

Spatial Statistics Toolbox 2.0

R. Kelley Pace¹
LREC Endowed Chair of Real Estate
Department of Finance
E.J. Ourso College of Business Administration
Louisiana State University
Baton Rouge, LA 70803-6308
OFF: (225)-578-6256, FAX: (225)-578-6095
kelley@pace.am, www.spatial-statistics.com

February 15, 2003

¹I have many individuals and organizations to thank for supporting the toolbox. First, I would like to gratefully acknowledge the research support received from the National Science Foundation (BCS-0136193 and BCS-0136229). Note, Any opinions, findings and conclusions or recommendations expressed in this material are mine and do not necessarily reflect the views of the National Science Foundation (NSF). Second, I would like to thank James LeSage (www.spatial-econometrics.com) for his remarks and advice and Ron Barry at the University of Alaska for help with the previous version. Third, I would like to gratefully acknowledge the research support received from Louisiana State University. Finally, I would like to thank Ming-Long Lee, Darren Hayunga, and Baris Kazar for their help.

Contents

1	Why the Toolbox Exists	5
2	Using the Toolbox	8
2.1	Hardware and Software Requirements	8
2.2	Installation	8
2.3	Help and documentation	10
2.4	Known Limitations	11
2.5	Tips on Using the Toolbox	11
2.6	Included Examples	13
2.7	Included Datasets	14
2.8	Included Manuscripts	14
3	A Brief, Selected Tour of the Toolbox	15
4	References	35

List of Tables

3.1	SAR Estimation Results using Chebyshev Indet approximation and likelihood dominance inference	22
3.2	SAR Estimation results using exact Indet	22
3.3	Estimates on Election Data	24
3.4	Signed Root Deviances using Election Data	25
3.5	Timing for operations to Election data (n=3,107)	26
3.6	Distribution of local estimates	31
3.7	Times for Different Methods for 57,647 Observations	32
3.8	Likelihoods across ρ for Doubly Stochastic Scaling	33
3.9	Times for Optimizing the Likelihood over ρ for Both Scalings	33
3.10	Likelihoods Across Doubly Stochastic and Regular Scalings	34
3.11	OLS versus MESS Results Using Optimal ρ for Doubly Stochastic Scaling	34

List of Figures

3.1	Connections among Counties via Delaunay	16
3.2	Connections among Counties via Eight Nearest Neighbors . . .	17
3.3	Plot of Non-zeros of Delaunay Weight Matrix	18
3.4	Plot of Non-zeros of Permuted Delaunay Weight Matrix	19
3.5	Plot of Exact log-determinant with Chebyshev Approximation and Taylor Bounds	20
3.6	Plot of Exact log-determinant with Monte Carlo Approximation and Limits	21
3.7	SAR Profile Likelihoods by Model	23
3.8	Plot of Fringe prediction error versus subsample size	27
3.9	Plot of prediction error on center of area versus subsample size	28
3.10	Spatial dependence parameter estimate versus subsample size	29
3.11	Map of influence of homeownership on voting	30

Chapter 1

Why the Toolbox Exists

Individual data arise at a time and a place. Randomization can destroy and aggregation can obscure spatial and temporal information, but the original data points potentially exhibit spatiotemporal dependence. Often simple models fitted to these data will produce spatially, temporally, or spatiotemporally correlated errors, provided reality is more complex than the model. Ignoring the spatial, temporal, or spatiotemporal dependence among errors results in inefficient parameter estimation, biased inference, and ignores information that can greatly improve prediction accuracy. In addition, if the data generating process is an autoregressive one, ignoring the dependence will lead to biased estimates and inference.

Historically, spatial statistics software floundered with problems involving even thousands of observations. For example, Li (1995) required 8,515 seconds to compute a 2,500 observation spatial autoregression using an IBM RS6000 Model 550 workstation. The culprit for the difficulty lies in the maximum likelihood estimator's need for the determinant of the n by n matrix of the covariances among the spatially scattered observations.

The conventional computational approach relies upon eigenvalues (Ord (1975)). Even with faster computers the calculation of eigenvalues requires substantial time and memory. For example, on a 1700 Athlon it requires 24.75 minutes to compute the eigenvalues of the spatial weight matrix based

on 30 nearest neighbors. The resulting matrix takes 77 megabytes of storage as well.

