R. Kelley Pace
James P. LeSage

Likelihood Dominance Spatial Inference

Many users estimate spatial autoregressions to perform inference on regression parameters. However, as the sample size or the number of potential models rise, computational exigencies make exact computation of likelihood-based inferences tedious or even impossible. To address this problem, we introduce a lower bound on the likelihood ratio test that can allow users to conduct conservative maximum likelihood inference while avoiding the computationally demanding task of computing exact maximum likelihood point estimates. This form of inference, known as likelihood dominance, performs almost as well as exact likelihood inference for the empirical examples examined. We illustrate the utility of the technique by performing likelihood-based inference on parameters from a spatial autoregression involving 890,091 observations in less than a minute (given the spatial weight matrix).

The computational challenges of likelihood-based spatial estimation problems have led to the introduction of a variety of alternative approaches such as the pseudo-likelihood technique of Besag (1975) or the instrumental variable technique of Kelejian and Prucha (1998). Such techniques may produce good point estimates, but they fail to determine the statistical significance of selected regression parameters, an important goal for many users of spatial autoregressions.

We set forth a computationally simple approach to inference that employs estimates based on lower and upper bounds for the spatial autoregressive parameter in the model. Our approach maps these bounds on the autoregressive parameter to a lower bound on the likelihood ratio associated with testing hypotheses on regression parameters. Of course, this also translates into a lower bound on the deviance, which equals twice the difference between the profile log-likelihood of the overall model and the restricted model. By construction, a deviance has a minimum of 0 and is distributed as chi-squared with degrees-of-freedom equal to the number of hypothesis under test.[1] The lower bound on the deviance constructed in this fashion represents

We would like to gratefully acknowledge the research support received from the National Science Foundation (BCS-0136193 and BCS-0136229). In addition, Pace would like to thank Ana Militino, Lola Ugarte, and the other seminar participants at the Universidad Pública de Navarra.

1. See Fan, Hung, and Wong (2000) for a recent derivation.

R. Kelley Pace is LREC Endowed Chair of Real Estate, Department of Finance, E. J. Ourso College of Business Administration, Louisiana State University. E-mail: kelley@pace.am. *James P. LeSage is a professor in the Department of Economics, University of Toledo. E-mail:* jlesage@spatial-econometrics.com.

Geographical Analysis, Vol. 35, No. 2 (April 2003) The Ohio State University
Submitted: February 13, 2002. Revised version accepted: October 8, 2002.

a form of "likelihood dominance" (Pollack and Wales [1991]), which permits rejection of restricted models in some cases without computation of the unrestricted model.

Bounds on the spatial autoregressive parameter required for this method could come from a number of sources. For example, bounds could arise from a combination of point estimates and prior information derived from Monte Carlo experiments. Estimates from similar models based on alternative data samples might represent another source for bounding the spatial autoregressive parameter. Even simple bounds that restrict the spatial autoregressive parameter to the interval [0,1) can prove adequate for conducting inference in some situations. Finally, better bounds exist such as the Monte Carlo bounds proposed by Barry and Pace (1999) and the quadratic bounds proposed by Pace and LeSage (2002). Both methods provide greater accuracy for somewhat higher computational cost.

The relation between the bounds on the spatial dependence parameter and the resulting bounded deviance tests derives from a partitioning of the overall deviance concerning the significance of some hypothesis into two separate deviances. The first deviance depends only on the ratio of two quadratic forms, which does not require computation of the troublesome log-determinant term used in maximum likelihood estimation. The second deviance depends on the difference between (a) the value of the profile log-likelihood from the unrestricted model evaluated at the optimal autoregressive parameter estimate, and (b) the same profile log-likelihood evaluated at a restricted autoregressive parameter estimate. Significance of either deviance implies significance of the overall deviance test, since the minimum of any deviance equals 0 by construction. The first deviance (a) is easily calculated and the second deviance (b) is relatively small in many applied situations. Section 1 develops these ideas and provides a graphical illustration of these relations.

To illustrate the method, we apply four simple bounded intervals to three different data sets to explore performance of the proposed bounded deviance tests in an empirical setting. For two of the empirical examples we compute both exact and bounded deviance tests to assess the relative performance of both forms of inference. These two examples involved over 1,000 hypotheses tests, with the results from this evaluation suggesting that bounded deviance tests often yield the same inferences as those from exact deviance tests. We term any two testing procedures that produce the same set of inferences as qualitatively identical. In the first example based on a sample of county-level election data, up to 100% of the bounded deviance tests yielded qualitative results identical to those from actual deviance tests. In the second example based upon census tract-level data on consumer expenditures, up to 99.8% of the bounded deviance tests yielded qualitative results identical to those from actual deviance tests.

The third example of the proposed approach illustrates the ability of the method to quickly produce inferential results for a large problem (a hedonic housing model using 890,091 observations from U.S. census blocks). Given a spatial weight matrix, the subsequent estimation took under a minute and provided decisive inference on all model parameters.

Section 1 develops these ideas and provides a graphical illustration of these relations. Section 2 examines the performance of both forms of inference in three empirical examples, and Section 3 concludes with the key implications.

1. SPATIALLY ROBUST INFERENCE

In this section we describe two likelihood dominance relations that can produce inferences regarding hypothesized parameter values in spatial autoregressive models. These likelihood inferences do not require computing maximum likelihood estimates for alternative models involved in the tests.

To begin, let the covariance matrix depend on the scalar parameter α, representing the magnitude of spatial dependence in the model. If the errors, ε, follow a normal distribution with the parameterized covariance matrix $\Omega(\alpha)$, (1) expresses the profile log-likelihood (Anselin [1988, p. 182]),

$$L(\alpha) = C - \frac{1}{2}\ln\left|\Omega(\alpha)\right| - \frac{n}{2}\ln\left(\left(y - X\tilde{\beta}(\alpha)\right)'\Omega(\alpha)^{-1}\left(y - X\tilde{\beta}(\alpha)\right)\right) \tag{1}$$

where C designates a constant, y represents n observations on the dependent variable, and X represents n observations on k independent variables (one of these columns equals a vector of ones when using an intercept in the model).[2] Equation (2),

$$\tilde{\beta}(\alpha) = \left(X'\Omega(\alpha)^{-1}X\right)^{-1}X'\Omega(\alpha)^{-1}y \tag{2}$$

represents a generalized least-squares (GLS) estimate parameterized by α. At the optimal value of α, denoted by $\tilde{\alpha}$, the estimate $\tilde{\beta}(\tilde{\alpha})$ becomes the maximum likelihood estimate of β.

