Lecture 1: Maximum likelihood estimation of spatial regression models

James P. LeSage
University of Toledo
Department of Economics
Toledo, OH 43606
jlesage@spatial-econometrics.com

March 2004

1 Introduction

This material should provide a reasonably complete introduction to traditional spatial econometric regression models. These models represent relatively straightforward extensions of regression models. After becoming familiar with these models you can start thinking about possible economic problems where these methods might be applied.

2 Spatial dependence

Spatial dependence in a collection of sample data implies that observations at location i depend on other observations at locations $j \neq i$. Formally, we might state:

$$y_i = f(y_j), i = 1, \ldots, n \quad j \neq i \tag{1}$$

Note that we allow the dependence to be among several observations, as the index i can take on any value from $i = 1, \ldots, n$.

Spatial dependence can arise from theoretical as well as statistical considerations.

1. **A theoretical motivation for spatial dependence.**

   From a theoretical viewpoint, consumers in a neighborhood may emulate each other leading to spatial dependence. Local governments might engage in competition that leads to local uniformity in taxes and services. Pollution can create systematic patterns over space, and clusters of consumers who travel to a more distant store to avoid a high crime zone would also generate these patterns. A concrete example will be given in Section 3.

2. **A statistical motivation for spatial dependence.**

   Spatial dependence can arise from unobservable latent variables that are spatially correlated. Consumer expenditures collected at spatial locations such as Census tracts exhibit spatial dependence, as do other variables such as housing prices. It seems plausible that difficult-to-quantify or unobservable characteristics such as the quality of life may also exhibit spatial dependence. A concrete example will be given in Section 3.

2.1 Estimation consequences of spatial dependence

In some applications, the spatial structure of the dependence may be a subject of interest or provide a key insight. In other cases, it may be a nuisance similar to serial correlation. In either case, inappropriate treatment of sample data with spatial dependence can lead to inefficient and/or biased and inconsistent estimates.

For models of the type: $y_i = f(y_j) + X_i\beta + \varepsilon_i$

Least-squares estimates for β are biased and inconsistent, similar to the simultaneity problem.

For model of the type: $y_i = X_i\beta + u_i, \quad u_i = f(u_j) + \varepsilon_i,$

Least-squares estimates for β are inefficient, but consistent, similar to the serial correlation problem.

3 Specifying dependence using weight matrices

There are several ways to quantify the structure of spatial dependence between observations, but a common specification relies on an nxn spatial weight matrix D with elements $D_{ij} > 0$ for observations $j = 1 \dots n$ sufficiently close (as measured by some metric) to observation i.

As a **theoretical motivation** for this type of specification, suppose we observe a vector of utility for 3 individuals. For the sake of concreteness, assume this utility is derived from expenditures on their homes. Let these be located on a regular lattice in space such that individual 1 is a neighbor to 2, and 2 is a neighbor to both 1 and 3, while individual 3 is a neighbor to 2. The spatial weight matrix based on this spatial configuration takes the form:

$$D = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{2}$$

The first row in D represents observation #1, so we place a value of 1 in the second column to reflect that #2 is a neighbor to #1. Similarly, both #1 and #3 are neighbors to observation #2 resulting in 1's in the first and third columns of the second row. In (2) we set $D_{ii} = 0$ for reasons that will become apparent shortly. Another convention is to normalize the spatial weight matrix D to have row-sums of unity, which we denote as W, known as a "row-stochastic matrix". We might express the utility, y as a function of observable characteristics $X\beta$ and unobservable characteristics ε, producing a spatial regression relationship:

$$\begin{aligned} (I_n - \rho W)y &= X\beta + \varepsilon \\ y &= \rho W y + X\beta + \varepsilon \\ y &= (I_n - \rho W)^{-1} X\beta + (I_n - \rho W)^{-1}\varepsilon \end{aligned} \tag{3}$$

Where the implied data generating process for the traditional *spatial autoregressive* (SAR) model is shown in the last expression in (3). The second expression for the SAR model makes it clear why $W_{ii} = 0$, as this precludes an observation y_i from directly predicting itself. It also motivates the use of row-stochastic W, which makes each observation y_i a function of the "spatial lag" Wy, an explanatory variable representing an average of spatially neighboring values, e.g., $y_2 = \rho(1/2y_1 + 1/2y_3)$.