However, many problems of practical importance generate large spatial data sets. Obvious examples include census data (over 200,000 block groups for the US) and housing sales (many millions sold per year). The Spatial Statistics Toolbox addresses the need to quickly estimate large problems. In contrast to the eigenvalue approach, by focusing upon direct computation of determinants and using sparsity, the same operation takes 3.38 seconds in the toolbox. Using a Chebyshev approximation reduces the time to under 0.2 seconds. As the eigenvalue computations rise at the cube of n while the log-determinant functions in the toolbox rise with n or $n \ln(n)$, large problems further increase the performance of the toolbox relative to conventional approaches.

The speed of Spatial Statistics Toolbox 2.0 permits users to explore alternative specifications, spatial and aspatial, in a timely fashion. For example, one can estimate local spatial autoregressions (spatial autoregressive local estimation or SALE) as in Pace and LeSage (forthcoming). In the SALE example provided, the toolbox was able to estimate a sequence of over 400 spatial autoregressions for around each of 3,107 points in under four minutes.

The Spatial Statistics Toolbox 2.0 conserves on memory as well. It is not difficult to estimate a spatial autoregression with over one million observations. In fact, the toolbox provides an example under the dataset directory whereby a one million observation spatial autoregression is estimated in just under 20 seconds. It took 130.63 seconds to find the weight matrix, 60.24 seconds to simulate the dependent variable, and 19.42 seconds to estimate the autoregression.

Relative to the previous incarnation, the Spatial Statistics Toolbox 2.0 introduces multidimensional and spatiotemporal weight matrices, new scalings of the weight matrices (doubly stochastic), faster computation of exact log-determinants (actually interpolated from exact computations of a smooth function), a couple of approximations to log-determinants, and some new

computationally and theoretically interesting models such as the aforementioned SALE and the matrix exponential spatial specification (MESS).

Chapter 2

Using the Toolbox

2.1 Hardware and Software Requirements

The toolbox has been developed under 6.5 and tested under 6.5 and 6.1 across W2K and Windows ME. However, some routines run faster under 6.5 than under 6.1. The total installation takes around 15 megabytes. The routines have been tested on PC compatibles. The routines should run on other platforms, but have not been tested on non-PC compatibles.

2.2 Installation

For users who can extract files from zip archives, follow the instructions for your product (e.g., Winzip) and extract the files into the drive in which you wish to install the toolbox. The installation program will create the following directory structure in whichever drive you choose.

```
drive_letter\space\...
+---articles
+---datasets
    +---big_one
    +---election
```



```
    +---housing
    +---space_time
+---documentation
+---examples
    +---CAR
    +---CAR_SIM
    +---CHEBYSHEV
    +---CHEBYSHEV_SEQUENCE
    +---CLOSEST_NEIGHBOR
    +---DELAUNAY2
    +---DOUBLY
    +---LNDET_INTERP
    +---LNDET_INTERP_SEQUENCE
    +---LNDET_MONTECARLO
    +---MESS_AR
    +---MESS_CAR
    +---MESS_SIM
    +---MIXED
    +---MULTIVARIATE
    +---NEAREST_NEIGHBORS
    +---OLS
    +---SALE
    +---SAR
    +---SAR_SIM
    +---SPACE_TIME
        +---fmex_code
+---functions
    +---fmex_code
```

To see whether the installation has succeeded, change the directory in Matlab to one of the supplied examples and type `run x_m`. For ex-

ample, go to the `...\examples\car` subdirectory and type `run x_car2_ga1`. The use of `...` in this context serves as a wildcard or placeholder for the intervening characters. This should cause the example script `x_car2_ga1.m` to run.

The multidimensional weight matrix routines require the installation of TSTOOL. Go to www.physik3.gwdg.de/tstool/ to find this useful package and follow the instructions to install it in Matlab. If you have Matlab 6.5, you can easily add the relevant paths to the mex functions by going to the File menu, selecting Set Path, under the applet selecting Add Folder, and add the paths so Matlab can find the functions. On my machine I added:

```
...\opentstool\tstoolbox\mex\dll
...\opentstool\tstoolbox\mex
...\opentstool\tstoolbox\utils
...\tstool\opentstool\tstoolbox\gui
...\tstool\opentstool\tstoolbox
```

2.3 Help and documentation

All the example scripts should follow the form `x_... .m` (e.g., `x_car2_ga1.m`, `x_sar2_ga1.m`). Functions follow the form `f_... .m` (e.g., `fsar2.m`, `fols2.m`). Matlab matrix files (which may include multiple matrices) have the form `... .mat`.