This formulation can represent a variety of specific spatial models. For example, the simultaneous autoregression (SAR) specification arises when $\Omega(\alpha)^{-1} = (I - \alpha D)'(I - \alpha D)$ where D represents a spatial weight matrix.[3] The conditional autoregression (CAR) specification arises when $\Omega(\alpha)^{-1} = (I - \alpha D)$ for symmetric D.[4] The moving average autoregression (MA) specification arises when $\Omega(\alpha) = (I - \alpha D)$ for symmetric D.[5] Similar profile log-likelihoods exist for the Gaussian, exponential, and spherical covariance specifications as well as for spatially autoregressive models or for mixed regressive spatially autoregressive models.

Likelihood ratio hypotheses tests involve a comparison of restricted and unrestricted estimates. One can impose linear restrictions of the form $R\beta = r$ involving J linear restrictions where R represents a J by k restriction matrix, and r represents a J by 1 vector. In this case, the restricted estimates take the form shown in (3), with the associated restricted profile log-likelihood shown in (4).

$$\tilde{\beta}_R(\alpha_R) = \tilde{\beta}(\alpha_R) - \left(X'\Omega(\alpha_R)^{-1}X\right)^{-1}R'\left[R\left(X'\Omega(\alpha_R)^{-1}X\right)^{-1}R'\right]^{-1}(R\beta - r) \tag{3}$$

$$L_R(\alpha_R) = C - \frac{1}{2}\ln\left|\Omega(\alpha_R)\right| - \frac{n}{2}\ln\left(\left(y - X\tilde{\beta}_R(\alpha_R)\right)'\Omega(\alpha_R)^{-1}\left(y - X\tilde{\beta}_R(\alpha_R)\right)\right) \tag{4}$$

Equation (4) expresses the profile log-likelihood as a function of the parameter α_R. The value, $\tilde{\alpha}_R$ maximizing (4) represents the restricted likelihood estimate for the parameter α_R. Twice the difference between the unrestricted and restricted profile log-likelihoods, $2(L(\tilde{\alpha}) - L_R(\tilde{\alpha}_R))$, is often referred to as the deviance. For normal maximum likelihood, the deviance follows a $\chi^2(J)$ distribution asymptotically. Critical

2. One can form similar decompositions between the log-determinant and a function of the error term for other continuous densities.

3. See Anselin (1988), Bavaud (1998), as well as Griffith and Lagona (1998) for definitions of spatial weight matrices. See Anselin (1988) and Ord (1975) for more on the SAR model.

4. See Besag (1975) for more on the CAR model.

5. See Haining (1990) for more on the moving average model as well as a discussion of the other types of spatial autoregressions.

values based on this distribution can be used to assess the probability of obtaining a higher test statistic under repeated sampling. If V^* represents the critical value associated with the chosen level of significance, a hypothesis test is defined as statistically significant when the deviance exceeds V^*. For example, imposing a single restriction, $\alpha = 0$ (equivalent to OLS), permits testing the statistical significance of the spatial autoregressive parameter α. The associated deviance test would follow a $\chi^2(1)$ distribution and a calculated deviance exceeding the critical value of $V^* = 6.63$ for $\chi^2(1)$ would allow rejection at the 1% level of the least-squares model associated with an hypothesis of no spatial dependence.

The development of bounded deviance tests relies on this background as well as the existence of bounds on the spatial dependence parameter. One possible set of bounds would be the interval defined by $\lambda_{\min}^{-1}$ and $\lambda_{\max}^{-1}$ where the λ denote eigenvalues of D, the spatial weight matrix. This represents the domain of α subject to $|I - \alpha D| > 0$ for all α.

Another bound could be based on restricted generalized least-squares (EGLS) estimates along with a maximizing choice of α, $\hat{\alpha}_{EGLS}$. For the common case of row-stochastic weight matrices and positive dependence, $0 \leq \tilde{\alpha} < \min(1,\hat{\alpha}_{EGLS})$. This occurs because the EGLS estimate, $\hat{\alpha}_{EGLS}$, exhibits positive bias relative to maximum likelihood and the maximum value for the autoregressive parameter is 1 for row-stochastic weight matrices.

The quadratic bounds proposed by Pace and LeSage (2002) provide another approach. These bounds provide shorter intervals than the simpler bounds above while avoiding computational difficulties associated with determining the eigenvalues for large spatial weight matrices. The confidence interval associated with the computationally efficient approximation to the log-determinant proposed by Barry and Pace (1999) constitutes another slightly more elaborate and accurate set of bounds.

Finally, other approaches exist. Techniques proposed by Golub and Von Matt (1995) to compute confidence bounds might work well and approximations such as Besag (1975), Griffith and Sone (1995), Kelejian and Prucha (1998), as well as Smirnov and Anselin (2001) might lead to bounds on the spatial parameter.

To develop the bounded tests, assume the existence of bounds to the autoregressive parameter that lie in the interval associated with a positive determinant.

Assumption. $\lambda_{\min}^{-1} < \alpha_L \leq \alpha_R \leq \alpha_U < \lambda_{\max}^{-1}$.

Proposition 1 formalizes the bounded deviance tests.

Proposition 1. *If* $2(L(\alpha_B) - L_R(\alpha_B)) > V^*$ *for all* $\alpha_B \in [\alpha_L,\alpha_U]$, *then* $2(L(\tilde{\alpha}) - L_R(\tilde{\alpha}_R)) > V^*$.

Proof. The overall deviance $2(L(\tilde{\alpha}) - L_R(\tilde{\alpha}_R))$ partitions into the sum of two deviances $2(L(\tilde{\alpha}) - L_R(\tilde{\alpha}_R)) + 2(L(\tilde{\alpha}_R) - L_R(\tilde{\alpha}_R))$. By the premise of this proposition and Assumption 1, $2(L(\alpha_B) - L_R(\alpha_B)) > V^*$ for all α_B in the interval $[\alpha_L,\alpha_U]$ containing $\tilde{\alpha}_R$. Thus, the second deviance is significant while the first deviance must be non-negative since unrestricted log-likelihoods must exceed or match restricted log-likelihoods. Therefore, the overall deviance is significant and this proves the proposition. QED.