Assigning a spatial correlation parameter value of $\rho = 0.5$, $(I_3 - \rho W)^{-1}$ is as shown in (4).

$$S^{-1} = (I - 0.5W)^{-1} = \begin{pmatrix} 1.1667 & 0.6667 & 0.1667 \\ 0.3333 & 1.3333 & 0.3333 \\ 0.1667 & 0.6667 & 1.1667 \end{pmatrix} \tag{4}$$

This model reflects a data generating process where $S^{-1}X\beta$ indicates that individual (or observation) #1 derives utility that reflects a linear combination of the observed characteristics of their own house as well as characteristics of both other homes in the neighborhood. The weight placed on own-house characteristics is slightly less than twice that of the neighbor's house (observation/individual #2) and around 7 times that of the non-neighbor (observation/individual #3). An identical linear combination exists for individual #3 who also has a single neighbor.

For individual #2 with two neighbors we see a slightly different weighting pattern to the linear combination of own and neighboring house characteristics in generation of utility. Here, both neighbors characteristics are weighted equally, accounting for around one-fourth the weight associated with the own-house characteristics.

Spatial models take this approach to describing variation in spatial data observations. Note that unobservable characteristics of houses in the neighborhood (which we have represented by ε) would be accorded the same weights as the observable characteristics by the data generating process.

Other points to note are that:

1. Increasing the magnitude of the spatial dependence parameter ρ would lead to an increase in the magnitude of the weights as well as a decrease in the decay as we move to neighbors and more distant non-neighbors (see the inverse in expression (6)).

2. The addition of another observation/individual not in the neighborhood, represented by another row and column in the weight matrix with zeros in all positions will have no effect on the linear combinations.

3. All connectivity relationships come into play through the matrix inversion process. By this we mean that house #3 which is a neighbor to #2 influences the utility of #1 because there is a connection between #3 and #2 which is a neighbor of #1.

4. We need not treat all neighboring (contiguity) relationships in an equal fashion. We could weight neighbors by distance, length of adjoining property boundaries, or any number of other schemes that have been advocated in the literature on spatial regression relationships (see Bavaud, 1998). In these cases, we might temper the comment above to reflect the fact that some connectivity relations may have very small weights, effectively eliminating them from having an influence during the data generating process.

Regarding points 1) and 4), the matrix inverse for the case of $\rho = 0.3$ is:

$$S^{-1} = \begin{pmatrix} 1.0495 & 0.3297 & 0.0495 \\ 0.1648 & 1.0989 & 0.1648 \\ 0.0495 & 0.3297 & 1.0495 \end{pmatrix} \tag{5}$$

while that for $\rho = 0.8$ is:

$$S^{-1} == \begin{pmatrix} 1.8889 & 2.2222 & 0.8889 \\ 1.1111 & 2.7778 & 1.1111 \\ 0.8889 & 2.2222 & 1.8889 \end{pmatrix} \tag{6}$$

3.1 Computational considerations

Note that we require knowledge of the location associated with the observational units to determine the closeness of individual observations to other observations. Sample data that contains address labels could be used in conjunction with Geographical Information Systems (GIS) or other dedicated software to measure the spatial proximity of observations. Assuming that address matching or other methods have been used to produce x-y coordinates of location for each observation in Cartesian space, we can rely on a Delaunay triangularization scheme to find neighboring observations. To illustrate this approach, Figure 1 shows a Delaunay triangularization centered on an observation located at point A. The space is partitioned into triangles such that there are no points in the interior of the circumscribed circle of any triangle. Neighbors could be specified using Delaunay contiguity defined as two points being a vertex of the same triangle. The neighboring observations to point A that could be used to construct a spatial weight matrix are B, C, E, F.

