Some Fast Methods for Fitting Some One-parameter Spatial Models

R. J. Martin

Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand

Abstract: It is common in geographic modelling to use a one-parameter spatial model to specify the inverse covariance matrix in terms of I-βW, for some known matrix W. Exact Gaussian maximum likelihood estimation of β requires evaluation of the determinant of the covariance matrix. For large data sets, this evaluation of the determinant can be slow and good approximations can be useful. Seventy regional configurations are used to consider some approximations to the determinant of I-βW that are fast to evaluate, and their usefulness is compared.

Key words: Conditional autoregression, geographical spatial models, simultaneous autoregression

INTRODUCTION

In much geographical spatial regression modelling it is assumed that for n sites or regions the n-vector of observations y has mean E(y)=Xθ for an n×p matrix of regressors X and a p-vector parameter θ; and has covariance matrix Vσ^2=var(y), where σ^2 is a scale parameter. Usually, one-parameter spatial models, with V depending on T=I_n-βW for a known weights matrix W (e.g. [1]) and Gaussian maximum likelihood estimation, are used.

Then $\hat{\theta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$, where $\hat{V}$ uses $\hat{T}$ = I_n-$\hat{\beta}$W. Exact estimation of β (e.g. [2]) minimizes with respect to β the profile likelihood $|V^{-1}|^{-1/n}\times(e'V^{-1}e)$, or its logarithm, where e=y-X$\hat{\theta}$. Iterative estimation of θ and β is usually needed. This requires numerous evaluations of the quadratic form e'V^{-1}e and the determinant $|V^{-1}|$. Neglecting this determinant term can lead to inconsistency in the estimator of β[3] and serious bias for finite n. If p/n is not small, REML estimation may be better (e.g. [4]), where θ is estimated as above, but β is estimated by minimizing $\{|X'V^{-1}X|/|V^{-1}|\}^{1/(n-p)}\times(e'V^{-1}e)$. This additionally requires the evaluation of $|X'V^{-1}X|$.

The simultaneous autoregression (SAR) model has V^{-1}=T'T, so $|V^{-1}|=|T|^2$. The conditional (CAR) form, with conditional variance var($y_i|\{y_j, j\neq i\})=v_i\sigma^2$ for known $\{v_i\}$, has V$^{-1}=\{D_i(v_i^{-1})\}$T, where $D_i(s_i)$ denotes diag($s_1,...,s_n$), and so $|V^{-1}|=\{\Pi v_i^{-1}\}\times|T|$. In both cases, e'V^{-1}e is easy and quick to calculate. For the SAR, estimation of θ, for a given β, can use ordinary least squares on the transformed model for which Ty has mean TXθ. Unless p is large, $|X'V^{-1}X|$ can be calculated quickly. If necessary for the CAR, $|X'V^{-1}X| =|X' \{D_i (v_i^{-1})\} X|\times$

$|I-\beta(X'\{D_i(v_i^{-1})\}X)^{-1}X'\{D_i(v_i^{-1})\}WX|$ can be used, where the second term can be calculated as below using the eigenvalues of $(X'\{D_i(v_i^{-1})\}X)^{-1}X'\{D_i(v_i^{-1})\}WX$. Thus, the most time-consuming element in calculating the log profile likelihood is usually the evaluation of $|T|$, considered here in the form J(β)=$ln(|T|^{-1/n})=$ -n$^{-1}ln|T|$.

Let the eigenvalues of W be $\{\lambda_i(W)\}$, or $\{\lambda_i\}$ when it is clear what W is, and let $\lambda=\max_i\{\lambda_i\}$ and $\lambda_{min}=\min_i\{\lambda_i\}$. Then J($\beta$) can be found using the $\{\lambda_i\}$ since $|T|=\Pi(1-\beta\lambda_i)$ and J(β)=-n$^{-1}\Sigma ln(1-\beta\lambda_i)$. However, the $\{\lambda_i\}$ need to be found, which can be very slow for large n and each evaluation of J(β) requires an n-fold product, or, more accurately but slower, the sum of n log terms.

Some quicker numerical methods for obtaining J(β) exactly have been discussed by Martin[5] and Pace and Barry[6,7]. However, good quick approximations to J(β) should suffice in preliminary investigations, for instance when different sets of regression variables and different forms of the spatial dependence, are being considered. Griffith[8] considered approximations to J(β) using a specified functional form. Apart from a postulated general form, those in Griffith[8] require the $\{\lambda_i\}$ to be found first, or $|T|$ evaluated for a set of values of β. Martin[5] considered some approximations based on using the first few terms of the series expansion of J(β). The 4th-order approximation is

$$J(\beta) \approx \frac{1}{n}\sum\nolimits_{k=1}^{4}\{tr(W^k)\beta^k / k\}, \qquad (1)$$

where tr(W) denotes the trace of W. Martin[9] (M05 henceforth), looked at the theoretical results available for known types of graphs. Some comparisons of the timings of some different approximations are in Griffith and Sone[10]. Berry and Pace[11] give a Monte Carlo

Corresponding Author: R. J. Martin, Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand

method for estimating $J(\beta)$ for a sparse stochastic matrix W.

Another possibility for approximating $J(\beta)$ is to regard the $\{\lambda_i\}$ as a discrete probability distribution on $[\lambda_{min}, \lambda]$ of the random variable Λ, with a probability of r_i/n at each different value λ_i (where each λ_i is repeated r_i times) and then to approximate this discrete distribution by a continuous distribution (of Λ^*) with probability density function (pdf) $f(x)$ on $[\lambda_{min}, \lambda]$. Then, $J(\beta)$ is the average value of $-ln(1-\beta\lambda_i)$, which is approximated by $J^*(\beta)$, the expected value of $-ln(1-\beta\Lambda^*)$,

$$J^*(\beta) = - \int_{\lambda_{min}}^{\lambda} \ln(1-\beta x)f(x)dx .$$

The use of $J^*(\beta)$ for a suitable approximate pdf $f(x)$ may produce worthwhile approximations for geographical regions. Since approximate results may suffice, the integral can itself be approximated. Various simple numerical rules of the form $K(\beta)=$ $-\Sigma w_i\, f(x_i)\, ln(1-\beta x_i)$ could be used (for example, §10 of Fröberg[12]). Griffith's[13] approximation of the form $w_1 g(\lambda)ln(1-\beta\lambda)-w_2 g(\lambda_{min})ln(1-\beta\lambda_{min})$ with weights w_1, w_2, can, if $f(x)$ is finite at the endpoints and $g(x)=f(x)$, be regarded as a 2-point rule using the endpoints (or a 3-point including x=0). That $J^*(\beta)$ need not be of the form $g(\lambda)ln(1-\beta\lambda)-g(\lambda_{min})ln(1-\beta\lambda_{min})$, for a function $g(x)$, can be seen for equidistant points on a line (the path graph P_n - M05, §4.2), for which $J(\beta)$ tends to $-ln[\{1 + (1-a^2\beta^2)^{1/2}\}/2]$ as $n\to\infty$ (a=4 or 1 for W a contiguity or standardized contiguity matrix, respectively).

As well as the CAR and SAR models, other one-parameter assumptions using T could be tried. These include having V=T (T symmetric), or V=T'T. Provided $|\beta|$ is not too large, T^{-1} can be approximated by, for example, $I-\beta W + \beta^2 W^2$, to give a quick approximation to $e'V^{-1}e$.

In this paper, seventy readily available regional configurations are used to suggest and compare some possibilities for $K(\beta)$, with particular interest in the shape of the eigenvalue distributions. The W matrices to be considered are defined in section 2. The configurations are given in section 3, their statistical properties described in section 4, and their eigenvalue properties discussed in section 5. Section 6 gives the possible approximations, and section 7 contains two examples.