Corollary. *If* $S(\alpha_B) > V^*$, *where* $S(\alpha_B) = n\ln\left[\frac{\left(y-X\tilde{\beta}_R(\alpha_B)\right)'\Omega(\alpha_B)^{-1}\left(y-X\tilde{\beta}_R(\alpha_B)\right)}{\left(y-X\tilde{\beta}(\alpha_B)\right)'\Omega(\alpha_B)^{-1}\left(y-X\tilde{\beta}(\alpha_B)\right)}\right]$ for all α_B, *then* $2(L(\tilde{\alpha}) - L_R(\tilde{\alpha}_R)) > V^*$.

Proof. Substitution of the definitions of the profile log-likelihoods into the premise of Proposition 1 leads to the cancellation of the constants and determinants from the component profile log-likelihoods. Rearrangement of the expanded premise yields the result. QED.

To provide insight regarding use of these bounds for inference, the data in Table 1 is used as an example. The table contains latitude, longitude coordinates for nine census divisions, associated information on the log of commuting time per capita, and the log of the median gross rent for nine observations. We fit the model model $y = \alpha Wy + X\beta + \varepsilon$, where y is the log of median gross rent and the single independent variable X is the log of commuting time per capita along with a constant vector. A row-stochastic weight matrix W was constructed using a Delaunay triangle routine.

Figure 1 shows the resulting profile log-likelihoods for the unrestricted model and a restricted model based on deletion of the commuting time per capita independent variable. Deletion of the only non-constant independent variable leads to an autoregressive model $y = \text{intercept} + \alpha Wy + \varepsilon$. In the figure, dot symbols were used to represent points on the profile log-likelihoods and asterisks symbols denote the optimal value for both unrestricted and restricted profile log-likelihoods. Naturally, the curve with higher values represents the unrestricted profile log-likelihood.

TABLE 1
Commuting Example Data

Division	Latitude	Longitude	Commute	Rent
1	44.3928	70.6069	4.0616	6.3154
2	41.8597	76.4974	3.9475	6.1800
3	42.5678	86.8034	4.1309	6.0088
4	42.6922	97.0084	4.2818	5.9081
5	33.9472	80.8570	4.0489	6.1137
6	34.5472	87.0706	4.2093	5.8051
7	32.2965	97.3402	4.1822	5.9375
8	40.3386	110.2123	4.2418	6.0162
9	47.7500	123.5000	4.0193	6.3630

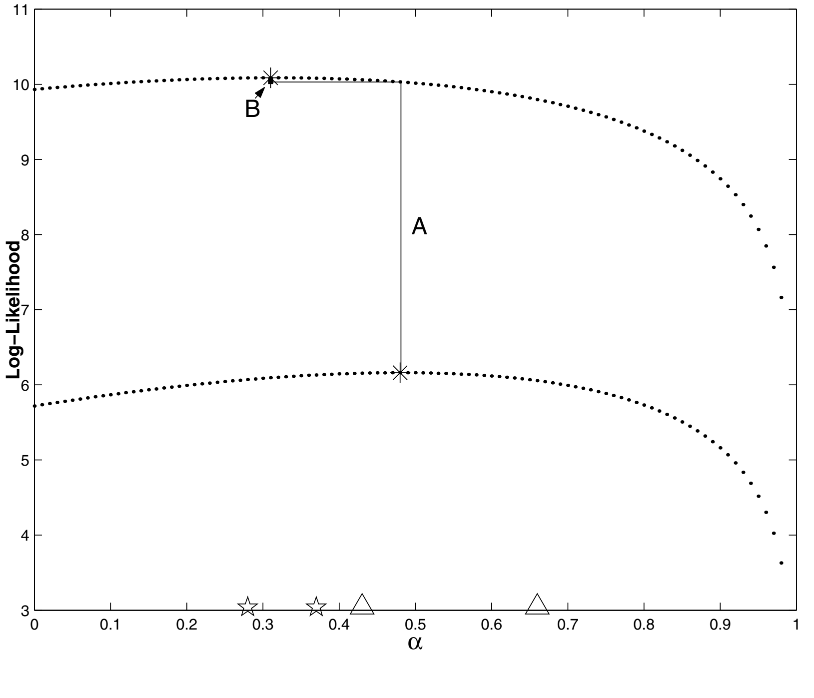

FIG. 1. Commuting Data Profile Log-likelihood

In the figure, two sets of bounds on the value of α are depicted. The two star symbols show bounds for the unrestricted model and the triangle symbols represent bounds for the restricted model. Both sets of bounds on the autoregressive parameters in the two models were constructed using quadratic bounds set forth in Pace and LeSage (2002).

The term $(L(\tilde{\alpha}_R) - L_R(\tilde{\alpha}_R))$ corresponds to the vertical distance between the restricted profile log-likelihood and the unrestricted model profile log-likelihood curves. In Figure 1, the vertical line segment labeled "A" at $\alpha = \tilde{\alpha}_R$ shows the distance between the two curves.

The term $(L(\tilde{\alpha}) - L(\tilde{\alpha}_R))$ corresponds to the vertical distance between a point $\tilde{\alpha}_R$ on the unrestricted profile log-likelihood and the optimum point $\tilde{\alpha}$ on the unrestricted profile log-likelihood. In Figure 1, this is depicted by the very short (almost imperceptible) vertical line segment labeled as "B" extending below the asterisk on the upper curve associated with the unrestricted profile log-likelihood.

In Figure 1, the vertical distance between the restricted model profile likelihood and the unrestricted model profile likelihood at the point $\tilde{\alpha}_R$ (segment A) greatly exceeds the vertical distance between a point $\tilde{\alpha}_R$ on the unrestricted profile log-likelihood and the optimum point $\tilde{\alpha}$ on this unrestricted profile log-likelihood (segment B).

Taking this a stage further, a user could easily compute the minimum vertical distance between the profile log-likelihood of the unrestricted and restricted models using a range of values based on the interval between the apex of the two triangle symbols in Figure 1, which represent the bounds on $\tilde{\alpha}_R$. This minimum vertical distance between the unrestricted and restricted log-likelihood curves over the bounded interval serves as a lower bound to the overall likelihood difference. In Figure 1, this minimum vertical difference almost equals the entire difference between the unrestricted and restricted profile log-likelihood maxima. For the bounds based on the interval between the two triangle symbols in Figure 1, the minimum difference over this interval is 3.75, and twice this distance is 7.5. The test statistic of 7.5 exceeds the one percent critical value of 6.63 for the chi-squared distribution with one degree-of-freedom.