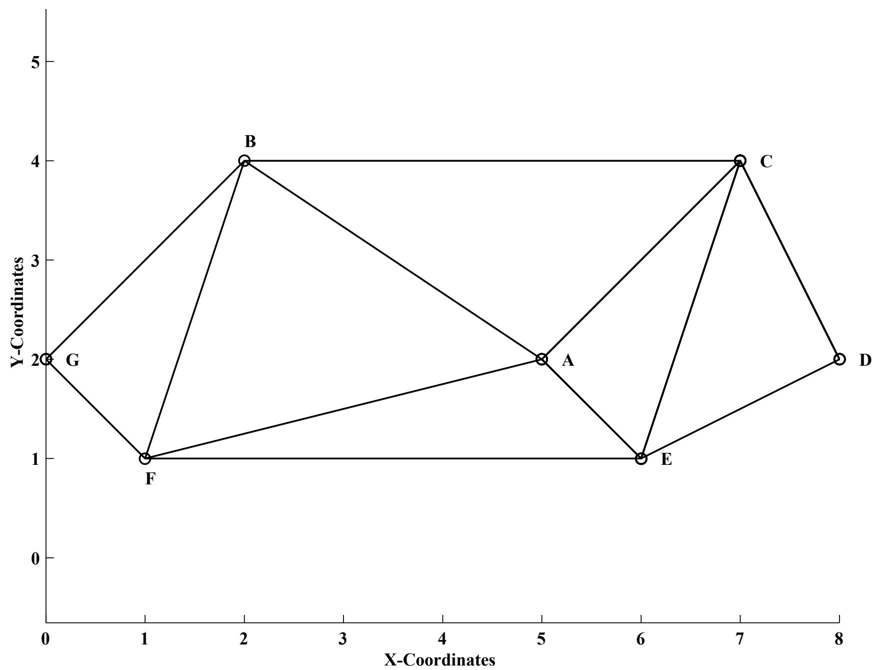

Figure 1: Delaunay triangularization

One way to specify a spatial weight matrix would be to set column elements associated with neighboring observations B, C, E, F equal to 1 in row A. This would reflect that these observations are neighbors to observation A. Typically, the weight matrix is standardized so that row sums equal unity, producing a row-stochastic weight matrix. Row-stochastic spatial weight matrices, or multidimensional linear filters, have a long history of application in spatial statistics (e.g., Ord, 1975).

An alternative approach would be to rely on neighbors ranked by distance from observation A. We can simply compute the distance from A to all other observations and rank these by size. In the case of figure 1 we would have: a nearest neighbor E, the nearest 2 neighbors E, C, nearest 3 neighbors E, C, D, and so on. Again, we could set elements $D_{Aj} = 1$ for observations j in row A to reflect any number of nearest neighbors to observation A, and transform to row-stochastic form. This approach might involve including a determination of the appropriate number of neighbors as part of the model estimation problem. (More will be said about this in Lecture #3.)

3.2 Spatial Durbin and spatial error models

We can extend the model in (3) to a *spatial Durbin model* (SDM), that allows for explanatory variables from neighboring observations, created by WX as shown in (7).

$$\begin{aligned}
(I_n - \rho W)y &= X\beta + WX\gamma + \varepsilon \\
y &= \rho Wy + X\beta + WX\gamma + \varepsilon \\
y &= (I_n - \rho W)^{-1}X\beta + (I_n - \rho W)^{-1}WX\gamma + (I_n - \rho W)^{-1}\varepsilon
\end{aligned} \tag{7}$$

The kx1 parameter vector γ measures the marginal impact of the explanatory variables from neighboring observations on the dependent variable y. Multiplying X by W produces "spatial lags" of the explanatory variables that reflect an average of neighboring observations $X-$values.

Another model that has been used is the *spatial error model* (SEM):

$$\begin{aligned}
y &= X\beta + u \\
u &= \rho W + \varepsilon \\
y &= X\beta + (I_n - \rho W)^{-1}\varepsilon
\end{aligned} \tag{8}$$

3.3 A statistical motivation for spatial dependence

Recent advances in address matching of customer information, house addresses, retail locations, and automated data collection using global positioning systems (GPS) have produced very large datasets that exhibit spatial dependence. The source of this dependence may be unobserved variables whose spatial variation over the sample of observations is the source of spatial dependence in y. Here we have a potentially different situation than described previously. There may be no underlying theoretical motivation for spatial dependence in the data generating process. Spatial dependence may arise here from census tract boundaries that do not accurately reflect neighborhoods which give rise to the variables being collected for analysis.

Intuitively, one might suppose solutions to this type of dependence would be: 1) to incorporate proxies for the unobserved explanatory variables that would eliminate the spatial dependence; 2) collect sample data based on alternative administrative jurisdictions that give rise to the information being collected; or 3) rely on latitude and longitude coordinates of the observations as explanatory variables, and perhaps interact these location variables with other explanatory variables in the relationship. The goal here would be to eliminate the spatial dependence in y, allowing us to proceed with least-squares estimation.

Note this is similar to the time-series case of serial dependence, where the source may be something inherent in the data generating process, or excluded important explanatory variables. However, I argue that it is much more difficult to "filter out" spatial dependence than it is to deal with serial correlation in time series.