PROPERTIES OF GEOGRAPHICAL ONE-PARAMETER PROCESSES

Assume, as usual, that W has zero diagonal elements, so that $\Sigma\lambda_i=0$ and non-negative off-diagonal elements $w_{i,j}$ which reflect the dependency between regions i and j. Often only a small proportion of the $w_{i,j}$ are non-zero.

Here, for simplicity, the two main forms of W which are frequently used in practice, are concentrated on (other W may be better for a given data set). The binary symmetric *contiguity* or *adjacency* matrix, denoted C, has $c_{i,j}=1$ if sites (or regions) i and j are connected in some sense, assumed here to be adjacency or sharing a common boundary, and 0 otherwise. Let c be the vector of row sums of C, $C1_n=c$, where 1_n is an n-vector of ones and let $D=D_i(c_i)$. Then the asymmetric *row standardized* form of C is $R=D^{-1}C$. The C matrix is equivalent to a graph, with c_i the vertex degree. The eigenvalue (spectral) properties of C and R have been studied theoretically - see M05.

For four configurations the row-standardization U of a symmetric weight matrix G is also considered, where the weights are based on inter-centroid distances. Both R and U have all row sums equal to 1, $W1_n=1_n$ and so are *stochastic* matrices. Their eigenvalues are real and $\lambda(W)=1$. For all of C, R, U, $|\lambda_{min}(W)| < \lambda(W)$, unless the graph is bipartite (M05), i.e. the only edges are between two disjoint sets, which is very unlikely for two-dimensional geographical configurations (except for rectangular grids).

For the one-parameter CAR, $\{D_i(v_i^{-1})\}W$ must be symmetric, so that W must be asymmetric with $w_{i,j}v_j=w_{j,i}v_i$ unless the v_i are constant[14]. Although it is usually preferable that the conditional variance $v_i\sigma^2$ is smaller for interior regions, geographic modelling often uses W=C, implying the v_i are constant. If W=R is used, $var(y_i|\{y_j, j\neq i\})$ must be σ^2/c_i, which seems more appealing. Then $v_i=c_i$, $V^{-1}=D-\beta C=D(I_n-\beta R)$ and $|V^{-1}|=\{\Pi c_i\}\times|I_n-\beta R|$. Using W=U is similar. Since T must be positive definite, β must lie between $1/\{\lambda_{min}(W)\}$ and $1/\{\lambda(W)\}$. For the one-parameter SAR, W does not need to be symmetric or positive definite and any of C, R, U can be used. Often β is restricted to the CAR range to avoid $|T|=0$ when $\beta=1/\{\lambda_i W\}$.

THE REGIONAL CONFIGURATIONS

The possible relevance of the exact and approximate results in M05 is considered and the distribution of the eigenvalues for geographic regions is now investigated. The 70 regional configurations which are used to consider approximations to $J(\beta)$ are listed,

with their abbreviations, in Appendix 1. They were chosen by their availability and may not be representative of all regional sets.

The exact and approximate results on graphs that are available may be very limited for geographical applications because of the structures of the graphs, or the small value of n, or the number of badly-connected regions (end-vertices). Because of the nature of political regions, no precise general results are possible.

Forty-four of the configurations are census tracts for enumeration districts in Canadian cities in 1971 (27 cities) or 1986 (18 cities) (compiled using census tract maps from Statistics Canada). Another 12 configurations are formed from the 73 municipios of Puerto Rico and its subregions (the five agricultural administrative regions and six adjacent pairings of these). The remainder are for China (provinces); Henan province in China (counties); Eire (counties, Cliff and Ord[15]); Ghana (subregional units, Cliff and Ord[15], Fig. 8.4); Nigeria (subregional units, Cliff and Ord[15], Fig. 8.5); plus Columbus, Ohio (neighborhoods, Anselin[16], Fig. 12.1, p. 188); Houston (census tracts, US Bureau of the Census, 1980); South-Western Ohio (counties, Anselin[16], Fig. 12.2, p. 204); three sets using Syracuse (US Bureau of the Census, 1990)-Syracuse:CT (143 census tracts), Syracuse:BG (447 block groups), Syracuse:CB (7249 census blocks); Buffalo police precincts (Annual report of the Buffalo Police Department, 1978) and Thiessen polygons formed from weather stations in part of Kansas-Nebraska[17].

Most configurations have relatively few regions (between 11 and 53). There are a few larger ones (71, 73, 76, 77, 79, 99, 101, 101, 143, 192), four much larger (351, 363, 447, 731) and one extremely large (7249). There are C and R matrices are for all 70 configurations. Four of the configurations (Eire, Houston, Ottawa-Hull:86, Puerto Rico) also have U matrices.

Henan province has some unusual properties. It has five city counties, each completely within another county. It has three outlying counties, each only adjacent to one other. It also has two 'peninsulas' (attached path graphs), one of two counties and one of three.

Note that some common boundary lengths in the mapped configurations are very small and may be due to inaccuracies in the diagram that has been used. Conversely, a true adjacency with a small common boundary may not be apparent in the diagram. For this exercise it should not be important whether all adjacencies are correct or not.

SIMPLE DESCRIPTIONS OF THE CONFIGURATIONS

The properties of the 70 regional configurations are summarized and compared with those for the graphs in M05. Several ways of summarising the graphs and the spectra, were discussed in M05. Note that some of the graphs in M05 are bipartite (§§3.2, 3.3, 3.7, 3.9, 3.11) but, as noted earlier, this would be unusual for geographical regions. Two descriptive measures of graphs are used in M05. The *index of (planar) connectedness* is $I_C = \{1_n'C1_n - 2(n-1)\}/\{2(2n-5)\}$, which ranges from 0 for a tree (a minimally connected graph) to 1 for a triangulation (a maximally-connected planar graph). The *index of regularity*, which can be defined as the sample standard deviation of the degrees (number of neighbours), is $I_R = \sqrt{\{c'c/n - (c'1_n/n)^2\}}$. The values of I_C and I_R for the 70 configurations are given in Appendix 1.

Other functions of the degrees which may be of interest are the extremes $\min_i\{c_i\}$ and $\max_i\{c_i\}$, the range $r(c) = \max_i\{c_i\} - \min_i\{c_i\}$ and the standardized third and fourth moments $b_1(c) = m_3(c)/\{m_2(c)\}^{3/2}$, $b_2(c) = m_4(c)/\{m_2(c)\}^2 - 3$, where $m_r(c) = n^{-1}\Sigma(c_i - 1_n'c/n)^r$, with $m_2(c) = I_R^2$. Note that in the following descriptions of sample distributions, the mean and standard deviation (sd) are used as measures of centrality and spread respectively when there are no large 'outliers'. Otherwise, the median and inter-quartile range (IQR) are used, respectively.

Index of (planar) connectedness I_C: I_C ranges from 0.348 (PR4) to 0.971 (C86:1), with mean 0.712 and sd 0.125. Only 7 values are below 0.50. This suggests that results in M05 for trees and other graphs with low connectedness may not be appropriate for irregular geographical regions. Also, the very high connectedness of some graphs in M05 rarely occurs in practice.

Index of regularity I_R: I_R ranges from 0.877 (C71:4) to 2.402 (O12), with mean 1.414 and sd 0.305. There is a suggestion of a small increase in the mean of I_R with n. Using a log-log regression with and without Syracuse:CB (n=7249) suggests that I_R approximately increases with $n^{1/8}$. Of the graphs in M05, only the Fisher (§3.9.4) and Reduced Fisher (§3.9.5) graphs are consistent with a constant, but non-zero I_R, whilst none are consistent with a very slight increase with n.

Simple functions of the degrees: The $\min_i\{c_i\}$ is 1, 2 or 3, with mean 1.93 and sd 0.688. Apart from four large values of 17 (O11), 23 (C86:11), 26 (O12, C86:23), $\max_i\{c_i\}$ ranges from 5 to 13, with median 8 and IQR 2. The value of r(c) ranges from 3 to 12, apart from the four large values of 16 (O11), 22 (C86:11), 25 (O12, C86:23), with median 6 and IQR 2.25.