To make this even more intuitive, we can recast these statistics as signed root deviances. The signed root deviance applies the sign of the coefficient to the square root of the deviance (Chen and Jennrich 1996). These statistics behave similar to t-statistics for large samples, and can be used in lieu of t-statistics for hypothesis testing. We will follow this approach in our empirical illustrations. In this example, the signed root deviance is 2.74, since the corresponding regression parameter has a positive sign. For this simple example, these results compare favorably to those based on the exact likelihood ratio of 7.85, and the value of 2.80 taken by the exact signed root deviance.

In summary, this example illustrates how inferences based on computationally simple and fast bounds can produce qualitatively identical inferences to those derived from exact maximum likelihood estimates. If interest centered only on a test for the impact of commuting time on rental values, users could accurately draw such an inference without actually computing exact maximum likelihood estimates. If users desire point estimates as well, they can employ a number of log-determinant approximations (e.g., Griffith and Sone 1995, Barry and Pace 1999, as well as Pace and LeSage 2003). The likelihood dominance inferential techniques proposed here allow users to engage in certain types of inference without producing maximum likelihood estimates. For example, model diagnostics often reduce to inferences regarding a supermodel. Examples include the RESET test of Ramsey (1974) for model misspecification, testing for spatial autocorrelation, which centers on rejection of $\alpha = 0$, and examination of different spatial weight matrices. In the latter problem, changing the weight matrix requires recomputation of the log-determinant for exact estimates, whereas likelihood dominance would allow more elaborate specification testing without these computations.

There are numerous ways of constructing bounds on the spatial dependence parameter α, so the ideas presented here do not require use of bounds determined with the method of Pace and LeSage (2002). Indeed, we hope that the ability to draw accurate inferences using bounds will stimulate future research into deriving alternative bounds.

We conclude this discussion with some comments on cases where the approach based on bounded deviances should succeed, and cases where the approach may fail. The best case for bounded deviances occurs with variables whose deletion has no effect on the optimum dependence parameter (i.e., $\tilde{\alpha}_R = \tilde{\alpha}$). In this case, the omitted term $(L(\tilde{\alpha}) - L(\tilde{\alpha}_R))$ vanishes and the bounds are sharp. Exclusion of variables with little spatial character (or whose spatial character may not affect the model parameters (cf., Getis and Griffith 2002) might fall into this category, as these should not cause the restricted and unrestricted autoregressive parameters to vary. Another case involves low autoregressive parameter values, since the bounds become exact when $\alpha = 0$.

A final case may arise in large samples. If the test statistics increase with the number of observations, *n* (Leamer 1988, pp. 290–95), this may serve to offset underestimation of the actual deviance by the bounded deviance. Given a large exact test statistic, underestimation of this statistic may still lead to a value that exceeds the critical value. This suggests that bounded deviances, or the associated signed root deviances, might perform well in large problems, one of their intended applications.[6]

The worst case for the deviance bounds occurs for variables whose deletion produces little change to the overall goodness-of-fit, while leading to a change in the optimum $\tilde{\alpha}_R$. In these cases, $(L(\tilde{\alpha}) - L(\tilde{\alpha}_R))$ is relatively large. Exclusion of variables with a primarily spatial character, such as a polynomial in the locational coordinates, might change the optimum autoregressive parameter while not greatly affecting the fit. However, an empirical illustration with such variables in the next section did not reveal this type of problem.

Proposition 1 and the Corollary represent a form of likelihood dominance result described by Pollack and Wales (1991). One can also determine another type of likelihood dominance using bounds on the log-determinant. The following result appeared informally in Pace and LeSage (2002). We repeat it here more formally using notation consistent with the previous development.

Definition 2. Let $L^U(\alpha)$ and $L^L(\alpha)$ represent bounded profile log-likelihood functions such that $L^U(\alpha) \geq L(\alpha) \geq L^L(\alpha)$ for all α. Let the bounded and exact profile log-likelihoods have optima such that $\tilde{\alpha}^U \geq \tilde{\alpha} \geq \tilde{\alpha}^L$. Note, the upper bounded log-likelihoods provide a smaller penalty to overfitting the autoregressive parameter and thus the upper bounded log-likelihood also results in an overestimate of the autoregressive parameter. The converse holds for the lower bounded log-likelihood. In addition, let $L_R^U(\alpha)$ and $L_R^L(\alpha)$ represent bounded restricted profile log-likelihood functions such that $L_R^U(\alpha) \geq L_R(\alpha) \geq L_R^L(\alpha)$ for all α. Let the bounded restricted and exact restricted profile log-likelihoods have optima such that $\tilde{\alpha}_R^U \geq \tilde{\alpha}_R \geq \tilde{\alpha}_R^L$.

Proposition 2. *If* $2(L^L(\tilde{\alpha}) - L_R^U(\tilde{\alpha}_R)) > V^*$, *then* $2(L(\tilde{\alpha}) - L_R(\tilde{\alpha}_R)) > V^*$.

Proof. Let $L^L(\tilde{\alpha}) = L(\tilde{\alpha}) - \Delta_1$ where Δ_1 is a non-negative scalar since $L^L(\tilde{\alpha}) \leq L(\tilde{\alpha})$ by Definition 1. Similarly, let $L_R^U(\tilde{\alpha}) = L_R(\tilde{\alpha}) + \Delta_2$ where Δ_2 is a non-negative scalar since $L_R^U(\tilde{\alpha}) \geq L(\tilde{\alpha})$ by Definition 1. Substituting these relations into $2(L^L(\tilde{\alpha}) - L_R^U(\tilde{\alpha}_R)) > V^*$ yields $2(L(\tilde{\alpha}) - L_R(\tilde{\alpha}_R)) - 2\Delta_1 - 2\Delta_2$. If $2(L(\tilde{\alpha}) - L_R(\tilde{\alpha}_R)) - 2\Delta_1 - 2\Delta_2 > V^*$ then $2(L(\tilde{\alpha}) - L_R(\tilde{\alpha}_R)) > V^* + 2\Delta_1 + 2\Delta_2 \geq V^*$ and this proves the proposition. QED.

6. Statistical significance provides a necessary, but not a sufficient condition for deeming a variable important. A statistically significant variable with an inconsequential magnitude may not meet the sufficient condition from a subject matter perspective.

To distinguish between the two results, we refer to likelihood dominance based on Proposition 1 as likelihood dominance of the first kind (abbreviated LD1). We label likelihood dominance using Proposition 2 as likelihood dominance of the second kind (abbreviated LD2). Taken together, these likelihood dominance results can permit inference without computing the full log-likelihood in various cases.