If you wish to access the functions in other directories, you can define a path to `...\space` and `...\space\functions` using the `set path` command under the file menu in Matlab 6.5, where `...` refers to the drive where the spatial statistics toolbox resides and provided `space` is the name of the directory you chose. This is the same procedure outlined in installing TSTOOL. On my machine I added:

```
...\space
...\space\functions
```

As in previous versions of Matlab, users can employ the `addpath` command to add these directories to the search path.

Both the functions and examples are well-documented internally. One can use the help functions of Matlab in the usual fashion, when the toolbox is on the search path or the functions are in the current directory. If the toolbox is on the search path, a user can type `help space` and see the functions contained in the Spatial Statistics Toolbox 2 (provided `space` is the name of the directory you chose). If the toolbox is on the search path, or the function lies in the current directory, a user can obtain help on the function by typing `help function_name` or `doc function_name`.

The examples provide the best means of understanding the toolbox functions. Each subdirectory under the examples subdirectory contains all the files needed for that particular example (except the multivariate subdirectory which requires installation of the TSTOOL package). Users can substitute their data for the example data, and see how the functions perform for their problem. Note, a few of these functions take several minutes to run (notably the SALE examples). Most, however, are quite fast.

2.4 Known Limitations

None. All are unknown!

2.5 Tips on Using the Toolbox

Typical sessions with the toolbox proceed in four steps. First, import the data into Matlab. If the file is fixed-format or tab-delimited ASCII, the command `load name.extension` (whatever that `name.extension` may be) will load the filename contents into memory into a Matlab variable `name`. Saving this will convert it into a matlab file (e.g., `save name name` will save variable `name` into matrix `name` stored in `name.mat`). Failure to specify both names will

result in saving all defined variables into one file. The data would include the dependent variable, the independent variables, and the locational coordinates. For example, suppose the user has the dependent variable in the text file `y.txt`. Issuing the command `load y.txt` results in a Matlab variable `y`. Issuing the command `save y y` results in saving `y.mat` in the directory.

Second, create a spatial weight matrix. Users can choose weight matrices based upon nearest neighbors (symmetric or asymmetric), multidimensional symmetric neighbors, spatiotemporal neighbors (asymmetric), and Delaunay triangles (symmetric). In almost all cases, one must make sure each location is unique. One may need to add slight amounts of random noise to the locational coordinates to meet this restriction (some of the latest versions of Matlab do this automatically – do not dither the coordinates in this case). Note, some estimators only use symmetric matrices. You can specify the number of neighbors used and their relative weightings.

Note, the Delaunay spatial weight matrix leads to a concentration matrix or a variance-covariance matrix that depends upon only one-parameter (α , the autoregressive parameter). In contrast, the nearest neighbor concentration matrices or variance-covariance matrices depend upon three parameters (α , the autoregressive parameter; m , the number of neighbors; and ρ , which governs the rate weights decline with the order of the neighbors with the closest neighbor given the highest weighting, the second closest given a lower weighting, and so forth). Three parameters should make this specification sufficiently flexible for many purposes.

Third, one computes the log-determinants for a grid of autoregressive parameters (prespecified by the routine as a default, or specified by the user as an option). Determinant computations proceed faster for symmetric matrices. You must choose the appropriate log-determinant routines for the type of spatial weight matrix you have specified. Computing the log-determinants is the slower than estimation, but only needs to be done when changing the spatial weight matrix. For example, one can use the same weight matrix and log-determinant files when exploring transformations or specifications of the

dependent and independent variables (for the same observations).

Fourth, pick a statistical routine to run given the data matrices, the spatial weight matrix, and the log-determinant vector. One can choose among conditional autoregressions (CAR), simultaneous autoregressions (SAR), matrix exponential spatial specifications (MESS), mixed regressive spatially autoregressive estimators (which include pure autoregressive models and spatially lagged independent variable models as special cases), and OLS. In addition, one can explore multivariate, spatiotemporal, and multivariate estimation. These routines require little time to run. One can change models, weightings, and transformations and reestimate in the vast majority of cases without rerunning the spatial weight matrix or log-determinant routines (you may need to add another simple Jacobian term when performing weighting or transformations of the dependent variables). This aids interactive data exploration.