To demonstrate this point, I will describe an experimental comparison of spatial and non-spatial models based on spatial data from a 1999 consumer expenditure survey along with 1990 US Census information. The consumer expenditure survey collects data on categories such as alcohol, tobacco, food, and entertainment in over 50,000 census tracts. To predict each of the 51 expenditure subcategories as well as overall consumer expenditures by census tract, 25 explanatory variables were used, including typical variables such as age, race, gender, income, age of homes, house prices, and so forth. The predictive performance of a simple least-squares model exhibited a median absolute percentage error of 8.21% across the categories (minimum 7.69% for furnishing expenditures and maximum of 8.91% for personal insurance expenditures) and the median R^2 was 0.94 across all categories. For comparison, a model using the variables and their squared values, along with these variables interacted with latitude and longitude, as well as latitude interacted with longitude, latitude squared, longitude squared, and 50 state dichotomous variables was constructed, producing a total of 306 variables. This is a more complex model of the type often employed by researchers confronted with spatial dependence.

This complex specification provides a way of incorporating both functional form and spatial information, allowing each parameter to have a quadratic surface over space. In addition, the constant term in this model can follow a quadratic over space, combined with a 50 state step function. As expected the addition of these 281 extra variables improved performance of the model, producing a median absolute percentage error of 6.99% across all categories. Both of these approaches yield reasonable results and one might wonder whether spatial modeling could substantially reduce the errors.

To address this, a spatial Durbin model was constructed using the original 25 explanatory variables plus spatial explanatory variables WX. This spatial Durbin model which is relatively parsimonious in comparison to the model containing 306 explanatory variables produced a median absolute percentage error of 6.23% across all categories, which is clearly better than the complicated model with 306 variables.

The spatial autoregressive estimate for ρ varied across the categories from a low of 0.72 to a high of 0.74, pointing to strong spatial dependence in the sample data. Note, the model subsumes the usual least-squares model, so testing for the presence of spatial dependence is equivalent to a likelihood ratio test for the autoregressive parameter $\rho = 0$. For this sample data all likelihood ratio tests overwhelmingly rejected spatial independence.

Relative to the non-spatial least-squares model based on 25 variables, the spatial model dramatically reduced prediction errors from 8.21% to 6.23%. These results point to some general empirical truths regarding spatial data samples.

1. A conventional regression augmented with geographic dichotomous variables (e.g., region or state dummy variables), or variables reflecting interaction with locational coordinates that allow variation in the parameters over space can rarely outperform

a simpler spatial model.

2. Spatial models provide a parsimonious approach to modeling spatial dependence, or filtering out these influences when they are perceived as a nuisance.

3. Gathering sample data from the appropriate administrative units is seldom a realistic option.

4 Maximum likelihood estimation of SAR, SEM, SDM models

Maximum likelihood estimation of the SAR, SDM and SEM models described here and in Anselin (1988) involves maximizing the log likelihood function (concentrated with respect to β and σ^2, the noise variance associated with ε) with respect to the parameter ρ. For the case of the SAR model we have:

$$
\begin{aligned}
\ln L &= C + \ln|I_n - \rho W| - (n/2)\ln(e'e) \\
e &= e_o - \rho e_d \\
e_o &= y - X\beta_o \\
e_d &= Wy - X\beta_d \\
\beta_o &= (X'X)^{-1}X'y \\
\beta_d &= (X'X)^{-1}X'Wy
\end{aligned} \tag{9}
$$

Where C represents a constant not involving the parameters. The computationally troublesome aspect of this is the need to compute the log-determinant of the nxn matrix $(I_n - \rho W)$. Operation counts for computing this determinant grow with the cube of n for dense matrices. This same approach can be applied to the SDM model by simply defining $X = [X \;\; WX]$ in (9).

The SEM model has a concentrated log-likelihood taking the form:

$$
\begin{aligned}
\ln L &= C + \ln|I_n - \rho W| - (n/2)\ln(e'e) \\
\tilde{X} &= X - \rho WX \\
\tilde{y} &= y - \rho Wy \\
\beta^{\star} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}\tilde{y} \\
e &= \tilde{y} - \tilde{X}\beta^{\star}
\end{aligned} \tag{10}
$$

5 Efficient computation

While W is a nxn matrix, in typical problems W will be sparse. For a spatial weight matrix constructed using Delaunay triangles among the n points in two-dimensions, the average number of neighbors for each observation will equal 6, so the matrix will have $6n$ non-zeros out of n^2 possible elements, leading to $(6/n)$ as the proportion of non-zeros. Matrix multiplication and the various matrix decompositions require $O(n^3)$ operations for dense matrices, but for sparse W these operation counts can fall as low as $O(n_{\neq 0})$, where $n_{\neq 0}$ denotes the number of non-zeros.

