RS tests are 17.9%, 5.3%, and 24.1%, respectively.

From the observation, I can conclude that Davies test has better power performance if we consider the alternatives around the neighborhood of the null hypothesis. In particular, with the presence of nuisance parameter ρ, the two RS tests lose the property of monotonicity in powers for some cases. For example, when $\tau = 0.8$ and $\lambda = 0.7$, the rejection probability of adjusted RS test is 6.2%, and when $\tau = 0.6$ and $\lambda = 0.7$, the estimated power of standard RS test is 2.8%, both are lower than their estimated sizes under $H_0 : \tau = \lambda = 0.7$. This feature is not seen in Davies test where we see that the power always increases when we move away from the null hypothesis.

3.8 Conclusion

In this chapter I propose a test procedure of model specification test for no spatial dependence against the alternative of a spatial ARMA process in the disturbances, with possible presence of spatial lag dependence. The derived test solves two problems simultaneously. First, the problem of nonidentification of the nuisance parameter and hence the singularity of information matrix under the null hypothesis. Second, it is robust to the possible impact of the presence of spatial lag dependence parameter ρ. Simulation studies show that my test has desired size, especially when there is spatial lag dependence. Compared to the adjusted RS test in Anselin et al. (1996), my test has higher power for alternatives local to the null hypothesis of spatial error independence. The power gain of my test is not symmetric around the null hypothesis, and is better on the right hand side, i.e., higher values of τ and λ. The suggested testing method is also applied to several spatial data sets, and I found that while there is possible spatial ARMA process in the error terms with the two similar values of spatial dependence parameters, our test is significant but the standard and adjusted RS tests fail to reject the null hypothesis of no spatial dependence.

Table 3.4: Rejection Probabilities: $\rho = 0$

	$\lambda = 0.0$			$\lambda = 0.1$			$\lambda = 0.2$			$\lambda = 0.3$			$\lambda = 0.4$		
τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ
0.0	0.044	0.043	0.045	0.082	0.104	0.091	0.171	0.230	0.195	0.225	0.358	0.312	0.356	0.550	0.490
0.1	0.046	0.040	0.040	0.054	0.051	0.058	0.104	0.102	0.111	0.156	0.206	0.183	0.247	0.402	0.332
0.2	0.061	0.115	0.102	0.058	0.056	0.061	0.053	0.053	0.049	0.098	0.103	0.087	0.142	0.204	0.180
0.3	0.110	0.229	0.193	0.059	0.123	0.113	0.056	0.054	0.057	0.050	0.054	0.052	0.087	0.109	0.104
0.4	0.195	0.400	0.344	0.118	0.232	0.220	0.064	0.124	0.119	0.048	0.044	0.045	0.048	0.041	0.052
0.5	0.358	0.618	0.543	0.234	0.434	0.373	0.155	0.299	0.262	0.082	0.132	0.121	0.054	0.057	0.051
0.6	0.591	0.823	0.757	0.457	0.690	0.619	0.279	0.482	0.437	0.188	0.326	0.292	0.108	0.184	0.171
0.7	0.762	0.935	0.874	0.636	0.871	0.787	0.532	0.723	0.667	0.423	0.607	0.547	0.258	0.405	0.356
0.8	0.902	0.981	0.952	0.832	0.952	0.910	0.778	0.936	0.894	0.667	0.841	0.779	0.535	0.728	0.666
0.9	0.968	0.998	0.987	0.968	0.996	0.989	0.937	0.984	0.966	0.910	0.970	0.955	0.822	0.927	0.868
	$\lambda = 0.5$			$\lambda = 0.6$			$\lambda = 0.7$			$\lambda = 0.8$			$\lambda = 0.9$		
τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ
0.0	0.461	0.694	0.616	0.580	0.823	0.744	0.692	0.915	0.839	0.746	0.955	0.905	0.824	0.974	0.944
0.1	0.335	0.545	0.466	0.471	0.725	0.658	0.574	0.826	0.757	0.666	0.907	0.837	0.786	0.961	0.923
0.2	0.238	0.382	0.326	0.337	0.565	0.510	0.480	0.733	0.655	0.562	0.842	0.757	0.726	0.932	0.885
0.3	0.133	0.193	0.180	0.218	0.365	0.322	0.304	0.551	0.493	0.467	0.735	0.656	0.584	0.844	0.755
0.4	0.094	0.102	0.102	0.147	0.219	0.195	0.246	0.385	0.341	0.317	0.582	0.495	0.466	0.733	0.652
0.5	0.045	0.048	0.050	0.098	0.104	0.084	0.149	0.218	0.210	0.228	0.421	0.353	0.325	0.598	0.535
0.6	0.049	0.068	0.058	0.046	0.054	0.052	0.081	0.116	0.111	0.148	0.260	0.230	0.219	0.431	0.384
0.7	0.129	0.195	0.174	0.048	0.061	0.063	0.057	0.056	0.050	0.080	0.132	0.118	0.144	0.297	0.265
0.8	0.362	0.527	0.448	0.206	0.272	0.254	0.068	0.102	0.093	0.055	0.054	0.053	0.097	0.157	0.164
0.9	0.756	0.875	0.823	0.560	0.692	0.628	0.334	0.430	0.390	0.122	0.162	0.135	0.043	0.050	0.048

1. The 5% significance level is used.
2. The p-values of Davies test are computed for the parameter space, $\lambda \in [-0.99, 0.99]$ with step length 0.005.