The standardized third and fourth moments: The value of $b_1(c)$ ranges from -0.846 to 3.554, with median 0.313 and IQR 0.564. There are 13 negative values. All

values are below 1.63 except for: 2.447 (O12), 2.593 (C71:6), 2.737 (C86:11) and 3.554 (C86:23).

The value of $b_2(c)$ ranges from -1.087 to 27.490, with median -0.161 and IQR 1.248. There are 40 negative values. There are four very large values: 9.905 (O12), 13.699 (C71:6), 16.338 (C86:11) and 27.490 (C86:23), with the remainder being below 2.57.

THE EIGENVALUE DISTRIBUTIONS

The above measures relate to the graph, the C matrix and the degrees. Here, the distribution of the eigenvalues of W is investigated and some measures of the distribution considered. These measures may be useful in themselves, for example to characterise different types of geographical structures. However, the distribution may also be useful if it can be approximated to give an approximation to the required determinant term, as discussed later.

Of particular interest in approximating the distribution of the eigenvalues are the extreme values $\lambda(C)$, $\lambda_{min}(C)$ and $\lambda_{min}(R)$. Simple descriptions of the distribution of the eigenvalues are given by the low-order moments and standardised moments - the measures of skewness and kurtosis. Recall that $\Sigma\lambda_i^k = tr(W^k)$. Since $tr(W)=0$, the mean is always 0 and the formulae for the moments are simplified, $m_r(W)=n^{-1}tr(W^k)$ where $m_r(W)$ denotes $m_r(\{\lambda_i(W)\})$. The variance is $m_2(W)=tr(W^2)/n$. The standardised third moment, $b_1(W)=n^{1/2}tr(W^3)/\{tr(W^2)\}^{3/2}$, describes the skewness. The standardised fourth moment, $b_2(W)=ntr(W^4)/\{tr(W^2)\}^2-3$, describes the kurtosis of the distribution. Some other intuitively appealing characteristics are also considered. These include the ratio of the extreme eigenvalues $r_{mm}(W)=-\lambda_{min}(W)/\lambda(W)$, and the ratio of the number of positive to negative eigenvalues $r_{pn}(W)$. The 70 values of $\lambda(C)$, $-\lambda_{min}(C)$, $b_1(C)$, $b_2(C)$, $-\lambda_{min}(R)$, $b_1(R)$ and $b_2(R)$ and the four $-\lambda_{min}(U)$, $b_1(U)$ and $b_2(U)$, are given in Appendix 1.

Histograms: Firstly, consider the distributions for each of the 70 C and R matrices and the four U matrices. For each of the 144 distributions, a histogram, using equal-width classes (usually 10), of the eigenvalues was formed. These were also smoothed, using a 3-point moving average (3-pt MA) with weights $(1, 2, 1)/4$ and kernel density estimates (see later) were considered. For n small it is difficult to determine the smoothed behaviour of the eigenvalues, particularly at the two extremes. All histograms suggest a positive skew distribution and none demonstrate any of the pathological behaviour that can occur with some of the graphs in M05. Most histograms for larger n are unimodal, but quite a number of the histograms, especially for R, suggest another possible mode at an intermediate positive value. Most histograms suggest something like a typical Gamma-distribution shape (with shape parameter bigger than 1). Some histograms suggest a non-negligible non-zero density at either or both of the extremes.

As examples, smoothed (3-pt MA) histograms for Winnipeg:71 (n=101), Syracuse:CT (n=143) and Syracuse:CB (n=7249) are shown in Fig. 1, 3, 5 (C) and 2, 4, 6 (R), respectively.

Extreme eigenvalues: The largest values of $\lambda(C)$ are around 7 (7.375 for O12, 7.084 for C86:23, 6.822 for C86:11) and the smallest values are around 3.5 (3.373 for PR4, 3.686 for PR1). The median value is 5.09 and the IQR is 0.862. As for I_R, there is evidence that $\lambda(C)$ slightly increases in general with n.

The range of $\lambda_{min}(C)$ is even smaller, with the lowest around -5 (-5.025 for O12, -4.080 for C86:11, -3.563 for O7) and the highest around -2 (-2.120 for PR1, -2.125 for PR4, -2.199 for PR3). The median value is -2.81 and the IQR is 0.497. Again, $|\lambda_{min}(C)|$ appears to slightly increase in general with n.

The smallest values of $\lambda_{min}(R)$ are around -0.9 (-0.936 for O12, -0.845 for C71:5, -0.798 for C71:6 and O4) and the highest around -0.5 (-0.510 for C86:7, -0.524 for C71:24, -0.526 for C86:1). The median value is -0.63 and the IQR is 0.119.

In M05, it is noted that the second largest eigenvalue $\lambda_{(2)}(C)$ may be of interest. Tiefelsdorf and Boots[18] show that $n\lambda_{(2)}(C)/1_n'C1_n \approx \lambda_{(2)}(C)/\lambda(C)$ is approximately the largest possible value of Moran's I. For many of the graphs in M05, there is a large gap between $\lambda(C)$ and $\lambda_{(2)}(C)$, but this occurs for none of the 70 configurations. For small n, $\lambda_{(2)}(C)/\lambda(C)$ or $\lambda_{(2)}(R)$ can be around 0.7, but for the larger n they are usually at least 0.95, with usually $\lambda_{(2)}(R) > \lambda_{(2)}(C)/\lambda(C)$.

Variance: Since $tr(C^2)/n$ is just $1_n'C1_n/n=1_n'c/n$, the C-variance is related to the overall connectedness and I_C. It ranges from 3.00 (PR4) to 5.45 (Syracuse:CT) with mean 4.57 and sd 0.542.

The R-variance uses $tr(R^2)$, the sum over neighbouring vertices (i,j) of $1/(c_ic_j)$. It ranges from 0.182 (C86:8) to 0.329 (PR4), with mean value 0.225 and sd 0.032. The four U's have three very small values (0.121, 0.130, 0.164) and one larger value (0.275).

Transformed mean and variance: Transforming from the interval (λ_{min}, λ) to the standard interval $(0, 1)$, the mean (m) for the seventy C matrices ranges from 0.323 to 0.405, with mean 0.358; and the variance ranges from 0.032 to 0.099 with mean 0.072. The mean for the seventy R and four U matrices ranges from 0.338 to

0.483, with mean 0.392; and the variance ranges from 0.045 to 0.124 with mean 0.082.

Skewness: Since $tr(C^3)/6$ is the number of triangles in the graph and R can be regarded as a probability transition matrix on the graph, both $b_1(C)$ and $b_1(R)$ are non-negative. They are zero if the graph contains no triangles, which includes all bipartite graphs. Otherwise, the distribution must be positive skew, with $|\lambda_{min}| < \lambda$.

The range of $m_3(C)$ is from 3.43 (PR4) to 11.10 (C86:1) with mean 7.34 and sd 1.63. The standardized $b_1(C)$ ranges from 0.509 (C71:6) to 0.910 (C86:1) with mean 0.750 and sd 0.082.

The range of $m_3(R)$ is from 0.038 (O12) to 0.115 (PR4) with mean 0.068 and sd 0.0135. The standardized $b_1(R)$ ranges from 0.401 (C71:6) to 0.837 (C86:1) with mean 0.638 and sd 0.0915.

The four values of $m_3(U)$ are 0.029, 0.029, 0.043, 0.049 and those of $b_1(U)$ are 0.342, 0.618, 0.644, 0.694.