Fifth, these procedures provide a wealth of information. Many of these routines yield the profile likelihood in the autoregressive parameter for each submodel (corresponding to the deletion of individual variables or the spatial term). All of the inference, even for the OLS routine, uses likelihood ratio statistics in the form of signed root deviances. This is just the square root of twice the difference in likelihoods given the sign of the parameter estimate. It has a t -like interpretation (Chen and Jennrich (1996)). The use of signed root deviances (SRDs) facilitates comparisons among different models.

2.6 Included Examples

The Spatial Statistics Toolbox comes with many examples. These are found in the subdirectories under `...spatial_toolbox_2\examples`. To run the examples, change the directory in Matlab into the many subdirectories that illustrate individual routines. Look at the documentation in each example directory for more detail. Almost all of the specific models have examples. In addition, the simulation routine examples serve as minor Monte Carlo stud-

ies which also help verify the functioning of the estimators. The examples use the 3,107 observation dataset from the Pace and Barry (1997) *Geographical Analysis* article.

2.7 Included Datasets

The `...spatial_toolbox_2\datasets` subdirectory contains subdirectories with individual data sets in Matlab file formats as well as their documentation. The data sets include example programs and output. Note, due to the many improvements incorporated into the Spatial Statistics Toolbox over time, the running times have greatly improved over those in the articles.

Hopefully, these data sets should provide a good starting point for exploring applications of spatial statistics.

2.8 Included Manuscripts

In the manuscript subdirectory we provide pdf versions of the *Geographical Analysis* 1997 and 2000 articles. I would like to thank the publishers (Ohio State Press and Elsevier) for having given us copyright permission to distribute these works. One can also go to www.spatial-statistics.com to access some other articles (e.g., the *Linear Algebra and its Applications* article which proposed the Monte Carlo log-determinant estimator).

Chapter 3

A Brief, Selected Tour of the Toolbox

The weight matrix specifies the dependence among observations. One form of weight matrix (Delaunay) uses the notion of contiguity to specify dependence as depicted in Figure [3.1](#).