Table 3.5: Rejection Probabilities: $\rho = 0.1$

	$\lambda = 0.0$			$\lambda = 0.1$			$\lambda = 0.2$			$\lambda = 0.3$			$\lambda = 0.4$		
τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ
0.0	0.053	0.067	0.058	0.083	0.051	0.123	0.175	0.098	0.212	0.276	0.217	0.335	0.357	0.349	0.510
0.1	0.055	0.086	0.049	0.038	0.038	0.057	0.090	0.050	0.111	0.161	0.091	0.219	0.232	0.203	0.352
0.2	0.075	0.223	0.103	0.044	0.089	0.048	0.037	0.049	0.058	0.088	0.054	0.105	0.157	0.106	0.230
0.3	0.107	0.355	0.175	0.060	0.199	0.092	0.051	0.110	0.040	0.050	0.063	0.039	0.079	0.053	0.116
0.4	0.200	0.567	0.299	0.131	0.389	0.188	0.070	0.222	0.102	0.055	0.102	0.048	0.054	0.067	0.046
0.5	0.317	0.717	0.465	0.210	0.595	0.331	0.154	0.434	0.221	0.075	0.238	0.106	0.067	0.111	0.060
0.6	0.537	0.889	0.693	0.393	0.770	0.548	0.288	0.643	0.388	0.202	0.490	0.269	0.114	0.299	0.147
0.7	0.707	0.962	0.815	0.660	0.916	0.758	0.500	0.859	0.635	0.395	0.758	0.514	0.249	0.556	0.319
0.8	0.860	0.992	0.917	0.830	0.979	0.895	0.761	0.956	0.834	0.645	0.911	0.729	0.522	0.827	0.615
0.9	0.964	0.997	0.983	0.934	0.995	0.965	0.910	0.993	0.937	0.867	0.986	0.906	0.830	0.976	0.882
	$\lambda = 0.5$			$\lambda = 0.6$			$\lambda = 0.7$			$\lambda = 0.8$			$\lambda = 0.9$		
τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ
0.0	0.510	0.536	0.658	0.622	0.678	0.775	0.730	0.805	0.859	0.807	0.888	0.927	0.877	0.945	0.969
0.1	0.386	0.382	0.536	0.493	0.525	0.682	0.638	0.721	0.813	0.759	0.829	0.890	0.815	0.896	0.930
0.2	0.258	0.219	0.363	0.376	0.359	0.508	0.529	0.560	0.722	0.662	0.743	0.836	0.740	0.828	0.887
0.3	0.159	0.096	0.213	0.237	0.213	0.388	0.355	0.372	0.540	0.496	0.556	0.700	0.650	0.709	0.832
0.4	0.097	0.045	0.120	0.167	0.119	0.230	0.248	0.225	0.383	0.379	0.401	0.555	0.521	0.571	0.713
0.5	0.056	0.068	0.058	0.097	0.057	0.129	0.154	0.118	0.232	0.259	0.237	0.397	0.423	0.429	0.605
0.6	0.062	0.134	0.066	0.053	0.055	0.058	0.085	0.060	0.129	0.150	0.128	0.260	0.298	0.293	0.470
0.7	0.113	0.323	0.146	0.074	0.154	0.060	0.049	0.039	0.050	0.088	0.055	0.142	0.200	0.156	0.302
0.8	0.351	0.615	0.415	0.212	0.437	0.248	0.091	0.187	0.087	0.055	0.075	0.066	0.103	0.072	0.166
0.9	0.708	0.923	0.756	0.568	0.824	0.605	0.343	0.574	0.360	0.143	0.265	0.120	0.048	0.047	0.055

1. The 5% significance level is used.
2. The p-values of Davies test are computed for the parameter space, $\lambda \in [-0.99, 0.99]$ with step length 0.005.

Table 3.6: Rejection Probabilities: $\rho = 0.2$

	$\lambda = 0.0$			$\lambda = 0.1$			$\lambda = 0.2$			$\lambda = 0.3$			$\lambda = 0.4$		
τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ
0.0	0.058	0.090	0.063	0.121	0.040	0.134	0.197	0.047	0.241	0.293	0.092	0.364	0.396	0.184	0.514
0.1	0.058	0.163	0.047	0.058	0.083	0.046	0.115	0.050	0.125	0.186	0.051	0.232	0.299	0.108	0.390
0.2	0.058	0.301	0.064	0.066	0.170	0.048	0.053	0.077	0.054	0.134	0.047	0.123	0.192	0.053	0.243
0.3	0.091	0.434	0.108	0.072	0.279	0.078	0.061	0.169	0.070	0.054	0.074	0.069	0.112	0.050	0.133
0.4	0.181	0.641	0.228	0.120	0.490	0.138	0.075	0.331	0.073	0.088	0.183	0.078	0.055	0.064	0.063
0.5	0.272	0.773	0.327	0.216	0.687	0.224	0.146	0.508	0.159	0.107	0.372	0.094	0.082	0.179	0.072
0.6	0.463	0.877	0.530	0.397	0.833	0.437	0.265	0.731	0.305	0.169	0.597	0.196	0.118	0.399	0.103
0.7	0.642	0.977	0.723	0.556	0.943	0.619	0.451	0.890	0.499	0.367	0.779	0.411	0.208	0.615	0.219
0.8	0.820	0.991	0.857	0.788	0.979	0.821	0.709	0.968	0.745	0.616	0.922	0.639	0.482	0.825	0.489
0.9	0.943	0.998	0.941	0.935	1.000	0.944	0.889	0.994	0.906	0.856	0.985	0.855	0.787	0.977	0.793
	$\lambda = 0.5$			$\lambda = 0.6$			$\lambda = 0.7$			$\lambda = 0.8$			$\lambda = 0.9$		
τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ	Davies	RS_τ	RS^*_τ
0.0	0.514	0.308	0.680	0.608	0.454	0.783	0.697	0.591	0.859	0.804	0.724	0.919	0.861	0.818	0.954
0.1	0.387	0.179	0.513	0.511	0.317	0.679	0.618	0.463	0.760	0.756	0.632	0.890	0.812	0.750	0.927
0.2	0.287	0.090	0.400	0.368	0.193	0.514	0.503	0.310	0.673	0.661	0.484	0.822	0.752	0.604	0.890
0.3	0.186	0.058	0.251	0.276	0.112	0.402	0.391	0.185	0.554	0.509	0.332	0.693	0.636	0.475	0.811
0.4	0.121	0.038	0.148	0.179	0.053	0.241	0.280	0.098	0.407	0.406	0.213	0.576	0.563	0.338	0.740
0.5	0.049	0.071	0.061	0.121	0.036	0.144	0.196	0.053	0.261	0.303	0.127	0.434	0.444	0.233	0.620
0.6	0.081	0.202	0.076	0.054	0.083	0.069	0.116	0.029	0.154	0.185	0.064	0.282	0.316	0.148	0.470
0.7	0.147	0.464	0.141	0.089	0.237	0.067	0.058	0.100	0.074	0.122	0.026	0.161	0.226	0.077	0.356
0.8	0.349	0.734	0.354	0.190	0.506	0.175	0.084	0.273	0.062	0.055	0.082	0.076	0.122	0.023	0.189
0.9	0.697	0.933	0.704	0.556	0.829	0.516	0.362	0.666	0.308	0.149	0.371	0.111	0.055	0.096	0.065