All distributions are moderately positively skew, with the distribution for C always more skew than that of R, i.e. $b_1(C) > b_1(R)$. The difference $b_1(C)-b_1(R)$ ranges from 0.036 (C86:4) to 0.267 (O6), with mean 0.112 and sd 0.048. The ratio $b_1(C)/b_1(R)$ ranges from 1.052 (C86:19) to 1.443 (O6), with mean 1.184 and sd 0.097.

Ratio of minimum to maximum eigenvalues: The ratio $r_{mm}(C)$ ranges from 0.477 (C86:1) to 0.681 (O12) with mean 0.559 and sd 0.049. It is smaller than $r_{mm}(R)=-\lambda_{min}(R)$ in all but two cases (C71:7, C86:7). The difference $-\lambda_{min}(R)-r_{mm}(C)$ ranges from -0.071 to 0.255, with one exceptional value of 0.398 (O6). This is caused by the two 'peninsulas' in Henan province (M05). Without Henan, its mean is 0.085 and its sd is 0.061. The two measures $r_{mm}(C)$ and $-\lambda_{min}(R)$ are reasonably well correlated (0.685 without Henan).

Ratio of numbers of positive to negative eigenvalues: The ratio $r_{pn}(W)=\{$number of $\lambda_i(W)>0\}/\{$number of $\lambda_i(W)<0\}$ is the same for C and R (M05), but differs for U. It is 1 for bipartite graphs and can exceed 1 for some graphs (M05, §5.3.2). For C and R, it ranges from 0.556 (C71:1) to 0.889 (C71:21) with mean 0.702 and sd 0.071. It is only fairly lowly correlated with $r_{mm}(C)$ (0.372) and $-\lambda_{min}(R)$ (0.318). It is unusual for a geographical configuration to have any zero eigenvalues. This occurs in just 5 cases: China, Columbus, Henan and Winnipeg:71 have 1 (China, Columbus and Winnipeg:71 each have 2 regions with the same neighbours, so C and R have two rows the same (M05, §4.1); for Henan, one region has the same neighbours as two others combined and these two have no neighbours in common, so one row of C is the sum of two others), but Syracuse:CB has 70.

Kurtosis: Large values of $b_2(W)$, which is always at least -2, are associated with a large number of eigenvalues concentrated in a small range (e.g. many near or equal to 0) and a few (even just 1) eigenvalues separated from the clump.

M05 notes that $b_2(C)$ is highly related to I_R. The range of $m_4(C)$ is from 19.00 (PR4) to 83.50 (C86:1) with mean 54.95 and sd 14.45; while $b_2(C)$ ranges from -0.889 (PR4) to 0.037 (O3) with mean -0.388 and sd 0.164. Just one $b_2(C)$ is positive.

The range of $m_4(R)$ is from 0.080 (O13) to 0.203 (PR4) with mean 0.114 and sd 0.026; while $b_2(R)$ ranges from -1.128 (PR4) to -0.397 (C86:1) with mean -0.751 and sd 0.140. The four $m_4(U)$ are 0.043, 0.045, 0.068, 0.150 and the $b_2(U)$ are -1.017, -0.478, -0.336, -0.092.

Almost all distributions had negative kurtosis. The kurtosis is always larger for the C distribution than that of R, i.e. $b_2(C)>b_2(R)$. The difference $b_2(C)-b_2(R)$ ranges from 0.163 (C71:10) to 0.697 (O3), with mean 0.363 and sd 0.118. The ratio $b_2(C)/b_2(R)$ ranges from -0.056 (O3) to 0.788 (PR4), with mean 0.504 and sd 0.178, with only one other value below 0.15 (C86:1).

USING AN APPROXIMATE EIGENVALUE DISTRIBUTION TO APPROXIMATE J(β)

Introduction: Although the eigenvalues $\{\lambda_i(W)\}$ themselves may be of interest in some applications, the aim here is to approximate $J(\beta)$. Approximations may be possible which are quite coarse, but still sufficiently good to allow reasonably accurate estimates of β to be found.

Two different situations can be envisaged with different levels of knowledge. In the first, all the eigenvalues $\{\lambda_i(W)\}$ are known and just a quick way of approximating $J(\beta)$ is required. In the second, there may originally be no eigenvalue information and a good estimate of $J(\beta)$ is required. Since usually there are good quick approximations to $\lambda_{min}(W)$ and $\lambda(W)$, they can be assumed known (see later). The second case is perhaps the most realistic and would apply when fitting exploratory regression models with a large data set where spatial dependence is possible but not certain. If the fast method of fitting suggests spatial dependence, more accurate methods may then be worthwhile.

Assume then that it is reasonable to approximate the distribution of the eigenvalues by a pdf over the range $[\lambda_{min}, \lambda]$. Extensions are possible to cases with a few isolated eigenvalues. In particular, n should be sufficiently large, at least 30, say. For smaller n, there are few problems in evaluating $J(\beta)$ either directly, or by using the eigenvalues of W.

All the eigenvalues $\{\lambda_i(W)\}$ known: In this case, the continuous pdf approximation to the eigenvalue

distribution can be estimated by histograms (as earlier), or more sophisticated methods of density estimation. Particular families of distributions could be used, for example the Beta or Gamma distributions which are discussed later.

A simple possibility if a numerical integration rule is to be used is to use a hand drawn (or distribution-free estimated) density to estimate its values at the few points needed.

At most $\lambda_{\min}(W)$ and $\lambda(W)$ known: If $\lambda_{\min}(W)$ and $\lambda(W)$ are not known, it is usually possible to get very good approximations to them quite quickly. For R and other row-standardized W, $\lambda=1$. For C, the power method $x_{(r)}=Cx_{(r-1)}$ (M05, §4.1.2) gives upper and lower bounds for $\lambda(C)$ and usually converges quickly from $x_{(0)}=1_n$ (with an improved lower bound from the Rayleigh quotient, $\lambda_{(r)}=x_{(r)}'Cx_{(r)}/x_{(r)}'x_{(r)}$, M05, §4.1.2). Some simple lower bounds are often good approximations (M05, §4.1.2). It is not usually quite so easy to approximate $\lambda_{\min}(W)$. One method is the power method on $W-aI_n$, for some a, with the initial vector having some negative elements, e.g. the alternating vector. The value of a>0 must be large enough for -$\{\lambda_{\min}(W) + a\}$ to be the largest of the $\{|\lambda_i(W)-a|\}$, but otherwise as small as possible. Assuming $\lambda_{\min}(W) <$ -$\lambda(W)/2$, a reasonable value for a is $\lambda(W)/4$. Unless $\lambda_{\min}(W)$ is a repeated eigenvalue, this method will converge to $\lambda_{\min}(W)-\lambda(W)/4$, but convergence will be slow if, as is usual, there is another eigenvalue close to $\lambda_{\min}(W)$. It can be better to use $W-aI_n-\lambda(W)x(W)x(W)'$, where $x(W)$ is the eigenvector corresponding to $\lambda(W)$. For $\lambda_{\min}(R)$, it may be better to use the symmetric $R^*=D^{-1/2}CD^{-1/2}$, which has the same eigenvalues as R. An alternative method, which usually converges quicker, but needs a matrix inverse, is to use the power method on $(W+\lambda I_n)^{-1}$. Another method, usually slower, is to use a numerical routine to minimise $x'Cx/x'x$.

Assume now that λ and $\lambda_{\min}$ have been determined, or well approximated. A possible method for using $J^*(\beta)$ is to approximate the pdf by using an appropriate family whose parameters can easily be estimated. Families which may have the right shape include the Beta, Gamma and lognormal distributions. None of these would lead to a simple theoretical integral for $J^*(\beta)$.