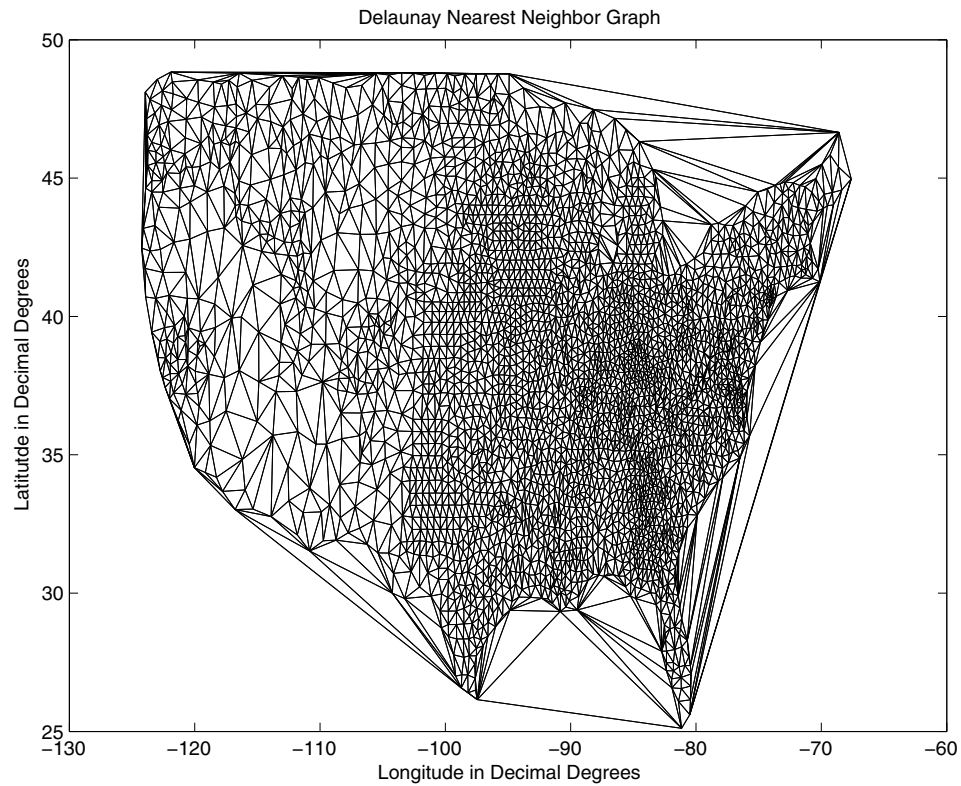


Figure 3.1: Connections among Counties via Delaunay

Note the somewhat strange behavior of connections to outlying observations in Figure 3.1. This arises to the geometric nature of contiguity. Using nearest neighbors based upon some metric can avoid this as shown in Figure 3.2.

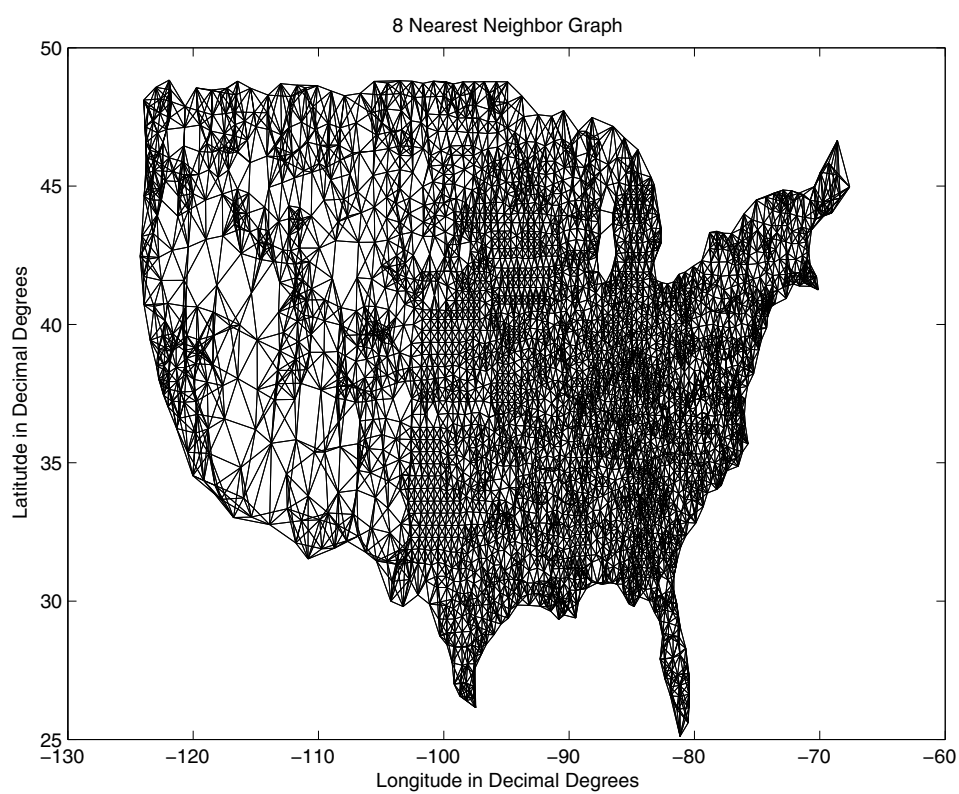


Figure 3.2: Connections among Counties via Eight Nearest Neighbors

Using only nearby observations implies that the weight matrix has main zeros or is sparse as shown in Figure 3.3.