1. The 5% significance level is used.
2. The p-values of Davies test are computed for the parameter space, $\lambda \in [-0.99, 0.99]$ with step length 0.005.

Appendix A

TECHNICAL APPENDIX FOR CHAPTER 3

A.1 Score Functions

The model we consider is

$$\begin{aligned} y &= \rho W y + X\beta + u \\ u &= \tau W u + \epsilon - \lambda W \epsilon, \end{aligned} \tag{A.1}$$

and the log-likelihood function is

$$\begin{aligned} l(\theta) =& -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\epsilon'\epsilon + \log|I - \rho W| + \log|I - \tau W| + \log|(I - \lambda W)^{-1}| \\ =& \text{Constant} - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(AY - X\beta)'B'C^{-1\prime}C^{-1}B(AY - X\beta) \\ &+ \log|A| + \log|B| + \log|C^{-1}|, \end{aligned} \tag{A.2}$$

where $\theta = (\beta', \sigma^2, \rho, \tau, \lambda)$ and $A = I - \rho W$, $B = I - \tau W$, $C = I - \lambda W$.
The first derivatives are

$$\frac{\partial l}{\partial \beta} = \frac{1}{\sigma^2}X'B'C^{-1\prime}\epsilon$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\epsilon'\epsilon.$$

$$\frac{\partial l}{\partial \rho} = -\text{tr}(A^{-1}W) + \frac{1}{\sigma^2}\epsilon'C^{-1}BWy$$

$$\frac{\partial l}{\partial \tau} = -\text{tr}(B^{-1}W) + \frac{1}{\sigma^2}\epsilon'C^{-1}W(Ay - X\beta)$$

$$\frac{\partial l}{\partial \lambda} = -\text{tr}(WC^{-1}) - \frac{1}{\sigma^2}\epsilon'C^{-1}W\epsilon$$

Under $H_0 : \tau = \lambda$, we have $B = C$ and hence

$$\frac{\partial l}{\partial \beta}|_{H_0} = \frac{1}{\sigma^2} X'\epsilon$$

$$\frac{\partial l}{\partial \sigma^2}|_{H_0} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\epsilon'\epsilon.$$

$$\frac{\partial l}{\partial \rho}|_{H_0} = -tr(A^{-1}W) + \frac{1}{\sigma^2}\epsilon' W y$$

$$\frac{\partial l}{\partial \tau}|_{H_0} = -tr(B^{-1}W) + \frac{1}{\sigma^2}\epsilon' C^{-1} W(Ay - X\beta)$$

$$\frac{\partial l}{\partial \lambda}|_{H_0} = -tr(WC^{-1}) - \frac{1}{\sigma^2}\epsilon' C^{-1} W\epsilon$$

Moreover, if we put $\rho = 0$, then we have $A = I$ and under H_0, the above becomes

$$\frac{\partial l}{\partial \beta}|_{H_0} = \frac{1}{\sigma^2} X'\epsilon$$

$$\frac{\partial l}{\partial \sigma^2}|_{H_0} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\epsilon'\epsilon.$$

$$\frac{\partial l}{\partial \rho}|_{H_0} = -tr(A^{-1}W) + \frac{1}{\sigma^2}\epsilon' W y$$

$$\frac{\partial l}{\partial \tau}|_{H_0} = -tr(B^{-1}W) + \frac{1}{\sigma^2}\epsilon' C^{-1} W\epsilon$$

$$\frac{\partial l}{\partial \lambda}|_{H_0} = -tr(WC^{-1}) - \frac{1}{\sigma^2}\epsilon' C^{-1} W\epsilon$$

A.2 Information Matrices

From the previous section, the second derivatives of the log-likelihood function are

$$\frac{\partial^2 l}{\partial \beta \partial \beta'} = -\frac{1}{\sigma^2} X'B'C^{-1'}C^{-1}BX$$

$$\frac{\partial^2 l}{\partial \beta \partial \sigma^2} = \frac{\partial^2 l}{\partial \sigma^2 \partial \beta'} = -\frac{1}{\sigma^4} X'\epsilon$$

$$\frac{\partial^2 l}{\partial \beta \partial \rho} = \frac{\partial^2 l}{\partial \rho \partial \beta'} = -\frac{1}{\sigma^2} X'B'C^{-1'}C^{-1}BWy$$

$$\frac{\partial^2 l}{\partial\beta\partial\tau} = \frac{\partial^2 l}{\partial\tau\partial\beta'} = -\frac{1}{\sigma^2}X'W'C^{-1'}\epsilon - \frac{1}{\sigma^2}X'B'C^{-1'}C^{-1}W(Ay - X\beta)$$

$$\frac{\partial^2 l}{\partial\beta\partial\lambda} = \frac{\partial^2 l}{\partial\lambda\partial\beta'} = \frac{1}{\sigma^2}X'B'C^{-1'}W'C^{-1'}B\epsilon + \frac{1}{\sigma^2}X'B'C^{-1'}C^{-1}WC^{-1}B(Ay - X\beta)$$

$$\frac{\partial^2 l}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\epsilon'\epsilon$$