Two possible strategies would be either to calculate the means of the n order statistics for the fitted distribution and use these as approximations to the λ_i, or to use a numerical rule to approximate the integral. The order statistics are usually not easy to calculate exactly. Even the approximation to these using the inverse cumulative density function (cdf) at i/n (or, e.g., (i-0.5)/n) is not easy to calculate for the Beta and Gamma distributions except when the shape parameters are integers, since the cdf does not have a simple form. This suggests that a numerical rule may be the best

method. These rules only need the pdf value and can often give very accurate results with only a very small number of points.

Beta and Gamma distributions: The Beta Be(ν_1, ν_2) distribution on (0, 1) can be transformed to the interval ($\lambda_{\min}$, λ). The pdf is asymmetric if $\nu_1 \neq \nu_2$. Aτ $\lambda_{\min}$, the pdf is unbounded if $\nu_1 < 1$ and is 0 if $\nu_1 > 1$, and its gradient is $+\infty$ if $1 < \nu_1 < 2$. Similar properties hold at λ in terms of ν_2. Efficient estimation of ν_1, ν_2 is not simple and as only a guide is needed here, the less efficient method of moments has been used. Let m=-$\lambda_{\min}/(\lambda - \lambda_{\min})$=$r_{mm}/(1+r_{mm})$ and t=-n$\lambda\lambda_{\min}/tr(W^2)$-1. Then $\hat{\nu}_1$=tm, $\hat{\nu}_2$=t(1-m).

Although ($\hat{\nu}_1$, $\hat{\nu}_2$) may be useful summaries of the eigenvalue distribution, the Beta distribution usually fits badly. A large number of the $\hat{\nu}_i$ are less than 1, giving the wrong behaviour at $\lambda_{\min}$ or λ. In many of the remaining cases, both $\hat{\nu}_1$ and $\hat{\nu}_2$ are between 1 and 2, which results in the pdf having a 'hump' shape, which is again inappropriate.

Transforming the Beta distribution to the interval ($\lambda_{\min}-\delta_1$, $\lambda + \delta_2$) for some δ_1, $\delta_2 \geq 0$, but truncated to the interval ($\lambda_{\min}$, λ), may be better. However, this would require the estimation of δ_1, δ_2 also.

The Gamma Ga(ν, ζ) distribution (shape parameter ν, scale parameter ζ, mean $\nu\zeta$) can be transformed to the interval ($\lambda_{\min}$, ∞). The pdf at $\lambda_{\min}$ is unbounded for $\nu<1$ and is 0 for $\nu>1$. Again, using moments for simplicity, $\hat{\nu}$=n$\lambda_{\min}^2/tr(W^2)$ and $\hat{\zeta}$=-$tr(W^2)/(n\lambda_{\min})$.

The estimated Gamma distributions are more satisfactory than the estimated Beta distributions. The fitted pdf usually appears to be an adequate, if coarse, approximation to the eigenvalue distribution. Estimated Gamma distributions for Winnipeg:71, Syracuse:CT and Syracuse:CB are shown in Fig. 1, 3, 5 (C) and 2, 4, 6 (R), respectively. For all four combinations of C, R with $\hat{\nu}$, $\hat{\zeta}$, there appears to be a good linear log-log relationship with n (correlations of 0.859, 0.789, -0.659, -0.824). For $\hat{\nu}$ the approximate relationship is $0.9n^{1/5}$ for both C and R. The approximate relationship for $\hat{\zeta}$ is $2n^{-1/14}$ for C and $0.5n^{-1/8}$ for R. Alternatively, $\hat{\zeta}$ can be obtained for a given $\hat{\nu}$ using $\hat{\nu}\hat{\zeta}^2$=$tr(W^2)/n$.

The Gamma distribution could be truncated to the interval ($\lambda_{\min}$, λ), but the incomplete gamma integral would have to be estimated. Again, to allow a non-zero density at $\lambda_{\min}$, but, for $\nu>1$, a mode bigger than $\lambda_{\min}$, the Gamma distribution could be transformed to the interval ($\lambda_{\min}-\delta$, ∞) for some $\delta \geq 0$, but truncated to the interval ($\lambda_{\min}$, ∞) or ($\lambda_{\min}$, λ). This would require the estimation of δ also.
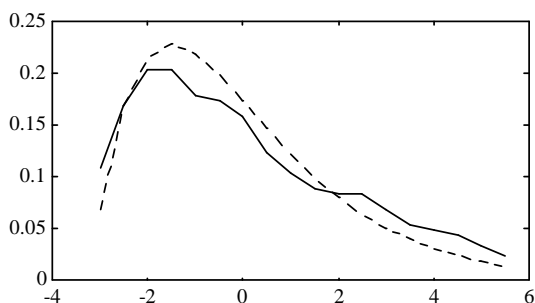
Fig. 1: Plot of smoothed histogram (solid) and fitted Gamma (dashed) for C for Winnipeg: 71

Fig. 2: Plot of smoothed histogram (solid), fitted Gamma (dashed) and kernel density estimate (+) for R for Winnipeg: 71

Nonparametric density estimation: If a numerical rule is to be used, then a simple nonparametric estimate of the pdf may suffice. Possibilities include a histogram, a smoothed histogram (see earlier) and a kernel density estimate. A kernel density estimate using a Gaussian kernel with smoothing parameter h close to that given by equation (3.31) in Silverman[19] has been tried. Figure 2 shows a kernel density estimate (h=0.16, limited to $[\lambda_{min}, \lambda]$) with the Gamma fit and the smoothed histogram for the R-distribution for Winnipeg: 71.
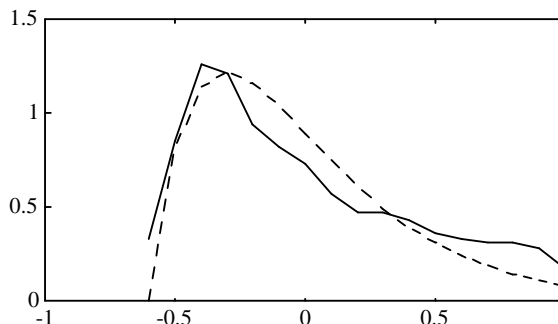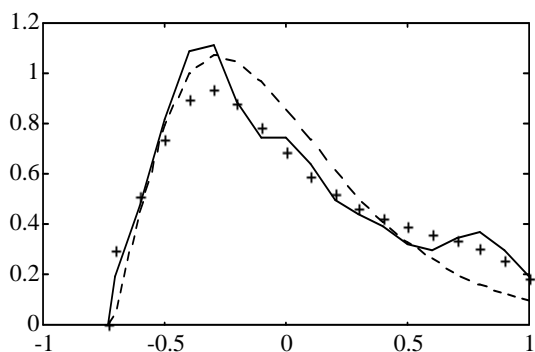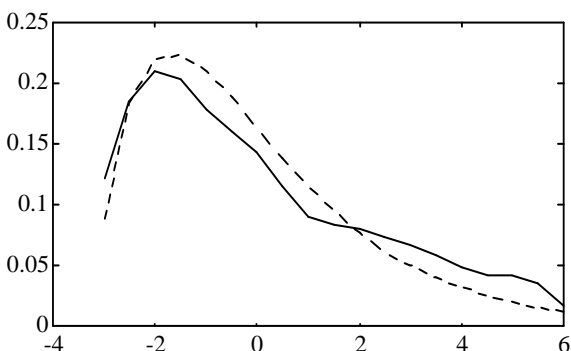

Fig. 3: Plot of smoothed histogram (solid) and fitted Gamma (dashed) for C for Syracuse:CT

Fig. 4: Plot of smoothed histogram (solid) and fitted Gamma (dashed) for R for Syracuse:CT

Estimating J(β) by numerical quadrature: Although the Gamma fits do not appear to be too bad, it turns out that biases in the fitted pdf can seriously affect the estimation of J(β). This seems to be because $ln(1-\beta x)f(x)$ is both positive and negative over the range, having the sign of βx. For example, for Winnipeg:71, the R matrix, the true value of J(0.6) is 0.0452. The fitted Gamma (λ_{min}=-0.7267, $\hat{\nu}$ = 2.593, $\hat{\zeta}$ = 0.2803) and a 3-pt MA smoothed histogram (from a histogram centred at 0.7(0.1)1) estimate of the eigenvalue distribution, are shown in Fig. 2. The fitted Gamma has area 0.9655 over the range. The mode is shifted to the right with respect to the eigenvalue distribution. Using a quadrature routine with the fitted Gamma, the value of J(0.6) is estimated to be -0.0002.