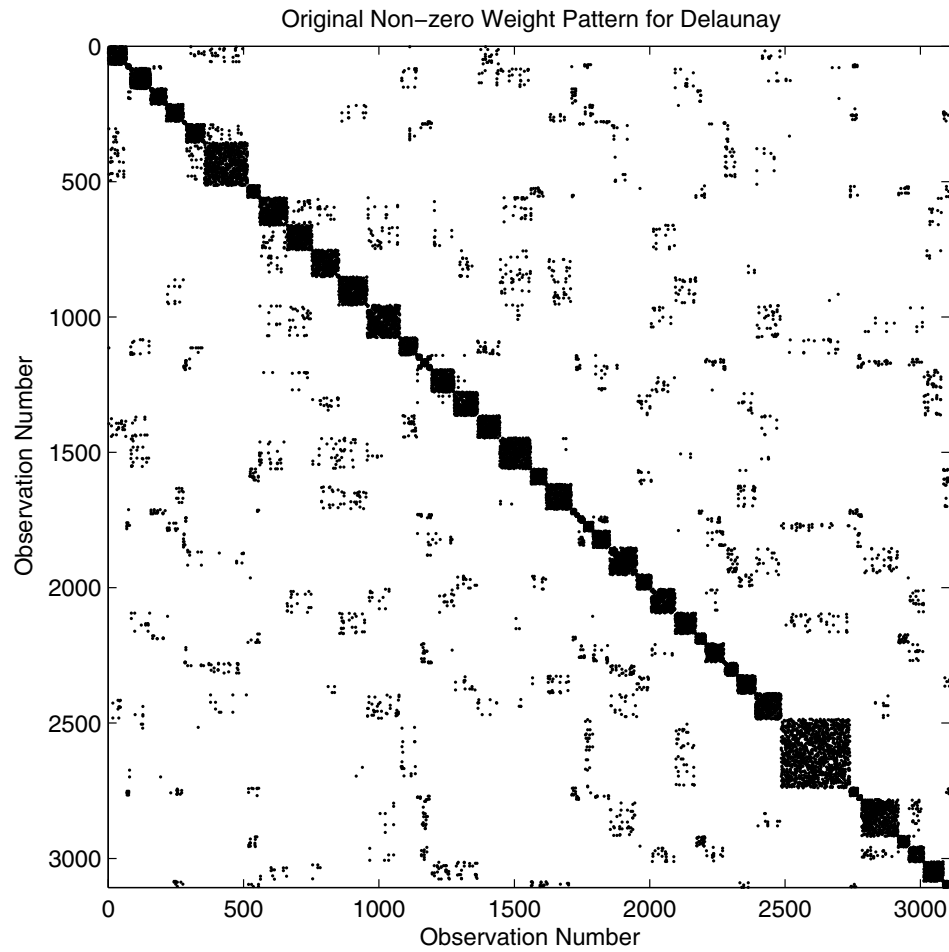


Figure 3.3: Plot of Non-zeros of Delaunay Weight Matrix

This becomes even more apparent when reordering the observations as in Figure 3.4.