5.1 Sparse matrix algorithms

One of the earlier computationally efficient approaches to solving for estimates in models involving a large number of observations was proposed by Pace and Barry (1997). They suggested using direct sparse matrix algorithms such as the Cholesky or LU decompositions to compute the log-determinant over a grid of values for the parameter ρ restricted to the interval $(-1/\lambda_{min}, 1/\lambda_{max})$, where $\lambda_{min}, \lambda_{max}$ represent the minimum and maximum eigenvalues of the spatial weight matrix. It is well known that for row-stochastic W, $\lambda_{\min} < 0, \lambda_{\max} > 0$, and that ρ must lie in the interval $[\lambda_{\min}^{-1}, \lambda_{\max}^{-1}]$ (see for example Lemma 2 in Sun et al., 1999). However, since negative spatial autocorrelation is of little interest in many cases, restricting ρ to the interval $[0, 1)$ can accelerate the computations. This along with a vector evaluation of the SAR or SDM log-likelihood functions over this grid of log-determinant values can be used to find maximum likelihood estimates. Specifically, for a grid of q values of ρ in the interval $[0, 1)$,

$$\begin{pmatrix} \mathrm{Ln}L(\beta, \rho_1) \\ \mathrm{Ln}L(\beta, \rho_2) \\ \vdots \\ \mathrm{Ln}L(\beta, \rho_q) \end{pmatrix} \propto \begin{pmatrix} \mathrm{Ln}|I_n - \rho_1 W| \\ \mathrm{Ln}|I_n - \rho_2 W| \\ \vdots \\ \mathrm{Ln}|I_n - \rho_q W| \end{pmatrix} - (n/2) \begin{pmatrix} \mathrm{Ln}(\phi(\rho_1)) \\ \mathrm{Ln}(\phi(\rho_2)) \\ \vdots \\ \mathrm{Ln}(\phi(\rho_q)) \end{pmatrix} \tag{11}$$

where $\phi(\rho_i) = e_o'e_o - 2\rho_i e_d'e_o + \rho_i^2 e_d'e_d$. (For the SDM model, we replace X with $[X \quad WX]$ in (9).) Note that the SEM model cannot be vectorized, and must be solved using more conventional optimization, such as a simplex algorithm. Nonetheless, a grid of values for the log-determinant over the feasible range for ρ can be used to speed evaluation of the log-likelihood function during optimization with respect to ρ.

The computationally intense part of this approach is still calculating the log-determinant, which takes around 201 seconds for a sample of 57,647 observations representing all Census tracts in the continental US. This is based on a grid of 100 values from $\rho = 0$ to 1 using sparse matrix algorithms in MATLAB version 6.0 on a 600 Mhz Pentium III computer. Note, if the optimum ρ occurs on the boundary (i.e., 0), this indicates the need to consider negative values of ρ.

In applied settings one may not require the precision of the direct sparse method, so it seems worthwhile to examine other approaches to the problem that simply approximate the log-determinant.

5.2 A Monte Carlo approximation to the log-determinant

An improvement based on a Monte Carlo estimator for the log determinant suggested by Barry and Pace (1999) allows larger problems to be tackled without the memory requirements or sensitivity to orderings associated with the direct sparse matrix approach. The estimator for $\ln|I_n - \rho W|$ is based on an asymptotic 95% confidence interval, $(\bar{V} - F, \bar{V} + F)$

for the log-determinant constructed using the mean $\bar{V}$ of p generated independent random variables taking the form:

$$V_i = -n \sum_{k=1}^{m} \frac{x_i' W^k x_i}{x_i' x_i} \frac{\rho^k}{k}, \quad i = 1, \ldots, p \tag{12}$$

where $x_i \sim N(0,1)$, x_i independent of x_j if $i \neq j$. This is just a Taylor series expansion of the log-determinant that relies on the random variates x_i to compute the trace without multiplying n by n matrices. Note that $tr(W) = 0$ and we can easily compute $tr(W^2) = \sum\sum W \odot W'$ using element-wise multiplication represented by $\odot$. In the case of symmetric W this reduces to taking the sum of squares of all the elements. This requires a number of operations proportional to the non-zero elements which are small since W is a sparse matrix. By replacing the estimated traces with exact ones, the precision of the method can be improved at low cost. In addition,

$$F = \frac{n\rho^{m+1}}{(m+1)(1-\rho)} + 1.96\sqrt{s^2(V_1, \ldots, V_p)/p} \tag{13}$$

where s^2 is the estimated variance of the generated V_i. Values for m and p are chosen by the user to provide the desired accuracy for the approximation. The method provides not only an estimate of the log-determinant term, but an empirical measure of accuracy as well, in the form of upper and lower 95% confidence interval estimates. The log-determinant estimator yields the correct expected value for both symmetric and asymmetric spatial weight matrices, but the confidence intervals above require symmetry. For asymmetric matrices, a Chebyshev interval may prove more appropriate.