However, numerical methods on approximate distributions can give good results - see the examples. In practice, only a rough estimate of the distribution would be available and it may not be convenient to use many evaluation points. Possibilities would be a small-point Simpson rule or the more complicated but more efficient Gaussian quadrature.

Since $-ln(1-\beta x)f(x)$ is negative for βx<0 and positive for βx>0, it may be better to estimate the integrals for x<0 and x>0 separately. For example, using two 3-point Simpson rules would approximate J(β) by

$-\{f(\lambda)ln(1-\beta\lambda)+ 4f(\lambda/2)ln(1-\beta\lambda/2)-$

$4\lambda_{min}f(\lambda_{min}/2)ln(1-\beta\lambda_{min}/2)-$

$\lambda_{min}f(\lambda_{min})ln(1-\beta\lambda_{min})\}/3.$

Two 3-point Gaussian rules for a quadratic would approximate J(β) by

$-\{5f(a_1\lambda)ln(1-a_1\beta\lambda) + 8f(\lambda/2)ln(1-\beta\lambda/2) +$

$5f(a_2\lambda)ln(1-a_2\beta\lambda)-5\lambda_{min}f(a_2\lambda_{min})ln(1-a_2\beta\lambda_{min})-$

$8\lambda_{min}f(\lambda_{min}/2)ln(1-\beta\lambda_{min}/2)-$

$5\lambda_{min}f(a_1\lambda_{min})ln(1-a_1\beta\lambda_{min})\}/18$, where $a_1 \approx 0.8873$, $a_2 \approx 0.1127$.

Unfortunately, these do not appear to work well. For Winnipeg:71, R, the estimated values of f(x) using the 3-pt MA smoothed histogram (Fig. 2) would be approximately 0.31, 0.32, 0.62, 0.74, 1.10, 0.35, at the points $a_1\lambda$, $\lambda/2$, $a_2\lambda$, $a_2\lambda_{min}$, $\lambda_{min}/2$, $a_1\lambda_{min}$, respectively. Then the Gaussian rule estimates J(0.6) as 0.007.
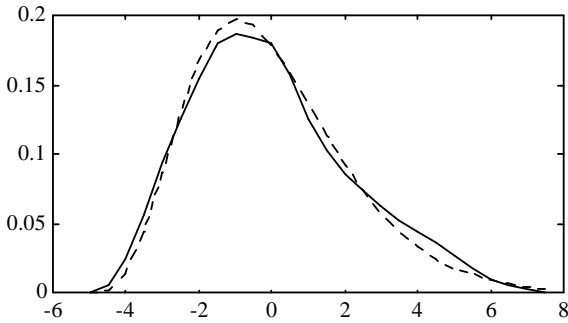
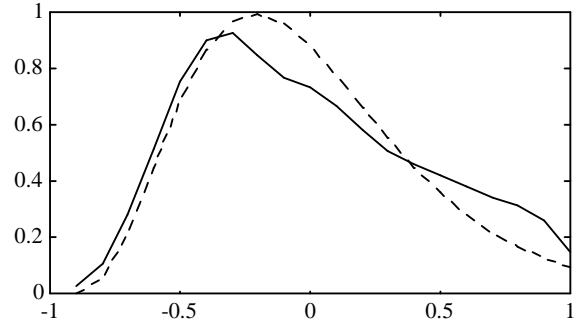Fig. 5: Plot of smoothed histogram (solid) and fitted Gamma (dashed) for C for Syracuse:CB

Fig. 6: Plot of smoothed histogram (solid) and fitted Gamma (dashed) for R for Syracuse:CB

Table 1: J(β) approximations for Winnipeg:71, C matrix

β	-0.3	-0.25	-0.2	-0.15	-0.1	-0.05	0.05	0.1	0.15
true	.2973	.1677	.0988	.0534	.0235	.0059	.0066	.0295	.0821
tr1	.2820	.1766	.1038	.0550	.0237	.0059	.0066	.0287	.0719
tr2	.2696	.1710	.1018	.0545	.0236	.0059	.0066	.0290	.0757
SS	.1870	.1100	.0616	.0294	.0093	.0004	.0118	.0388	.0934
SK	.2253	.1298	.0745	.0381	.0147	.0022	.0090	.0326	.0810

Table 2: J(β) approximations for Winnipeg:71, R matrix

β	-1	-0.8	-0.6	-0.4	0.2	0.4	0.6	0.8	0.9
True	.1052	.0644	.0355	.0157	.0043	.0182	.0452	.0942	.1360
tr1	.1078	.0658	.0359	.0158	.0043	.0180	.0436	.0840	.1110
tr2	.1067	.0654	.0358	.0157	.0043	.0181	.0439	.0861	.1168
SS	.0942	.0576	.0313	.0135	.0048	.0188	.0456	.0941	.1364
SK	.0831	.0487	.0249	.0093	.0064	.0218	.0492	.0970	.1376

Table 3: J(β) approximations for Syracuse:CT, C matrix

β	-0.3	-0.25	-0.2	-0.15	-0.1	-0.05	0.05	0.1	0.15
true	.3082	.1797	.1067	.0580	.0256	.0065	.0074	.0336	.0987
tr1	.3168	.1964	.1145	.0603	.0260	.0065	.0073	.0325	.0823
tr2	.3056	.1914	.1127	.0598	.0259	.0065	.0074	.0327	.0869
SS	.2013	.1181	.0659	.0313	.0097	-.001	.0134	.0445	.1115
SK	.2323	.1321	.0748	.0373	.0136	.0013	.0110	.0387	.0981

Table 4: J(β) approximations for Syracuse:CT, R matrix

β	-1	-0.8	-0.6	-0.4	0.2	0.4	0.6	0.8	0.9
True	.0885	.0555	.0310	.0139	.0038	.0165	.0410	.0853	.1224
tr1	.0930	.0573	.0315	.0140	.0038	.0163	.0395	.0762	.1007
tr2	.0922	.0570	.0315	.0140	.0038	.0164	.0397	.0777	.1048
SS	.0789	.0487	.0265	.0113	.0045	.0173	.0413	.0836	.1186
SK	.0738	.0448	.0237	.0095	.0054	.0190	.0436	.0865	.1217

EXAMPLES

The approximations to J(β) were compared for a range of β for two moderately large configurations: Winnipeg:71 (n=101) and Syracuse:CT (n=143). The following were calculated: the true value J(β); the 4th-order trace approximation (1) (tr1); an improved version if λ is known which is always better if $\beta > 0$[5] (tr2):

$$\frac{1}{n}\sum_{k=2}^{4}[\{tr(W^k) - \lambda^k\}\beta^k / k] - \frac{1}{n} \{ln(1 - \beta\lambda) + \beta\lambda\};$$

plus four estimated J*(β) using Simpson's rule (usually 19-point). These were on a histogram (SH); a 3pt-MA smoothed histogram (SS); the fitted Beta (SB) and

Gamma (SG) distributions; and the earlier kernel density estimate (SK). The SB and SG estimates were usually very bad and the SH estimates were usually much worse than the SS ones. The SB, SG and SH methods are therefore not discussed further.

Tables 1 to 4 give, respectively, the J(β) approximations for the C matrix for Winnipeg:71; the R matrix for Winnipeg:71; the C matrix for Syracuse:CT; the C matrix for Syracuse:CT. The valid ranges of β for the CAR are (-0.305, 0.174), (-1.376, 1) (-0.308, 0.165) (-1.666, 1) respectively.