$$\frac{\partial^2 l}{\partial\sigma^2\partial\rho} = \frac{\partial^2 l}{\partial\rho\partial\sigma^2} = -\frac{1}{\sigma^4}\epsilon'C^{-1}BWy$$

$$\frac{\partial^2 l}{\partial\sigma^2\partial\tau} = \frac{\partial^2 l}{\partial\tau\partial\sigma^2} = -\frac{1}{\sigma^4}\epsilon'C^{-1}W(Ay - X\beta)$$

$$\frac{\partial^2 l}{\partial\sigma^2\partial\lambda} = \frac{\partial^2 l}{\partial\lambda\partial\sigma^2} = -\frac{1}{\sigma^4}\epsilon'C^{-1}W\epsilon$$

$$\frac{\partial^2 l}{\partial\rho^2} = -\mathrm{tr}(A^{-1}WA^{-1}W) - \frac{1}{\sigma^2}y'W'B'C^{-1'}C^{-1}BWy$$

$$\frac{\partial^2 l}{\partial\rho\partial\tau} = \frac{\partial^2 l}{\partial\tau\partial\rho} = -\frac{1}{\sigma^2}(Ay - X\beta)'W'C^{-1'}C^{-1}BWy - \frac{1}{\sigma^2}\epsilon'C^{-1}WWy$$

$$\frac{\partial^2 l}{\partial\rho\partial\lambda} = \frac{\partial^2 l}{\partial\lambda\partial\rho} = \frac{1}{\sigma^2}(Ay - X\beta)'B'C^{-1'}W'C^{-1'}C^{-1}BWy + \frac{1}{\sigma^2}\epsilon'C^{-1}WC^{-1}BWy$$

$$\frac{\partial^2 l}{\partial\tau^2} = -\mathrm{tr}(B^{-1}WB^{-1}W) - \frac{1}{\sigma^2}(Ay - X\beta)'W'C^{-1'}C^{-1}W(Ay - X\beta)$$

$$\frac{\partial^2 l}{\partial\tau\partial\lambda} = \frac{\partial^2 l}{\partial\lambda\partial\tau} = \frac{1}{\sigma^2}(Ay - X\beta)'B'C^{-1'}W'C^{-1'}C^{-1}W(Ay - X\beta) + \frac{1}{\sigma^2}\epsilon'C^{-1}WC^{-1}W(Ay - X\beta)$$

$$\begin{aligned}\frac{\partial^2 l}{\partial\lambda^2} =& -\mathrm{tr}(WC^{-1}WC^{-1}) - \frac{1}{\sigma^2}(Ay - X\beta)'B'C^{-1'}W'C^{-1'}C^{-1}W\epsilon \\ &- \frac{1}{\sigma^2}\epsilon'C^{-1}WC^{-1}W\epsilon - \frac{1}{\sigma^2}\epsilon'C^{-1}WC^{-1}WC^{-1}B(Ay - X\beta)\end{aligned}$$

Therefore the information matrix can be derived as

$$I(\theta) = \frac{1}{n\sigma^2} \begin{bmatrix} J_{\beta} & J_{\beta\sigma^2} & J_{\beta\rho} & J_{\beta\tau} & J_{\beta\lambda} \\ J_{\sigma^2\beta} & J_{\sigma^2} & J_{\sigma^2\rho} & J_{\sigma^2\tau} & J_{\sigma^2\lambda} \\ J_{\rho\beta} & J_{\rho\sigma^2} & J_{\rho} & J_{\rho\tau} & J_{\rho\lambda} \\ J_{\tau\beta} & J_{\tau\sigma^2} & J_{\tau\rho} & J_{\tau} & J_{\tau\lambda} \\ J_{\lambda\beta} & J_{\lambda\sigma^2} & J_{\lambda\rho} & J_{\lambda\tau} & J_{\lambda} \end{bmatrix}, \tag{A.3}$$

where

$$J_{\beta} = X'B'C^{-1'}C^{-1}BX$$

$$J_{\beta\sigma^2} = J_{\sigma^2\beta} = J_{\beta\tau} = J_{\tau\beta} = J_{\beta\lambda} = J_{\lambda\beta} = 0$$

$$J_{\beta\rho} = J_{\rho\beta} = X'B'C^{-1'}C^{-1}BA^{-1}WX\beta$$

$$J_{\sigma^2} = \frac{n}{2\sigma^2}$$

$$J_{\sigma^2\rho} = J_{\rho\sigma^2} = tr(C^{-1}BWA^{-1}B^{-1}C)$$

$$J_{\sigma^2\tau} = J_{\tau\sigma^2} = tr(C^{-1}WB^{-1}C)$$

$$J_{\sigma^2\lambda} = J_{\lambda\sigma^2} = -tr(C^{-1}W)$$

$$\begin{aligned} J_{\rho} =& \sigma^2 tr(A^{-1}WA^{-1}W) + \beta'X'A^{-1'}W'B'C^{-1'}C^{-1}BWA^{-1}X\beta \\ &+ \sigma^2 tr(C'B^{-1'}A{-}1'W'B'C^{-1'}C^{-1}BWA^{-1}B^{-1}C) \end{aligned}$$

$$J_{\rho\tau} = J_{\tau\rho} = \sigma^2[tr(C'B^{-1'}W'C^{-1'}C^{-1}BWA^{-1}B^{-1}C) + tr(C^{-1}WWA^{-1}B^{-1}C)]$$

$$J_{\rho\lambda} = J_{\lambda\rho} = -\sigma^2[tr(C'B^{-1'}W'C^{-1'}C^{-1}BWA^{-1}B^{-1}C) + tr(C^{-1}WC^{-1}BWA^{-1}B^{-1}C)]$$

$$J_{\tau} = \sigma^2[tr(C^{-1}BWA^{-1}B^{-1}C) + tr(C'B^{-1'}B'C^{-1'}W'C^{-1'}C^{-1}W)]$$

$$J_{\tau\lambda} = J_{\lambda\tau} = -\sigma^2[\mathrm{tr}(C'B^{-1'}W'C^{-1'}C^{-1}BWA^{-1}B^{-1}C) + \mathrm{tr}(C^{-1}WC^{-1}WC^{-1}BB^{-1}C)]$$

$$J_{\lambda} = -\sigma^2[\mathrm{tr}(C'B^{-1'}B'C^{-1'}W'C^{-1'}C^{-1}W) + \mathrm{tr}(C^{-1}WC^{-1}WC^{-1}BB^{-1}C)].$$