The computational properties of this estimator are related to the number of non-zero entries in W which we denote as f. Each realization of V_i takes time proportional to fm, and computing the entire estimator takes time proportional to fmp. For cases where the number of non-zero entries in the sparse matrix W increase linearly with n, the required time will be order nmp.

In addition to the time advantages of this algorithm, memory requirements are quite frugal because the algorithm only requires storage of the sparse matrix W along with space for two nx1 vectors, and some additional space for storing intermediate scalar values. Additionally, the algorithm does not suffer from "fill-in" since the storage required does not increase as computations are performed.

Another advantage is that this approach works in conjunction with the calculations proposed by Pace and Barry (1997) for a grid of values over ρ, since a minor modification of the algorithm can be used to generate log-determinants for a simultaneous set of ρ values $\rho_1, \ldots, \rho_q$, with very little additional time or memory.

As an illustration of these computational advantages, the time required to compute a grid of log-determinant values for $\rho = 0, \ldots, 1$ based on 0.001 increments for the sample of $57,647$ observations was 3.6 seconds when setting $m = 20, p = 5$, which compares quite favorably to 201 seconds for the direct sparse matrix computations cited earlier. This approach yielded nearly the same estimate of ρ as the direct method (0.91 versus 0.92), despite the use of a spatial weight matrix based on pure nearest neighbors rather than the symmetricized nearest neighbor matrix used in the direct approach.

LeSage and Pace (2001) report experimental results indicating robustness of the Monte Carlo log-determinant estimator is rather remarkable, and suggest a potential for other approximate approaches. Smirnov and Anselin (2001) provide an approach based on a Taylor's series expansion and Pace and LeSage (2001) use a Chebyshev expansion.

5.3 Estimates of dispersion

So far, the estimation procedure set forth produces an estimate for the spatial dependence parameter ρ through maximization of the log-likelihood function concentrated with respect to β, σ. Estimates for these parameters can be recovered given $\hat{\rho}$, the likelihood maximizing value of ρ using:

$$\begin{aligned}
\hat{e} &= e_o - \hat{\rho} e_d \\
e_o &= y - X\beta_o \\
e_d &= Wy - X\beta_d \\
\hat{\sigma}^2 &= (\hat{e}'\hat{e})/(n-k) \\
\beta_o &= (X'X)^{-1}X'y \\
\beta_d &= (X'X)^{-1}X'Wy \\
\hat{\beta} &= \beta_o - \hat{\rho}\beta_d
\end{aligned} \tag{14}$$

An implementation issue is constructing estimates of dispersion for these parameter estimates that can be used for inference. For problems involving a small number of observations, we can use our knowledge of the theoretical information matrix to produce estimates of dispersion. An asymptotic variance matrix based on the Fisher information matrix shown below for the parameters $\theta = (\rho, \beta, \sigma^2)$ can be used to provide measures of dispersion for the estimates of ρ, β and σ^2. Anselin (1988) provides the analytical expressions needed to construct this information matrix.

$$[I(\theta)]^{-1} = -E[\frac{\partial^2 L}{\partial\theta\partial\theta'}]^{-1} \tag{15}$$

This approach is computationally impossible when dealing with large scale problems involving thousands of observations. The expressions used to calculate terms in the information matrix involve operations on very large matrices that would take a great deal of computer memory and computing time. In these cases we can evaluate the numerical hessian matrix using the maximum likelihood estimates of ρ, β and σ^2 and our sparse matrix representation of the likelihood. Given the ability to evaluate the likelihood function rapidly, numerical methods can be used to compute approximations to the gradients shown in (15).

6 Applied examples

The first example represents a hedonic pricing model often used to model house prices, using the selling price as the dependent variable and house characteristics as explanatory variables. Housing values exhibit a high degree of spatial dependence.