The values in Tables 1 to 4 confirm that the trace methods give simple accurate approximations provided |β| is not too large (β between about $0.6/\lambda_{min}$ and $0.6/\lambda$), with tr2 better if λ is known. For the larger

values of $\beta>0$, the SS and SK approximations are usually quite good. Although it is normally unimportant if β is appreciably negative, these two approximations are less good then, possibly because it is difficult for the histograms to pinpoint the start of the distribution and the fast ascent in the density there.

It is also necessary to know what effect an approximate $J(\beta)$ may have on the estimate of β. Using the Winnipeg:71 R matrix and taking e'We/e'e as 0.1, 0.2, 0.5, the effect of a 10% or a 20% relative error in $J(\beta)$ was considered for a CAR model (with $v=1_n$). The 10% error led to the relative errors in the estimate of about 8, 5 and 0.6% respectively and the 20% error gave approximately double these. Thus relative errors in approximating $J(\beta)$ do not appear to be so serious for large β, although for the CAR substantial spatial dependence means that β must be near its upper limit.

DISCUSSION

Simple quick methods for approximating the determinant term in Gaussian maximum likelihood appear to be possible given either all the eigenvalues of W, or a good approximation to their distribution, but much more problematic otherwise. It appears to be difficult to find good summaries of the eigenvalue distribution for a general regional configuration. The trace approximations are easy to use and good if β is not too close to its limits and n is not too large. Numerical quadrature using an approximation to the eigenvalue distribution can be satisfactory if the estimated pdf is not too bad and sufficient points are used in the numerical rule.

Appendix 1: Values of n, I_C, I_R, $\lambda(C)$, $-\lambda_{min}(C)$, $b_1(C)$, $b_2(C)$, $-\lambda_{min}(R)$, $b_1(R)$, $b_2(R)$, for the 70 configurations (plus, in the R columns, $-\lambda_{min}(U)$, $b_1(U)$, $b_2(U)$ for the four U matrices)

Region	n	I_C	I_R	$\lambda(C)$	$-\lambda_{min}(C)$	$b_1(C)$	$b_2(C)$	$-\lambda_{min}(R)$	$b_1(R)$	$b_2(R)$
PR1	11	0.471	1.052	3.686	2.120	0.737	-0.623	0.719	0.532	-1.041
PR2	13	0.476	1.389	3.948	2.199	0.741	-0.475	0.685	0.531	-1.035
C71:4	13	0.667	0.877	4.238	2.564	0.635	-0.615	0.620	0.554	-0.851
C71:1	14	0.783	1.116	4.800	2.363	0.828	-0.341	0.553	0.740	-0.712
O1	14	0.696	1.245	4.552	2.312	0.813	-0.428	0.551	0.682	-0.826
C71:12	14	0.783	1.237	4.700	2.655	0.782	-0.400	0.599	0.703	-0.651
PR4	14	0.348	0.926	3.373	2.125	0.660	-0.889	0.630	0.608	-1.128
C71:17	14	0.696	0.915	4.364	2.384	0.762	-0.545	0.553	0.717	-0.742
C71:7	16	0.593	1.166	4.266	2.613	0.688	-0.611	0.576	0.630	-0.960
PR3	16	0.630	1.225	4.410	2.279	0.844	-0.406	0.568	0.714	-0.785
C71:16	16	0.778	0.935	4.740	2.644	0.746	-0.494	0.671	0.625	-0.692
C86:12	17	0.724	1.234	4.746	2.680	0.738	-0.423	0.621	0.663	-0.745
C71:21	17	0.724	1.135	4.703	2.500	0.738	-0.473	0.660	0.597	-0.746
C71:19	18	0.774	1.461	5.070	2.627	0.823	-0.221	0.579	0.724	-0.642
C86:4	19	0.818	1.250	5.051	3.035	0.705	-0.378	0.602	0.669	-0.670
PR5	19	0.394	1.163	4.158	2.285	0.696	-0.578	0.754	0.520	-1.044
C71:20	19	0.667	1.673	4.643	2.651	0.731	-0.542	0.628	0.633	-0.857
C71:24	19	0.667	1.151	4.862	2.654	0.841	-0.162	0.600	0.659	-0.731
C86:1	20	0.971	1.646	5.767	2.751	0.910	-0.027	0.526	0.837	-0.397
C71:10	21	0.811	0.971	4.982	2.684	0.825	-0.354	0.599	0.772	-0.517
C86:17	21	0.757	1.330	4.945	2.729	0.760	-0.411	0.600	0.660	-0.745
C86:19	21	0.784	1.084	4.933	2.809	0.794	-0.354	0.578	0.754	-0.555
C71:18	22	0.744	1.269	5.006	2.571	0.788	-0.430	0.579	0.687	-0.757
C71:22	23	0.854	1.488	5.534	2.804	0.898	-0.101	0.551	0.828	-0.478
O10	25	0.689	1.386	4.994	2.536	0.806	-0.401	0.588	0.677	-0.800
PR7	25	0.467	1.414	4.532	2.482	0.738	-0.475	0.694	0.562	-1.055
O4	26	0.702	1.599	5.131	2.587	0.808	-0.287	0.634	0.646	-0.776
O4U	26	*	*	*	*	*	*	0.798	0.342	-1.017
C86:24	26	0.915	1.527	5.720	2.872	0.868	-0.200	0.524	0.799	-0.525
PR6	27	0.612	1.297	4.789	2.413	0.815	-0.426	0.580	0.708	-0.802
PR9	27	0.510	1.499	4.662	2.406	0.757	-0.452	0.594	0.618	-0.944
C71:13	28	0.667	1.315	4.789	3.079	0.660	-0.430	0.747	0.523	-0.757
C86:20	28	0.863	1.557	5.475	3.025	0.807	-0.309	0.557	0.740	-0.581
C86:22	28	0.804	1.747	5.646	3.001	0.801	-0.112	0.596	0.705	-0.570
C71:15	29	0.755	1.744	5.357	2.894	0.856	-0.146	0.566	0.711	-0.651
O2	29	0.679	1.867	5.281	2.835	0.826	-0.147	0.673	0.630	-0.727
PR8	30	0.455	1.306	4.604	2.448	0.732	-0.479	0.754	0.522	-0.983
PR11	30	0.491	1.209	4.546	2.408	0.776	-0.522	0.630	0.668	-0.923
C86:7	32	0.898	2.016	5.893	3.422	0.795	-0.143	0.510	0.742	-0.598
PR10	32	0.644	1.467	4.968	2.719	0.775	-0.382	0.654	0.613	-0.821
C86:16	33	0.820	1.403	5.413	2.973	0.788	-0.325	0.580	0.725	-0.627
C71:25	35	0.708	1.358	5.102	2.769	0.737	-0.473	0.601	0.641	-0.821
C86:10	36	0.701	1.091	4.867	3.064	0.634	-0.580	0.644	0.558	-0.790
C86:18	39	0.753	1.270	5.267	2.948	0.694	-0.473	0.717	0.582	-0.724
O5	40	0.720	1.476	5.385	2.783	0.793	-0.348	0.660	0.665	-0.748
C71:8	40	0.653	1.136	5.044	2.637	0.748	-0.515	0.622	0.670	-0.750
O8	41	0.766	1.323	5.402	2.591	0.814	-0.404	0.561	0.716	-0.721
C71:5	42	0.532	1.290	4.555	3.082	0.600	-0.564	0.845	0.427	-0.960
C86:13	43	0.630	1.490	4.899	3.167	0.620	-0.487	0.783	0.441	-0.863
C71:26	45	0.765	1.475	5.492	2.898	0.788	-0.362	0.648	0.637	-0.733