Both forms of likelihood dominance serve as lower bounds to the deviance test statistics. Hence, likelihood dominance test statistics exceeding their critical values imply rejection of restrictions under test. However, when likelihood dominance test statistics fall below critical values this does not imply failure to reject the restrictions, since the actual deviances may still exceed critical values. In these cases, likelihood dominance statistics are biased conservatively.[7]

2. EMPIRICAL ILLUSTRATIONS

To illustrate likelihood dominance, Sections A and B present two empirical illustrations where the sample size permits comparison of results from both exact maximum likelihood and likelihood dominance inferences. Section C illustrates the speed of likelihood dominance inference in an applied setting where the data set involves 890,091 observations, demonstrating that identical qualitative results arise from exact and bounded deviance tests.

2.1. Election Data

In this section we illustrate the two different forms of likelihood dominance using data on votes cast in the 1980 presidential election across 3,107 contiguous U.S. counties with complete data on selected variables. These sample data from Pace and Barry (1997) contain information on the number of recorded votes in the 1980 presidential election (*Votes*), the population 18 years of age or older (*Pop*), the population with a 12th grade or higher education (*Education*), the number of owner-occupied housing units (*Houses*), aggregate county-level income (*Income*), and locational coordinates (X_c, Y_c).[8]

A recurring theme in spatial econometrics is the interplay between modeling spatial heterogeneity and spatial dependence (Anselin 1988, pp. 119–36). Casetti (1972) devised one of the simplest schemes for modeling spatial heterogeneity based on interacting independent variables with a polynomial surface in the locational coordinates. Simultaneous autoregression (SAR) represents a common specification used to model spatial dependence. We employ a more general model that subsumes these two models and thus permits testing of the heterogeneity versus spatial dependence hypotheses. Under the null hypothesis that spatial heterogeneity drives observed spatial dependence, restricting the spatial autoregressive term to zero ($\alpha = 0$) should result in a relatively small increase in the log-likelihood. In contrast, under the null hypothesis that spatial dependence accounts for spatial heterogeneity, the terms interacted with locational coordinates and the locational coordinates themselves should not greatly augment the log-likelihood.[9]

7. If users believe the conservative bias of the likelihood dominance inference has led to an insufficiently low test statistic, they can elect to employ a more accurate approximation. For example, the Barry and Pace (1999) approximation allows continuous control on the trade-off between accuracy and speed. A sufficiently good approximation should minimize any conservative bias from likelihood dominance inference.

8. The coordinates came from a Lambert conformal conic projection of the internal point centroids of latitude and longitude of each county. The projection minimized the error along the 27.5 and 43.5 degree parallels.

9. Anselin (1988, p. 127) notes that Casetti's method assumes a trend surface without error. A trend surface with error would imply heteroskedasticity in the residuals. In our simple illustration, rejection of the null of no spatial heterogeneity is the focus of interest, rather than the exact form of spatial hetero-

To begin, we used the proportion of votes cast for both candidates in the 1980 presidential election as our dependent variable, denoted as ln(*PrVotes*), which also equals ln(*Votes/Pop*), or ln(*Votes*) − ln(*Pop*).[10] The general model uses two different groups of independent variables. Let U represent the 3,107 by 5 matrix formed by combining the four explanatory variables with a vector of ones denoted by **1**.

$$U = [1 \quad \ln(Pop) \quad \ln(Education) \quad \ln(Houses) \quad \ln(Income)]$$

Let S represent a quadratic polynomial in the locational coordinates,[11] where the multiplications are performed elementwise.

$$S = [1 \quad X_c \quad X_c^2 \quad X_cY_c \quad Y_c \quad Y_c^2]$$

The interaction of all columns in U and S defines the 3,107 by 30 matrix of independent variables.

$$X = U*S$$

We then estimate the model in (1) via maximum likelihood,

$$y = X\beta + \varepsilon$$

where $\varepsilon \sim N(0,\Omega(\alpha))$, $\Omega(\alpha) = \sigma^2((I - \alpha D)' (I - \alpha D))^{-1}$, and the spatial weight matrix D is constructed using Delaunay triangles, reweighted so that D is row-stochastic (i.e., $D1 = 1$).[12]

We employed four different types of bounds in computing likelihood dominance statistics. A first set of simple bounds, $0 \leq \alpha \leq 1$ were used, with the signed root deviance statistics produced by this approach labeled SRD01. The second set of bounds, $0 \leq \alpha < \min(1,\hat{\alpha}_{EGLS})$, produce statistics that we label SRDG. Result based on the quadratic bounds proposed by Pace and LeSage (2002) were labeled SRDQ. A final set of bounds were based on the endpoints from 99% confidence intervals produced by the Monte Carlo log-determinant estimator of Barry and Pace (1999).[13] We labeled likelihood dominance statistics produced by this approach SRDM.

As to the bounds proposed by Pace and LeSage (2002), they showed that:

$$(\alpha + \ln(1 - \alpha))tr(D^2) \leq \ln|I - \alpha D| \leq -0.5\alpha^2 tr(D^2) \,,$$

for a symmetric weight matrix D. Of course, $tr(D^2)$ equals the sum of all squared elements for a symmetric spatial weight matrix. Running SAR at both the upper and lower log-determinant bounds yields lower and upper bounds on the autoregressive parameter. Performing this for both models, one where the independent variables are restricted and the other where these are unrestricted, results in lower and upper

---

geneity. Rejection of an imperfect model relative to another imperfect model is in the spirit of likelihood dominance where a superior model may exist, but does not have to be computed explicitly.

10. While the dependent variable is the log of a ratio, the log of the denominator appears on the right-hand side, and thus the fit is exactly the same as a regression using ln (*Votes*) .

11. Due to the squaring, these were standardized for numerical stability.

12. The row-stochastic Delaunay spatial weight matrix D is similar (has the same eigenvalues) as a reweighted symmetric version (Ord 1975). We use the symmetric version for the log-determinant approximations and the row-stochastic version for statistical computations.

13. The Monte Carlo estimator used 10 iterations, exact computation for the first 2 moments and approximate computation for the remaining 28 moments.

bounds on the autoregressive parameter for both restricted and unrestricted parameters. Given these bounds, test statistics for likelihood dominance of the first kind (LD1) can be computed as described previously.