Under $H_0 : \tau = \lambda$, we have $B = c$, and hence (B.1) becomes

$$I(\theta) = \frac{1}{n\sigma^2}\left[\begin{array}{c|c} J_{11} & J_{12} \\ \hline J_{21} & J_{22} \end{array}\right], \tag{A.4}$$

where the partition matrices are

$$J_{11} = \begin{bmatrix} X'X & 0 \\ 0 & \frac{n}{2\sigma^2} \end{bmatrix},$$

$$J_{12} = J_{21}' = \begin{bmatrix} X'WA^{-1}X\beta & 0 & 0 \\ 0 & \mathrm{tr}(C^{-1}W) & -\mathrm{tr}(C^{-1}W) \end{bmatrix},$$

$$J_{22} = \begin{bmatrix} \sigma^2[\mathrm{tr}(A^{-1}WA^{-1}W) + \mathrm{tr}(A^{-1}W'WA^{-1})] + \beta'XA^{-1'}W'WA^{-1}X\beta & \sigma^2[\mathrm{tr}(W'C^{-1'}W) + \mathrm{tr}(C^{-1}WW)] & -\sigma^2[\mathrm{tr}(W'C^{-1'}W) + \mathrm{tr}(C^{-1}WW)] \\ \sigma^2[\mathrm{tr}(W'C^{-1'}W) + \mathrm{tr}(C^{-1}WW)] & \sigma^2[\mathrm{tr}(W'C^{-1'}C^{-1}W) + \mathrm{tr}(C^{-1}WC^{-1}W)] & -\sigma^2[\mathrm{tr}(W'C^{-1'}C^{-1}W) + \mathrm{tr}(C^{-1}WC^{-1}W)] \\ -\sigma^2[\mathrm{tr}(W'C^{-1'}W) + \mathrm{tr}(C^{-1}WW)] & -\sigma^2[\mathrm{tr}(W'C^{-1'}C^{-1}W) + \mathrm{tr}(C^{-1}WC^{-1}W)] & \sigma^2[\mathrm{tr}(W'C^{-1'}C^{-1}W) + \mathrm{tr}(C^{-1}WC^{-1}W)] \end{bmatrix}.$$

When $\tau = \lambda$ and $\rho = 0$, we further have $A = I$ and notice that $\mathrm{tr}(W) = 0$. Therefore (B.2) reduces to

$$I(\theta) = \frac{1}{n\sigma^2}\begin{bmatrix} X'X & 0 & X'WX\beta & 0 & 0 \\ 0 & \frac{n}{2\sigma^2} & 0 & \mathrm{tr}(C^{-1}W) & -\mathrm{tr}(C^{-1}W) \\ X'WX\beta & 0 & J_{\rho} & J_{\rho\tau} & J_{\rho\lambda} \\ 0 & \mathrm{tr}(C^{-1}W) & J_{\tau\rho} & J_{\tau} & J_{\tau\lambda} \\ 0 & -\mathrm{tr}(C^{-1}W) & J_{\lambda\rho} & J_{\lambda\tau} & J_{\lambda} \end{bmatrix}, \tag{A.5}$$

where

$$\begin{aligned} J_{\rho} &= \sigma^2[\mathrm{tr}(WW) + \mathrm{tr}(W'W)] + \beta'XW'WX\beta \\ J_{\rho\tau} &= J_{\tau\rho}^{J} = \sigma^2[\mathrm{tr}(W'C^{-1}W + \mathrm{tr}(C^{-1}WW)] \\ J_{\rho\lambda} &= J_{\lambda\rho}^{J} = -\sigma^2[\mathrm{tr}(W'C^{-1}W + \mathrm{tr}(C^{-1}WW)] \\ J_{\tau} &= \sigma^2[\mathrm{tr}(W'C^{-1'}C^{-1}W) + \mathrm{tr}(C^{-1}WC^{-1}W)] \\ J_{\tau\lambda} &= J_{\lambda\tau}^{J} = -\sigma^2[\mathrm{tr}(W'C^{-1'}C^{-1}W) + \mathrm{tr}(C^{-1}WC^{-1}W)] \\ J_{\lambda} &= \sigma^2[\mathrm{tr}(W'C^{-1'}C^{-1}W) + \mathrm{tr}(C^{-1}WC^{-1}W)]. \end{aligned}$$

A.3 Derivation of Test Statistics

We first derive adjusted RS test statistic assuming λ is given. Defining $\underline{\theta} = (\beta', \rho, \tau, \sigma^2)'$, the log-likelihood can be rewritten as

$$\begin{aligned} l(\underline{\theta}|\lambda) = & \text{Constant} - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}[(I-\rho W)Y - X\beta]'(I-\tau W)'C^{-1\prime}\cdot \\ & C^{-1}(I-\tau W)[(I-\rho W)Y - X\beta] + \log|I-\rho W| + \log|I-\tau W| + \log|C^{-1}|, \end{aligned} \tag{A.6}$$

And for a given value of λ, the score functions under joint null are

$$\begin{aligned} d_\beta(\lambda) &= \frac{1}{\sigma^2}X'(Y - X\beta) \\ d_{\sigma^2}(\lambda) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta)'(Y - X\beta) \\ d_\rho(\lambda) &= \frac{1}{\sigma^2}(Y - X\beta)'WY \\ d_\tau(\lambda) &= \frac{1}{\sigma^2}(Y - X\beta)'C^{-1}W(Y - X\beta) \end{aligned}$$