Here is a MATLAB program that uses functions from the **Spatial econometrics Toolbox** available at www.spatial-econometrics.com to carry out estimation of a least-squares model, spatial Durbin model and spatial autoregressive model.

```
% example1.m file
% An example of spatial model estimation compared to least-squares
load house.dat;
% an ascii datafile with 8 colums containing 30,987 observations on:
% column 1 selling price
% column 2 YrBlt, year built
% column 3 tla, total living area
% column 4 bedrooms
% column 5 rooms
% column 6 lotsize
% column 7 latitude
% column 8 longitude
y = log(house(:,1));      % selling price as the dependent variable
n = length(y);
xc = house(:,7);          % latitude coordinates of the homes
yc = house(:,8);          % longitude coordinates of the homes
% xy2cont() is a spatial econometrics toolbox function
[j W j] = xy2cont(xc,yc); % constructs a 1st-order contiguity
                          % spatial weight matrix W, using Delauney triangles
age = 1999 - house(:,2);  % age of the house in years
age = age/100;
x = zeros(n,8);           % an explanatory variables matrix
x(:,1) = ones(n,1);       % an intercept term
x(:,2) = age;             % house age
x(:,3) = age.*age;        % house age-squared
x(:,4) = age.*age.*age;   % house age-cubed
x(:,5) = log(house(:,6)); % log of the house lotsize
x(:,6) = house(:,5);      % the # of rooms
x(:,7) = log(house(:,3)); % log of the total living area in the house
x(:,8) = house(:,4);      % the # of bedrooms
vnames = strvcat('log(price)','constant','age','age2','age3','lotsize','rooms','tla','beds');
result0 = ols(y,x);   % ols() is a toolbox function for least-squares estimation
prt(result0,vnames);  % print results using prt() toolbox function
% compute ols mean absolute prediciton error
mae0 = mean(abs(y - result0.yhat));
fprintf(1,'least-squares mean absolute prediction error = %8.4f \n',mae0);
% compute spatial Durbin model estimates
info.rmin = 0;       % restrict rho to 0,1 interval
info.rmax = 1;
result1 = sdm(y,x,W,info); % sdm() is a toolbox function
prt(result1,vnames);       % print results using prt() toolbox function
% compute mean absolute prediction error
mae1 = mean(abs(y - result1.yhat));
fprintf(1,'sdm mean absolute prediction error = %8.4f \n',mae1);
% compute spatial autoregressive model estimates
result2 = sar(y,x,W,info); % sar() is a toolbox function
prt(result2,vnames);       % print results using prt() toolbox function
% compute mean absolute prediction error
mae2 = mean(abs(y - result2.yhat));
fprintf(1,'sar mean absolute prediction error = %8.4f \n',mae2);
```

It is instructive to compare the biased least-squares estimates for β to those from the

Table 1: A comparison of least-squares and SAR model estimates

Variable	OLS β	OLS t-statistic	SAR β	SAR t-statistic
constant	2.7599	35.6434	-0.3588	-10.9928
age	1.9432	28.5152	1.0798	22.4277
age2	-4.0476	-32.8713	-1.8522	-21.2607
age3	1.2549	18.9105	0.5042	10.7939
lotsize	0.1857	48.1380	0.0517	19.7882
rooms	0.0112	2.9778	-0.0033	-1.3925
tla	0.9233	74.8484	0.5317	70.6514
beds	-0.0150	-2.6530	0.0197	4.9746

SAR model shown in Table 1. We see upward bias in the least-squares estimates indicating over-estimation of the sensitivity of selling price to the house characteristics when spatial dependence is ignored. This is a typical result for hedonic pricing models. The SAR model estimates reflect the marginal impact of home characteristics after taking the spatial location into account.

In addition to the upward bias, there are some sign differences as well as different inferences that would arise from least-squares versus SAR model estimates. For instance, least-squares indicates that more bedrooms has a significant negative impact on selling price, an unlikely event. The SAR model suggest that bedrooms have a positive impact on selling prices.

Here are the results printed by MATLAB from estimation of the three models using the spatial econometrics toolbox functions.

```
Ordinary Least-squares Estimates
Dependent Variable =       log(price)
R-squared      =    0.7040
Rbar-squared   =    0.7039
sigma^2        =    0.1792
Durbin-Watson  =    1.0093
Nobs, Nvars    =  30987,     8
***************************************************************
Variable        Coefficient      t-statistic    t-probability
constant           2.759874        35.643410         0.000000
age                1.943234        28.515217         0.000000
age2              -4.047618       -32.871346         0.000000
age3               1.254940        18.910472         0.000000
lotsize            0.185666        48.138001         0.000000
rooms              0.011204         2.977828         0.002905
tla                0.923255        74.848378         0.000000
beds              -0.014952        -2.653031         0.007981

least-squares mean absolute prediction error = 0.3066
```

The fit of the two spatial models is superior to that from least-squares as indicated by both the $R-$squared statistics as well as the mean absolute prediction errors reported with the printed output.