Table 2 contains both exact and likelihood dominance test statistics, converted to the more intuitive signed root deviance (SRD) form. As expected, exact signed root deviances (SRD) exceed or match the various likelihood dominance signed root deviances in all cases. As an indication of the correspondence between the various likelihood dominance statistics and exact likelihood-based results, we computed the median of the ratios for these. The median of SRDM/SRD equals 0.99, that for SRDQ/SRD equals 0.93, whereas SRDG/SRD equals 0.72, and SRD01/SRD equals 0.67.

Using the common t-statistic of 2 as a cutoff, all nineteen cases where exact SRDs exceed 2 correspond to SRDMs statistics greater than 2. Thus, the exact and bounded deviance tests result in qualitatively identical inferences in 100% of the tests. A similar result arises from comparing results based on SRDs and SRDQs. Both methods result in qualitatively identical inferences in 100% of 19 cases. Turning to the SRDG statistics, in 14 of 19 or 73.68% of the cases both exact SRDs and SRDGs exceed 2, resulting in identical inferences. Finally, the SRD01 statistics also result in 14 of 19 cases where exact SRDs and SRDGs produce identical inferences.

TABLE 2
Likelihood Dominance for Election Data

Variables	SRD	SRDM	SRDQ	SRDG	SRD01
$Pop18$	4.97	4.96	4.93	4.45	4.45
Edu	0.57	0.53	0.32	0.01	0.01
$Home$	7.3	7.28	7.09	5.64	5.64
$Income$	0.99	0.97	0.89	0.65	0.6
X_c	1.52	1.49	1.29	0.78	0.64
X_c^2	0.85	0.84	0.76	0.01	0.01
X_cY_c	4.56	4.53	4.2	3.37	2.93
Y_c	0.13	0.12	0.04	0	0
Y_c^2	3.07	3.03	2.69	1.69	1.05
X_cPop18	3.93	3.92	3.83	3.63	3.63
X_c^2Pop18	3.97	3.96	3.87	3.63	3.59
X_cY_cPop18	0.9	0.9	0.89	0.71	0.71
Y_cPop18	3.51	3.5	3.4	2.31	2.31
Y_c^2Pop18	2.89	2.88	2.79	1.84	1.84
X_cEdu	4.07	3.97	3.59	2.56	2.4
X_c^2Edu	6.37	6.29	5.85	4.94	4.79
X_cY_cEdu	2.36	2.34	2.25	1.56	1.56
Y_cEdu	3.57	3.54	3.38	3.01	2.81
Y_c^2Edu	1.89	1.85	1.61	0.92	0.65
X_cHome	3.2	3.16	2.95	1.51	1.51
X_c^2Home	3.67	3.64	3.41	2.79	2.48
X_cY_cHome	10.38	10.38	10.32	9.44	9.44
Y_cHome	7.94	7.93	7.86	7.06	7.06
Y_c^2Home	1.48	1.48	1.48	1.47	1.47
$X_cIncome$	2.39	2.37	2.22	1.82	1.74
$X_c^2Income$	0.04	0.03	0	0	0
$X_cY_cIncome$	4.49	4.47	4.23	3.55	3.12
$Y_cIncome$	1	0.99	0.96	0.61	0.61
$Y_c^2Income$	4.09	4.06	3.84	3.2	2.85
$Intercept$	1.06	1.04	0.94	0.73	0.73

Based on the data examined here, it would seem prudent to use the easily computed quadratic Taylor bounds or the more accurate Monte Carlo log-determinant estimator to form the bounds, as the other techniques do not perform very well. However, these other techniques show how even small restrictions on α can lead to likelihood dominance inference.

This example can also be used to illustrate likelihood dominance of the second kind (LD2). We first note that using SAR on variables in U alone results in an interval for the likelihoods between -5761.28 and -5627.32. Since the log-likelihood from OLS equals -6307.1 for these same variables, we can reject OLS because it falls well outside this SAR interval. We can use likelihood dominance of the second kind to reject the Casetti expansion in favor of SAR based on U or *a fortiori* SAR based on $S{*}U$. To see this, note that the true log-likelihood arising from SAR applied to the Casetti expansion (X) is between -5392.78 and -5332.47. Computing the deviance of the second kind yields a test statistic of $2(-5392.78 - -5627.32) = 469.08$, which is considered in light of the presence of 25 extra variables in X that are not included in U. This results in rejection of the SAR model based on variables U, despite the presence of 25 extra variables in X that are not in U. A Casetti expansion by itself (SAR applied to X with the restriction $\alpha = 0$) has a likelihood of -5697.81, so a user could use likelihood dominance of the second kind to reject the Casetti expansion in favor of SAR based on U or *a fortiori* SAR based on $S{*}U$.

2.2. Consumer Expenditure Data

To illustrate likelihood dominance inference in a different context, we collected data on 51 consumer expenditures by census tract (taken from the 1999 Consumer Expenditure Survey) and matched this with an intercept and 25 explanatory variables from the 1990 Census. We selected typical explanatory variables such age, race, gender, income, age of homes, house prices, and so forth. The 51 dependent variables included observations on alcohol, tobacco, food, entertainment, and a variety of other expenditure categories. Using only records with complete data yielded 54,584 census tract observations. The unprojected internal point latitudes and longitudes served as the locational coordinates.

The total number of possible hypothesis tests associated with deleting a single explanatory variable was 1,352 (51·26) which should constitute a large enough number to provide an indication of the performance of likelihood dominance inference. We estimated a SAR model using a row-stochastic Delaunay spatial weight matrix for each of the 51 dependent variables. Table 3 contains the number of exact signed root deviance statistics whose absolute value exceeded a critical value V^* equal to 2 or 3 along with the proportion relative to an exact test. In addition, Table 3 contains analogous signed root deviance statistics based on likelihood dominance for the Barry and Pace bounds (SRDM), Pace and LeSage quadratic bounds (SRDQ), EGLS bounds (SRDG), and 0 to 1 bounds (SRD01).

TABLE 3
Likelihood Dominance for Consumer Expenditure Data

Test Critical Values	SRD	SRDM	SRDQ	SRDG	SRD01
# exceeding $V^* = 2$	1215	1212	1143	807	786
Proportion exceeding 2 Relative to an exact test	1	0.998	0.941	0.664	0.647
# exceeding $V^* = 3$	1124	1122	1036	705	700
Proportion exceeding 3 Relative to an exact test	1	0.998	0.922	0.627	0.623

For exact SRD statistics 1,215 of 1,352 possible parameters exceeded critical value V^* of 2. For SRDM bounds based on the Barry and Pace Monte Carlo method, likelihood dominance statistics exceeded the same critical value in 1,212 of 1,215 cases (99.8%). Likelihood dominance statistics constructed from the Pace and LeSage quadratic bounds SRDQ resulted in 1,143 or 94.1% of the exact number. Using the likelihood dominance statistics based on EGLS bounds, 807 exceeded the critical value $V^* = 2$, or 66.4% of the exact number. Finally, likelihood dominance statistics based on 0-1 bounds produced 786 significant statistics or 64.7% of the cases as a result. Results based on a critical value $V^* = 3$ showed a similar pattern of outcomes.