The information matrix under the null when $\rho = 0$ and λ is given, denoted as $I(\underline{\theta}|\lambda)|_{H_0}$, can be derived as

$$\begin{aligned} I(\theta|\lambda)|_{H_0^J} &= \frac{1}{n\sigma^2}\begin{bmatrix} J_\beta(\lambda) & J_{\beta\sigma^2}(\lambda) & J_{\beta\rho}(\lambda) & J_{\beta\tau}(\lambda) \\ J_{\sigma^2\beta}(\lambda) & J_{\sigma^2}(\lambda) & J_{\sigma^2\rho}(\lambda) & J_{\sigma^2\tau}(\lambda) \\ J_{\rho\beta}(\lambda) & J_{\rho\sigma^2}(\lambda) & J_\rho(\lambda) & J_{\rho\tau}(\lambda) \\ J_{\tau\beta}(\lambda) & J_{\tau\sigma^2}(\lambda) & J_{\tau\rho}(\lambda) & J_\tau(\lambda) \end{bmatrix} \\ &= \frac{1}{n\sigma^2}\begin{bmatrix} X'X & 0 & X'WX\beta & 0 \\ 0 & \frac{n}{2\sigma^2} & 0 & tr(C^{-1}W) \\ X'WX\beta & 0 & \sigma^2[tr(WW) + tr(W'W)] + \beta'XW'WX\beta & \sigma^2[tr(W'C^{-1}W + tr(C^{-1}WW)] \\ 0 & tr(C^{-1}W) & \sigma^2[tr(W'C^{-1}W + tr(C^{-1}WW)] & \sigma^2[tr(W'C^{-1\prime}C^{-1}W) + tr(C^{-1}WC^{-1}W)] \end{bmatrix}, \end{aligned} \tag{A.7}$$

Denoting $\gamma = (\beta', \sigma^2)$, the standard RS test on $(\gamma, 0, \tau)$ given λ has the form

$$RS(\lambda) = \frac{1}{n}d_\tau'(\lambda)J_{\tau\cdot\gamma}^{-1}(\lambda)d_\tau(\lambda),$$

where

$$\begin{aligned} J_{\tau\cdot\gamma}(\lambda) &= J_\tau(\lambda) - J_{\tau\gamma}(\lambda)J_\gamma^{-1}(\lambda)J_{\gamma\tau}(\lambda) \\ &= \frac{1}{n\sigma^2}\{\sigma^2[tr(W'C^{-1\prime}C^{-1}W) + tr(C^{-1}WC^{-1}W)] - \frac{2\sigma^2}{n}[tr(C^{-1}W)]^2\} \end{aligned}$$

using (C.2), and the standard RS test statistic for fixed λ can be derived as

$$\mathrm{RS}(\lambda) = \frac{\{\frac{1}{\hat{\sigma}^2}\hat{u}'C^{-1}W\hat{u} - \mathrm{tr}[C^{-1}W]\}^2}{\mathrm{tr}\{W'C^{-1\prime}C^{-1}W + [C^{-1}W]^2\} - \frac{2}{n}[\mathrm{tr}(C^{-1}W)]^2}. \quad (A.8)$$

Now we consider the RS test adjusted for the presence of ρ. Assuming λ given, the adjusted RS test statistic, denoted by $\mathrm{RS}^*_\tau(\lambda)$ has the form

$$\begin{aligned}\mathrm{RS}^*_\tau(\lambda) =& \frac{1}{n}[d_\tau(\lambda) - J_{\tau\rho\cdot\gamma}(\lambda)J^{-1}_{\rho\cdot\gamma}(\lambda)d_\rho(\lambda)]' \\ &\times [J_{\tau\cdot\gamma}(\lambda) - J_{\tau\rho\cdot\gamma}(\lambda)J^{-1}_{\rho\cdot\gamma}(\lambda)J_{\rho\tau\cdot\gamma}(\lambda)]^{-1} \\ &\times [d_\tau(\lambda) - J_{\tau\rho\cdot\gamma}(\lambda)J^{-1}_{\rho\cdot\gamma}(\lambda)d_\rho(\lambda)],\end{aligned}$$

where

$$\begin{aligned}J_{\tau\rho\cdot\gamma}(\lambda) &= J_{\rho\tau\cdot\gamma}(\lambda) = J_{\tau\gamma}(\lambda)J^{-1}_{\gamma}(\lambda)J_{\gamma\rho}(\lambda) \\ J_{\rho\cdot\gamma}(\lambda) &= J_\rho(\lambda) - J_{\rho\gamma}(\lambda)J^{-1}_{\gamma}(\lambda)J_{\gamma\rho}(\lambda).\end{aligned}$$

Using (C.2), we have

$$\begin{aligned}J_{\tau\rho\cdot\gamma}(\lambda) &= \frac{1}{n}[\mathrm{tr}(W'C^{-1\prime}W) + \mathrm{tr}(C^{-1}WW)] \\ J_{\rho\cdot\gamma}(\lambda) &= \frac{1}{n\sigma^2}\{\sigma^2[\mathrm{tr}(W'W + WW)] + \beta'X'W'MWX\beta\},\end{aligned}$$

where $M = I - X(X'X)^{-1}X'$, and hence the adjusted RS statistic for fixed λ can be derived as

$$\mathrm{RS}^*(\lambda) = \frac{\{\hat{u}'C^{-1}W\hat{u}/\hat{\sigma}^2 - \mathrm{tr}(C^{-1}W) - [\mathrm{tr}(W'C^{-1\prime}W + C^{-1}WW)](n\hat{J}_{\rho\cdot\gamma})^{-1}\hat{u}'Wy/\hat{\sigma}^2\}^2}{\mathrm{tr}(W'C^{-1\prime}C^{-1}W + C^{-1}WC^{-1}W) - \frac{2}{n}[\mathrm{tr}(C^{-1}W)]^2 - [\mathrm{tr}(W'C^{-1\prime}W + C^{-1}WW)]^2(n\hat{J}_{\rho\cdot\gamma})^{-1}} \quad (A.9)$$

where

$$(n\hat{J}_{\rho\cdot\gamma})^{-1} = \hat{\sigma}^2 \cdot \{\hat{\beta}X'W'MWX\beta + \hat{\sigma}^2[\mathrm{tr}((W' + W)W)]\}^{-1}.$$

References

Andersson, M. and U. Grasjo. Spatial dependence and the representation of space in empirical models. *The Annals of Regional Science*, 43:159–180, 2003.