C71:14	48	0.637	1.201	4.953	3.020	0.710	-0.433	0.667	0.606	-0.774
O3	49	0.731	1.946	5.908	3.097	0.903	0.037	0.651	0.714	-0.660
O9	50	0.800	1.575	5.654	2.834	0.805	-0.304	0.569	0.714	-0.662
C71:9	51	0.763	1.358	5.399	2.899	0.702	-0.453	0.691	0.580	-0.735
C86:26	53	0.802	1.523	5.624	2.980	0.775	-0.386	0.599	0.673	-0.726
C86:9	71	0.737	1.407	5.480	3.017	0.703	-0.459	0.733	0.605	-0.760
PR	73	0.716	1.544	5.502	2.762	0.812	-0.340	0.601	0.695	-0.737
PRU	73	*	*	*	*	*	*	0.618	0.644	-0.477
C86:8	76	0.884	1.686	5.923	3.274	0.788	-0.259	0.598	0.702	-0.576
C71:2	77	0.785	1.243	5.461	3.351	0.660	-0.506	0.656	0.600	-0.688
C71:3	79	0.693	1.146	5.168	3.327	0.597	-0.585	0.742	0.507	-0.802
C71:6	99	0.606	1.401	5.422	3.426	0.509	-0.534	0.798	0.401	-0.864
C71:11	101	0.777	1.650	5.742	3.384	0.726	-0.342	0.709	0.626	-0.708
C71:27	101	0.756	1.594	5.738	3.279	0.684	-0.411	0.727	0.580	-0.713
O6	130	0.804	1.888	6.111	3.029	0.868	-0.111	0.894	0.601	-0.637
O13	143	0.883	1.590	6.059	3.250	0.767	-0.355	0.600	0.708	-0.616
C86:11	192	0.831	2.334	6.822	4.080	0.733	-0.124	0.624	0.627	-0.728
C86:11U	192	*	*	*	*	*	*	0.627	0.618	-0.336
C71:23	351	0.746	1.277	5.582	3.434	0.615	-0.547	0.747	0.542	-0.752
O7	363	0.756	1.341	5.702	3.563	0.618	-0.536	0.778	0.539	-0.756
O7U	363	*	*	*	*	*	*	0.648	0.694	-0.092
O11	447	0.855	1.866	6.522	3.826	0.739	-0.295	0.696	0.664	-0.657
C86:23	731	0.776	1.819	7.084	4.749	0.651	-0.366	0.770	0.573	-0.710
O12	7249	0.711	2.402	7.375	5.025	0.590	-0.115	0.936	0.432	-0.756

Region codes (U for U matrix) - see the earlier text for details
Canada 1971 (C71) and Canada 1986 (C86): C71:1 & C86:1 Brantford; C71:2 Calgary; C71:3 Edmonton; C71:4 & C86:4 Guelph; C71:5 Halifax; C71:6 Hamilton; C71:7 & C86:7 Kingston; C71:8 & C86:8 Kitchener-Waterloo; C71:9 & C86:9 London; C71:10 & C86:10 Oshawa; C71:11 & C86:11 & C86:11U Ottawa-Hull; C71:12 & C86:12 Peterborough; C71:13 & C86:13 Regina; C71:14 St. Catherines - Niagara Falls; C71:15 St. Johns; C71:16 & C86:16 St. John's; C71:17 & C86:17 Sarnia; C71:18 & C86:18 Saskatoon; C71:19 & C86:19 Sault Ste. Marie; C71:20 & C86:20 Sherbrooke; C71:21 Sudbury; C71:22 & C86:22 Thunder Bay; C71:23 & C86:23 Toronto; C71:24 & C86:24 Trois Rivieres; C71:25 Victoria; C71:26 & C86:26 Windsor; C71:27 Winnipeg.

Puerto Rico: PR & PRU: Puerto Rico; PR1: Arecibo; PR2: Caguas; PR3: Mayaguez; PR4: Ponce; PR5: San Juan; PR6:Arecibo/Mayaguez; PR7: Arecibo/Ponce; PR8: Arecibo/San Juan; PR9: Caguas/Ponce; PR10: Caguas/San Juan; PR11: Mayaguez/Ponce.

Others: O1: Buffalo; O2: China; O3: Columbus, Ohio; O4 & O4U: Eire; O5: Ghana; O6: Henan, China; O7 & O7U: Houston; O8: Kansas-Nebraska; O9: Nigeria; O10: South Western Ohio; O11: Syracuse:BG; O12: Syracuse:CB; O13: Syracuse:CT.

With ever increasing computer power and the availability of better algorithms, it becomes easier and quicker to evaluate the determinant term for a given n, but at the same time, researchers then want to use data sets with increasingly large n. Thus, approximate fast techniques are always likely to have a use. At present, for a large n, a coarse approximation is recommended to see if significant spatial dependence is present and perhaps to choose a more appropriate form of the model (e.g. CAR or SAR), followed by exact methods if it is. The Monte Carlo approximation of Berry and Pace[11], which also involves finding or estimating $\lambda(C)$, would also be worth considering in many cases.

ACKNOWLEDGEMENT

REFERENCES

1. Cressie, N.A.C., 1991. Statistics for Spatial Data. Wiley: New York.
2. Martin, R.J., 1984. Exact maximum likelihood for incomplete data from a correlated Gaussian process. Commun. Statist. -Theor. Meth., 13: 1275-1288.
3. Whittle, P., 1954. On stationary processes in the plane. Biometrika, 41: 434-449.
4. Martin, R.J., 1987. Some comments on correction techniques for boundary effects and missing value techniques. Geog. Anal., 19: 273-282.
5. Martin, R.J., 1993. Approximations to the determinant term in Gaussian maximum likelihood estimation of some spatial models. Commun. Statist. -Theor. Meth., 22: 189-205.
6. Pace, R.K. and R. Barry, 1997a. Sparse spatial autoregressions. Statist. Probab. Lett., 33: 291-297.
7. Pace, R.K. and R. Barry, 1997b. Quick computation of regressions with a spatially autoregressive dependent variable. Geog. Anal., 29: 232-47.
8. Griffith, D.A., 1992. Simplifying the normalizing factor in spatial autoregressions for irregular lattices. Papers Reg. Sci., 71: 71-86.
9. Martin, R.J., 2005. Some results on graph spectra, with applications to geographic spatial modelling. Preprint. http://www.maths.otago.ac.nz/?martin

10. Griffith, D.A. and A. Sone, 1995. Trade-offs associated with normalizing constant computational simplifications for estimating spatial statistical models. J. Statist. Comput. Simul., 51: 165-183.
11. Barry, R. and R.K. Pace, 1999. A Monte Carlo estimator of the log determinant of large sparse matrices. Lin. Alg. Appl., 289: 41-54.
12. Fröberg, C.-E., 1969. Introduction to Numerical Analysis. Addison-Wesley: Reading, Ma.
13. Griffith, D.A., 1990. A numerical simplification for estimating parameters of spatial autoregressive models. In 'Spatial Statistics: Past, Present and Future', Ed. D.A. Griffith, pp: 109-127. Institute of Mathematical Geography: Ann Arbor.
14. Besag, J., 1975. Statistical analysis of non-lattice data. The Statistician, 24: 179-195.
15. Cliff, A. and J.K. Ord, 1981. Spatial processes - models and applications. Pion: London.
16. Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer: Dordrecht.
17. Haining, R.P., D.A. Griffith and R.J. Bennett, 1984. A statistical approach to the problem of missing data using a first order markov model. Prof. Geog., 36: 338-348.
18. Tiefelsdorf, M. and B. Boots, 1995. The exact distribution of Moran's I. Env. Plann., A 27: 985-999.
19. Silverman, B.W., 1986. Density estimation for statistics and data analysis. Chapman and Hall: London.