We note that these tests based on individual regression parameters are not independent. Hence, this example shows the technique meets a necessary condition (good performance in at least one case), but this is not a sufficient condition for good performance on other data. Nevertheless, the results suggest that likelihood dominance inference may perform well in practice.

2.3. Housing Data

To examine a data set where exact answers might prove tedious or even impossible (for given computing facilities) to produce, we collected 1990 Census observations on white population, black population, mean number of rooms, and mean price of housing from individual census blocks in the continental United States. This resulted in 890,091 complete data observations. As with previous examples, we used a spatial weight matrix based on Delaunay triangles and estimated a SAR model $y = X\beta + \varepsilon$, where y represented the log of mean housing prices, X contained an intercept plus log of white population, black population, mean number of rooms, an as explanatory variables, ε is distributed $N(0,\Omega(\alpha))$, and $\Omega(\alpha)^{-1} = (I - \alpha D)' (I - \alpha D)$.

The least-squares results presented in Table 4 suggest that white and black populations have disparate impacts on mean housing prices. Specifically, log of white population has a coefficient of 0.211 while log of black population has a coefficient of −0.022, a difference of 0.233. Spatial estimates based on the Monte Carlo log-determinant estimator bounds and quadratic bounds show smaller magnitudes for coefficients on all independent variables except the intercept. In terms of the difference between coefficients on log of white population and log of black population, this declines from 0.233 under ordinary least-squares (OLS) to a maximum of 0.073 using the Monte Carlo bounds or 0.086 using quadratic bounds. This suggests that disparate black/white impact falls by more than a factor of three when measured by Monte Carlo bounds relative to OLS, and thus shows the potentially important impact of spatial methods on estimation and inference. The quadratic bounds place α in an interval [0.79, 0.89], while the Monte Carlo bounds place α in the short interval [0.88, 0.88]. Note, the upper and lower determinant bounds lead to an interval estimate for β. For example, the estimated parameter associated with log of rooms fell into the interval [0.5494, 0.5817] for the quadratic log-determinant bounds. Thus, the technique provides both interval estimates as well as lower bounds on the deviance.

TABLE 4
Likelihood Dominance for Housing Data

Variables	OLS	SRD OLS	LBM	UBM	SRDM	LBQ	UBQ	SRDQ
WhitePop	0.2115	401.4510	0.0650	0.0650	176.6333	0.0761	0.0603	166.8463
BlackPop	−0.0220	−44.2715	−0.0147	−0.0147	−43.1965	−0.0189	−0.0128	−38.1505
Rooms	0.7691	236.1927	0.5600	0.5600	250.2206	0.5817	0.5494	248.9154
Intercept	9.1290	1062.8418	9.9600	9.9600	1265.2977	9.9122	9.9699	1258.6837
α	0.0000	0.0000	0.8800	0.8800	1040.0221	0.7900	0.8900	978.0480

The likelihood dominance SRDs are smaller for all but one non-intercept variable than corresponding OLS SRDS, but easily exceed conventional standards for significance.[14] For the mean number of rooms variable, likelihood dominance SRDs (SRDM, SRDQ) are actually higher than the corresponding one for OLS. Since likelihood dominance SRDs are a lower bound, an exact SRD would be higher still.

The SAR log-likelihood lies in the interval −5,240,332.8 to −5,136,523.0 for the quadratic bound, between −5,177,606.01 and −5,176,232.9 for the Monte Carlo log-determinant estimator bounds, while the OLS log-likelihood was −5,718,610.8. Likelihood dominance of the second kind leads to clear rejection of the hypothesis $\alpha = 0$. The SRD statistic in Table 4 is calculated by taking the square root of twice the smallest difference between the interval and OLS log-likelihoods.

It required just under 5 minutes to create the spatial weight matrix based on a Delaunay triangle algorithm, just under 0.6 seconds to compute the quadratic log-determinant approximations, and 5 seconds to perform SAR estimation with likelihood dominance inference.[15] In total, it required just seconds given the weight matrix to estimate a spatial autoregression with 890,091 observations using the quadratic bounds and to perform likelihood dominance inference that, in this case, proved as decisive as results based on exact log-likelihoods. The more accurate Monte Carlo log-determinant approximation required over a minute to compute. Nonetheless, it yielded qualitatively identical inferences as the computationally faster quadratic approximation in this case. Attempts at exact computation of the log-determinant failed, despite the presence of three gigabytes of memory.[16]

3. CONCLUSION

If the desire to conduct maximum likelihood inference on various parameters provides the motivation for estimating a spatial autoregression, paradoxically users may not need to compute actual maximum likelihood estimates to conduct such inference. In many cases users can compute lower bounds to the profile deviance that will exceed selected critical values needed to establish "significance." The virtue of this approach lies in the low effort needed for calculating bounds relative to maximum likelihood.

The development of bounded deviance tests relies on a partitioning of the exact deviance test into two separate deviances, where the first of these depends only on a ratio of two quadratic forms, eliminating the need to compute a log-determinant term. The second deviance depends on a difference in the log-likelihoods between the autoregressive parameter estimate in the unrestricted and restricted models. We show that this second deviance may not have a large magnitude in many cases. For the first deviance, we can use bounds on the autoregressive parameter to narrow the

14. We omitted estimates and SRDs for the 0 to EGLS bounds on the autoregressive parameter and for 0 to 1 bounds on the autoregressive parameter to conserve space. However, both bounds also yielded significant SRDs.

15. The times are for a 1700+ Athlon with three gigabytes of RAM using Matlab 6.5 on W2K. We used function fdelw2.m to generate the Delaunay triangle based weight matrices, the function fsarld2.m to compute SAR likelihood dominance inference and point estimates, and the function fdet_mc2.m to calculate the Monte Carlo log-determinant approximation with confidence bounds. All of these can be found in the Spatial Statistics Toolbox 2.0 at http://www.spatial-statistics.com.