Andrews, D.W.K. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, 1993.

Anselin, L. *Spatial econometrics: methods and models*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988.

Anselin, L. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2):153–166, 2003.

Anselin, L. and A. K. Bera. Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah, A. and D. E. A. Giles, editors, *Handbook of Applied Economic Statistics*, pages 237–289. Marcel Dekker, New York, 1998.

Anselin, L., A. K. Bera, R. Florax, and M. J. Yoon. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26:77–104, 1996.

Arnold, M. and D. Wied. Improved gmm estimation of the spatial autoregressive error model. *Economics Letters*, 108(1):65–68, 2010.

Balestra, P. and M. Nerlove. Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica*, 34(3):585–612, 1966.

Balke, N. S. and T. B. Fomby. Threshold cointegration. *International Economic Review*, 38(3): 627–645, 1997.

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

Baltagi, B. H., S. H. Song, B. C. Jung, and W. Koh. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics*, 140(1):5–51, 2007.

Bartlett, M. S. On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*, 8(1):27–41, 1946.

Basdas, U. Spatial econometric analysis of the determinants of location in turkish manufacturing industry. Working paper, NBER Series, 2009.

Behrens, K., C. Ertur, and W. Koch. Dual gravity: using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics*, 27(5):773–794, 2012.

Bera, A. K. and S. Ra. A test for white noise against arma(1,1). Paper presented at the Winter Meeting of the Econometric Society, Boston, January, 1994, 1993.

Bera, A. K. and M. J. Yoon. Specification testing with locally misspecified alternatives. *Econometric Theory*, 9:649–658, 1993.

Beveridge, S. and C. R. Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "business cycle.". *Journal of Monetary Economics*, 7(2):151–174, 1981.

Blonigen, B. A., R. B. Davies, G. R. Waddell, and H. Naughton. Fdi in space: Spatial autoregressive lags in foreign direct investment. Working paper, NBER Series, 2004.

Boarnet, M. G. and A. Glazer. Federal grants and yardstick competition. *Journal of Urban Economics*, 52:53–64, 2002.

Burridge, P. On the cliff-ord test for spatial correlation. *Journal of the Royal Statistical Society B*, 42:107–108, 1980.

Case, A. C., H. S. Rosen, and J. R. Hines. Budget spillovers and fiscal policy interdependence: evidence from the states. *Journal of Public Economics*, 52:285–307, 1993.

Cliff, A. and J. K. Ord. *Spatial processes: models and applications*. Pion, London, 1981.

Davidson, R. and J. G. MacKinnon. Implicit alternatives and the local power of test statistics. *Econometrica*, 55:1305–1329, 1987.

Davies, R. B. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2):247–254, 1977.

Davies, R. B. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1):33–43, 1987.

Ding, Z., C. W. J. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.

Dogan, O. and S. Taspinar. Gmm estimation of spatial autoregressive models with moving average disturbances. *Regional Science and Urban Economics*, 43(6):903–926, 2013.

D.W.K., Andrews and W. Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6):1383–1414, 1994.

D.W.K., Andrews and W. Ploberger. Testing for serial correlation against an arma(1,1) process. *Journal of the American Statistical Association*, 91:1331–1342, 1996.

Egger, P. and M. Pfaffermayr. Spatial convergence. *Papers in Regional Science*, 85(2):199–215, 2006.

Elhorst, J. P. Matlab software for spatial panels. *International Regional Science Review*, 37: 389–405, 2014.

Filiztekin, A. Regional unemployment in turkey. *Papers in Regional Science*, 88(4):863–879, 2009.

Fingleton, B. A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices. *Empirical Economics*, 34(1):35–37, 2008.

Fingleton, B. and E. Lopez-Bazo. Empirical growth models with spatial effects. *Papers of the Regional Science*, 85:177–198, 2006.

Florax, R., H. Folmer, and S. J. Rey. Specification searches in spatial econometrics: The relevance of hendry's methodology. *Regional Science and Urban Economics*, 33:557–579, 2003.

Florax, R. J. G. M. *The University: A Regional Booster? : Economic Impacts of Academic Knowledge Infrastructure*. Aldershot, Aldershot, Hants, England and Brookfield, Vt., USA, 1992.

Frank, B. and M. Botolf. The german itc industry: Spatial employment and innovation patterns. Discussion paper, German Institute for Economic Research (DIW), 2007.

Garcia, R. and P. Perron. An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, 78(1):111–125, 1996.

Garretsen, H. and J. Peeters. Fdi and the relevance of spatial linkages: Do third-country effects matter for dutch fdi? *Review of World Economics*, 145:319–338, 2009.

Garrett, T. A. and T. L. Marsh. The revenue impacts of cross-border lottery shopping in the presence of spatial autocorrelation. *Regional Science and Urban Economics*, 32:501–519, 2002.

Godfrey, L. G. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, 46(6):1293–1301, 1978.

Granger, C. W. J. and M. J. Morris. Time series modeling and interpretation. *Journal of the Royal Statistical Society. Series A (General):*, 139(2):246–257, 1976.

Griffith, D. A. and G. Arbia. Detecting negative spatial autocorrelation in georeferenced random variables. *International Journal of Geographical Information Science*, 24:417–437, 2010.

Haining, R. P. The moving average model for spatial interaction. *Transactions of the Institute of British Geographers*, 3(2):202–225, 1978.

Haining, R. P. *Spatial data analysis in the social and environmental sciences*. Cambridge University Press, Cambridge, 1990.