Turning attention to the SDM model estimates, here we see that the lotsize, number of rooms, total living area and number of bedrooms for neighboring (first-order contiguous) properties that sold have a negative impact on selling price. (These variables are reported using $W-$variable name, to reflect the spatially lagged explanatory variables WX in this model.) This is as we would expect, the presence of neighboring homes with larger lots, more rooms and bedrooms as well as more living space would tend to depress the selling price. Of course, the converse also is true, neighboring homes with smaller lots, less rooms and bedrooms as well as less living space would tend to increase the selling price.

```
Spatial autoregressive Model Estimates
Dependent Variable =       log(price)
R-squared          =    0.8537
Rbar-squared       =    0.8537
sigma^2            =    0.0885
Nobs, Nvars        =  30987,     8
log-likelihood     =       -141785.83
# of iterations    =     11
min and max rho    =    0.0000,   1.0000
total time in secs =   32.0460
time for lndet     =   19.6480
time for t-stats   =   11.8670
Pace and Barry, 1999 MC lndet approximation used
order for MC appr  =     50
iter  for MC appr  =     30
***************************************************************
Variable       Coefficient  Asymptot t-stat    z-probability
constant         -0.358770      -10.992793         0.000000
age               1.079761       22.427691         0.000000
age2             -1.852236      -21.260658         0.000000
age3              0.504158       10.793884         0.000000
lotsize           0.051726       19.788158         0.000000
rooms            -0.003318       -1.392471         0.163780
tla               0.531746       70.651381         0.000000
beds              0.019709        4.974640         0.000001
rho               0.634595      242.925754         0.000000

sar mean absolute prediction error =   0.2111

Spatial Durbin model
Dependent Variable =       log(price)
R-squared          =    0.8622
Rbar-squared       =    0.8621
sigma^2            =    0.0834
log-likelihood     =       -141178.27
Nobs, Nvars        =  30987,     8
# iterations       =     15
min and max rho    =    0.0000,   1.0000
total time in secs =   54.3180
time for lndet     =   18.7970
time for t-stats   =   29.6230
Pace and Barry, 1999 MC lndet approximation used
order for MC appr  =     50
iter  for MC appr  =     30
```

```
****************************************************************
Variable          Coefficient  Asymptot t-stat    z-probability
constant            -0.093823        -1.100533         0.271100
age                  0.553216         7.664817         0.000000
age2                -1.324653       -11.221285         0.000000
age3                 0.404283         6.988042         0.000000
lotsize              0.121762        26.722881         0.000000
rooms                0.004331         1.730088         0.083615
tla                  0.583322        74.771529         0.000000
beds                 0.017392         4.465188         0.000008
W-age                0.129150         1.370133         0.170645
W-age2               0.247994         1.507512         0.131679
W-age3              -0.311604        -3.609161         0.000307
W-lotsize           -0.095774       -17.313644         0.000000
W-rooms             -0.015411        -2.814934         0.004879
W-tla               -0.109262        -6.466610         0.000000
W-beds              -0.043989        -5.312003         0.000000
rho                  0.690597       245.929569         0.000000

sdm mean absolute prediction error =   0.2031
```

References

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

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

Bavaud, Francois, (1998), "Models for Spatial Weights: A Systematic Look," *Geographical Analysis*, Volume 30, pp. 153-171.

Ord, J.K., (1975), "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, Volume 70, pp. 120-126.

LeSage, James P., and R. Kelley Pace (2001), "Spatial Dependence in Data Mining," Chapter 24 in *Data Mining for Scientific and Engineering Applications*, Robert L. Grossman, Chandrika Kamath, Philip Kegelmeyer, Vipin Kumar, and Raju R. Namburu (eds.), Kluwer Academic Publishing.

Pace, R., Kelley, and Ronald Barry. (1997), "Quick computation of spatial autoregressive estimators," *Geographical Analysis*, Volume 29, pp. 232-246.

Pace, R., Kelley, and James P. LeSage (2001), "Spatial Estimation Using Chebyshev Approximation", unpublished manuscript available at:
www.econ.utoledo.edu/faculty/lesage/workingp.html

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*, Volume 35, p. 301-319.

Sun, D., R.K. Tsutakawa, P.L. Speckman. (1999), "Posterior distribution of hierarchical models using car(1) distributions", *Biometrika*, Volume 86, pp. 341-350.