16. One can extrapolate the exact results. As Smirnov and Anselin (2001) noted, the exact sparse matrix methods proposed by Pace and Barry (1997) show an approximate quadratic increase in time with n. For the 3,107 data set it required 0.64 seconds to compute the log-determinants. For 890,091 observations the time would go to 14.5 hours, based on the quadratic relation with n. Computing the eigenvalues for the 3,107 observation data set required 377.23 seconds. Given the cubic increase in time for computing the eigenvalues of a dense matrix, a straightforward computation of log-determinants via the eigenvalues would take over 280 years, assuming one could avoid the memory constraints.

possible range of values taken by this statistic. By selecting the lowest value of the statistics, we arrive at an implementable lower bound to the actual profile deviance. This procedure avoids the necessity of computing the log-determinant term that impedes computation of maximum likelihood estimates. The lower bound on the deviance constructed in this fashion represents a form of "likelihood dominance" (Pollack and Wales 1991).

We compared exact inference and likelihood dominance inference results using three empirical examples. In the first example, based on county-level election data, bounded inference techniques yielded the same qualitative inferences as exact techniques in the vast majority of cases, when using a common criterion of counting an estimate with a t-statistic of 2 as significant. This result held true for bounds based on two approaches, the quadratic bounds of Pace and LeSage (2002) and the Monte Carlo bounds of Barry and Pace (1999). A second example was constructed from a spatial data sample of 51 consumer expenditure categories and 54,584 census tract observations. Here the bounded inference technique (based on the common criterion of a t-statistic equal to 2 as significant) yielded results identical to those from exact techniques in 99.8% (Monte Carlo) and 94.1% (quadratic) of the 1,352 possible cases.

A third illustration examined housing at the census block level using a data set with 890,091 observations that would pose a computational challenge to deviance tests based on exact methods. The likelihood dominance approach to inference required under a minute to compute bounded inference when using quadratic bounds (given the weight matrix). The Monte Carlo technique for determining bounds required less than an additional minute. In both cases, bounded inference showed which variables were significant and provided bounds on the parameter values as well.

The likelihood dominance approach to inference has a number of potential uses. First, a mathematical relation exists between deviances and confidence intervals, so the technique could lead to the construction of confidence intervals. Second, the technique can provide likelihood-based inference for estimators that have either no inferential framework or have only an approximate inferential framework. Examples include the pseudo-likelihood estimator of the conditional autoregressive model of Besag (1975), or the instrumental variable estimators of Kelejian and Prucha (1998). One could use these estimators to compute point estimates, and rely on likelihood dominance for inference. Third, the ability to handle large sample sizes suggests that likelihood dominance inference would be useful in spatial data mining applications. Fourth, simplicity of the code facilitates its incorporation within GIS or other types of interactive software. Finally, the speed of the technique promotes exploration of different model specifications and thus has potential to improve the robustness of inferences drawn from spatial data.

LITERATURE CITED

Anselin, Luc (1988). *Spatial Econometrics: Methods and Models*. Dorddrecht: Kluwer Academic Publishers.

Barry, Ronald, and R. Kelley Pace. (1999). "A Monte Carlo Estimator of the Log Determinant of Large Sparse Matrices." *Linear Algebra and its Applications*, 289, 41–54.

Bavaud, Francois (1998). "Models for Spatial Weights: A Systematic Look." *Geographical Analysis*, 50, 155–71.

Besag, Julian (1975). "Statistical Analysis of Non-lattice Data," *The Statistician*, 24, 179–95.

Casetti, Emilio (1972). "Generating Models by the Expansion Method: Applications to Geographical Research." *Geographical Analysis*, 4, 81–91.

Chen, Jian-Shen, and Robert Jennrich (1996). "The Signed Root Deviance Profile and Confidence Intervals in Maximum Likelihood Analysis." *Journal of the American Statistical Association*, 91, 993–98.

Fan, Jianqing, Hui-Nien Hung, and Wing-Hung Wong (2000). "Geometric Understanding of Likelihood Ratio Statistics." *Journal of the American Statistical Association*, 95, p. 836–41.

Golub, Gene H., and Urs Von Matt (1995). "Generalized Cross-Validation for Large Scale Problems." *Journal of Computational and Graphical Statistics*, 6, 1–34.

Griffith, Daniel, and A. Getis (2002). "Comparative Spatial Filtering in Regression Analysis." *Geographical Analysis*, 34, 130–40.

Griffith, Daniel, and F. Lagona (1998). "On the Quality of Likelihood-based Estimators in Spatial Autoregressive Models When the Data Neighborhood Structure is Misspecified." *Journal of Statistical Planning and Inference*, 69, 153–74.

Griffith, Daniel, and Akio Sone (1995). "Trade-offs Associated with Normalizing Constant Computational Simplifications for Estimating Spatial Statistical Models." *Journal of Statistical Computation and Simulation*, 51, 165–83.

Haining, Robert (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.

Kelejian, Harry, and Igmar Prucha (1998). "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." *Journal of Real Estate Finance and Economics*, 17, 99–121.

Leamer, Edward E. (1988). "Model Choice and Specification Analysis." In *Handbook of Econometrics, Volume 1*, edited by Griliches and Intrilligator. Amsterdam: North-Holland.

Ord, J. K. (1975). "Estimation Methods for Models of Spatial Interaction." *Journal of the American Statistical Association*, 70, 120–26.

Pace, R. Kelley, and Ronald Barry (1997). "Quick Computation of Regressions with a Spatially Autoregressive Dependent Variable." *Geographical Analysis*, 29, 232–47.

Pace, R. Kelley, and James LeSage (2002). "Semiparametric Maximum Likelihood Estimates of Spatial Dependence." *Geographical Analysis*, 34, 75–90.

Pace, R. Kelley, and James LeSage (2003). "Chebyshev Approximation of Log-determinants of Spatial Weight Matrices." Forthcoming in *Computational Statistics and Data Analysis*. Elsevier Science.

Pollack, R. A., and T. J. Wales (1991). "The Likelihood Dominance Criterion: A New Approach to Model Selection." *Journal of Econometrics*, 47, 227–42.

Ramsey, J. (1974). "Classical Model Selection Throughthrough Specification Error Tests." In *Frontiers in Econometrics*, edited by P. Zarembka. Academic Press

Smirnov, Oleg, and Luc Anselin (2001). "Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: A Characteristic Polynomial Approach." *Computational Statistics and Data Analysis*, 35, 301–19.