Hallin, M. and D. Paindaveine. Optimal procedures based on interdirections and pseudo-mahalanobis ranks for testing multivariate elliptic white noise against arma dependence. *Bernoulli*, 8(6):787–815, 2002.

Hallin, M., J. F. Ingenbleek, and M. L. Puri. Linear serial rank tests for randomness against arma alternatives. *The Annals of Statistics*, 13(3):1156–1181, 1985.

Hansen, B. E. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–430, 1996.

Hansen, B. E. Threshold effects in non-dynamic panels: estimation, testing, and inference. *Journal of econometrics*, 93(2):345–368, 1999.

Harrison, D. and D. L. Rubinfeld. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5:81–102, 1978.

Hepple, L. W. Bayesian and maximum likelihood estimation of the linear model with spatial moving average disturbances. School of Geographical Sciences, University of Bristol, Working Papers Series, 2003.

Holtz-Eakin, D. Public-sector capital and the productivity puzzle. *Review of Economics and Statistics*, 76:12–21, 1994.

Jin, F. and L. F. Lee. Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. *Regional Science and Urban Economics*, 43(4):590–616, 2013.

Keim, D. B. and R. F. Stambaugh. Predicting returns in the stock and bond markets. *Journal of Financial Economics*, 17(2):357–390, 1986.

Kelejian, H. H. and I. R. Prucha. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533, 1999.

Kelejian, H. H. and I. R. Prucha. On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics*, 104(2):219–257, 2001.

Kendall, M. G. and A. Stuart. *The Advanced Theory of Statistics*, volume 1. Griffin, London, 1966.

Kim, C. W., T. T. Phipps, and L. Anselin. Measuring the benefits of air quality improvement: a spatial hedonic approach. *Journal of Environmental Economics and Management*, 45:24–39, 2003.

King, M. L. A point optimal test for autoregressive disturbances. *Journal of Econometrics*, 27: 21–37, 1985.

King, M. L. and T. S. Shively. Locally optimal testing when a nuisance parameter is present only under the alternative. *The Review of Economics and Statistics*, 75(1):1–7, 1991.

Kramer, W. and H. Zeisel. Finite sample power of linear regression autocorrelation tests. *Journal of Econometrics*, 43:363–372, 1990.

Lam, C. and Souza P. Cl. Regularization for spatial panel time series using adaptive lasso. Mimeo, 2013.

L'Esperance, W. L. and D. Taylor. The power of four tests of autocorrelation in the linear regression model. *Journal of Econometrics*, 3:1–21, 1975.

Lim, C. and M. McAleer. Time series forecasts of international travel demand for australia. *Tourism Management*, 23(4):389–396, 2002.

McLeod, A. I. and W. K. Li. Diagnostic checking arma time series models using squared?residual autocorrelations. *Journal of Time Series Analysis*, 4(4):269–273, 1983.

Mur, J. and A. Angulo. Model selection strategies in a spatial setting: Some additional results. *Regional Science and Urban Economics*, 39:200–213, 2009.

Myrdal, G. *Economic Theory of Underdeveloped Regions*. Gerard Dicksworth, London, 1957.

P., Moran. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):243–251, 1948.

Pace, R. K. and O. W. Gilley. Using the spatial configuration of the data to improve estimation. *Journal of the Real Estate Finance and Economics*, 14:333–340, 1997.

Pace, R. K. and J. P. LeSage. *Introduction to spatial econometrics*. Chapman and Hall/CRC, Boca Raton, 2009.

Park, S.B. On the small-sample power of durbin's h test. *Journal of the American Statistical Association*, 70:60–63, 1975.

Pavlyuk, D. Spatial analysis of regional employment rates in latvia. *Scientific Journal of Riga Technical University*, 2:56–62, 2011.

Poskitt, D. S. and A. R. Tremayne. Testing the specification of a fitted autoregressive-moving average model. *Biometrika*, 67(2):359–363, 1980.

Rayner, R. K. The small-sample power of durbin's h test revisited. *Computational Statistics and Data Analysis*, 17:87–94, 1994.

Rey, S. J. Montouri and B. D. U.s. regional income convergence: A spatial econometric perspective. *Regional Studies*, 33:143–156, 1999.

Saavedra, L. A. A model of welfare competition with evidence from afdc. *Journal of Urban Economics*, 47:248–279, 2000.

Saikkonen, P. Asymptotic relative efficiency of the classical tests under misspecification. *Journal of Econometrics*, 42:351–369, 1989.

Sen, M., A. K. Bera, and Y. H. Kao. A hausman test for spatial regression model. In Badi H Baltagi, R. Carter Hill, Whitney K. Newey, and Halbert L. White, editors, *Essays in Honor of Jerry Hausman (Advances in Econometrics 29)*, pages 547–559. Emerald Group Publishing Limited, 2012.

Sharpe, K. Some properties of the crossings process generated by a stationary chi-squared process. *Advances in Applied Probability*, 10:373–391, 1978.

Silvey, S. D. The lagrange multiplier test. *Annals of Mathematical Statistics*, 30:389–407, 1959.

Stock, J. H. and M. W. Watson. Business cycle fluctuations in us macroeconomic time series. In Stock, J. H., editor, *Handbook of Macroeconomics Vol.1 Part A*, pages 3–64. 1999.

Walker, G. T. On periodicity in series of related terms. *Proceedings of the Royal Society of London Series A*, 131:518–532, 1931.

Watson, M. W. and R. F. Engle. Testing for regression coefficient stability with a stationary ar(l) alternative. *Review of Economics and Statistics*, 67:341–346, 1985.

Wold, H. *A Study in the Analysis of Stationary Time Series*. Almqvist and Wiksell, Uppsala, Stockholm, 1938.

Yao, Q. and P. J. Brockwell. Gaussian maximum likelihood estimation for arma models ii: spatial processes. *Bernoulli*, 12(3):403–429, 2006.

Yule, G. U. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London Series A*, 226:267–298, 1927.