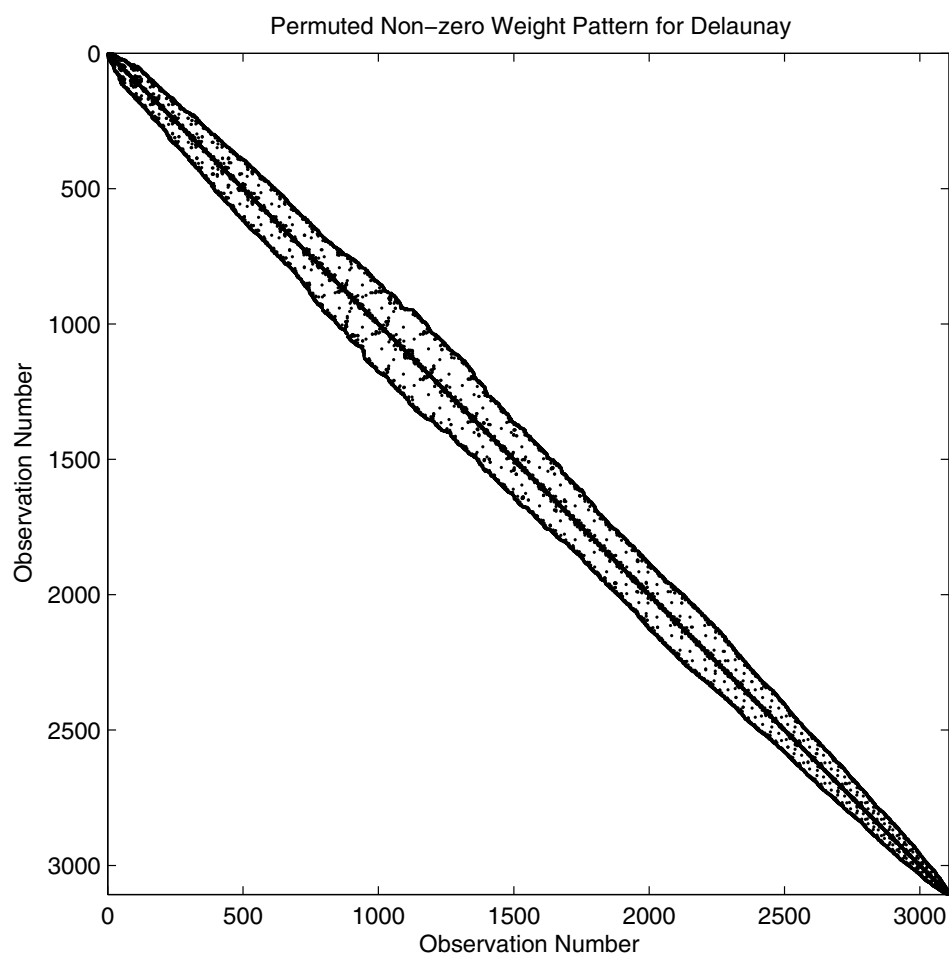


Figure 3.4: Plot of Non-zeros of Permuted Delaunay Weight Matrix

Sparsity as well as finding an appropriate ordering are key in quickly computing the log-determinants used in maximum likelihood. The toolbox has functions for exact computation of the log-determinants (actually interpolation of exact computations at various points). However, users can select approximations as well which depend only on sparsity and not upon orderings. The quadratic Chebyshev is the fastest and most approximate technique (Figure 3.5).

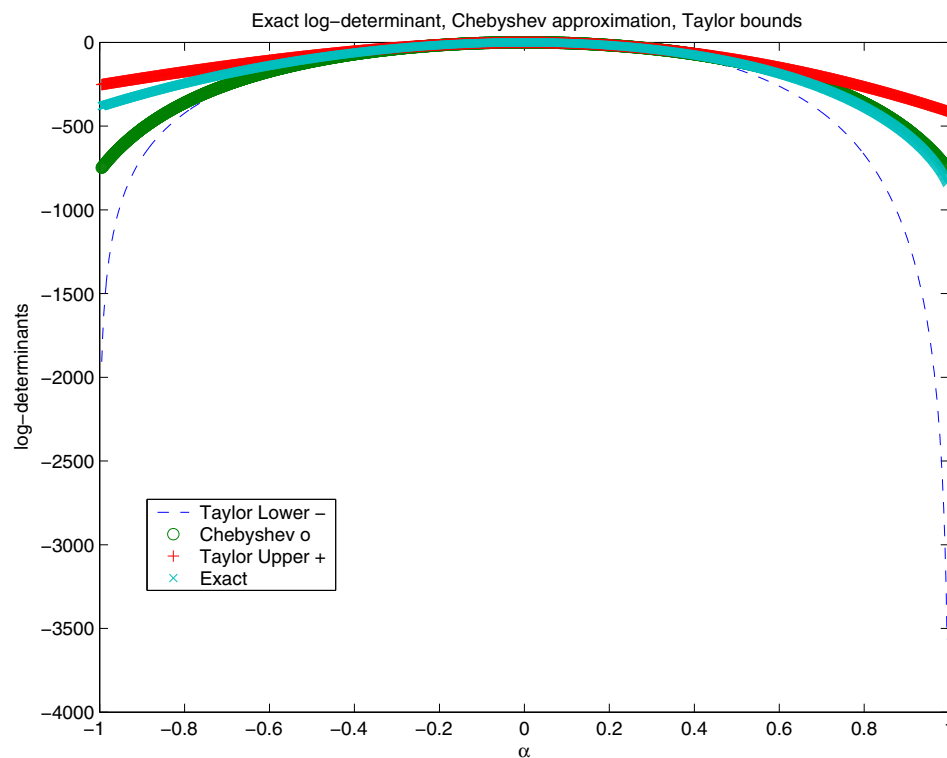


Figure 3.5: Plot of Exact log-determinant with Chebyshev Approximation and Taylor Bounds

The Chebyshev approximation appears quite good for positive, moderate values of the dependence parameter, but could use improvement for materially negative values of the spatial dependence parameter. Fortunately, such negative values seem rare in practice.

The Monte Carlo log-determinant estimator is quite fast, but more accurate (Figure 3.6).

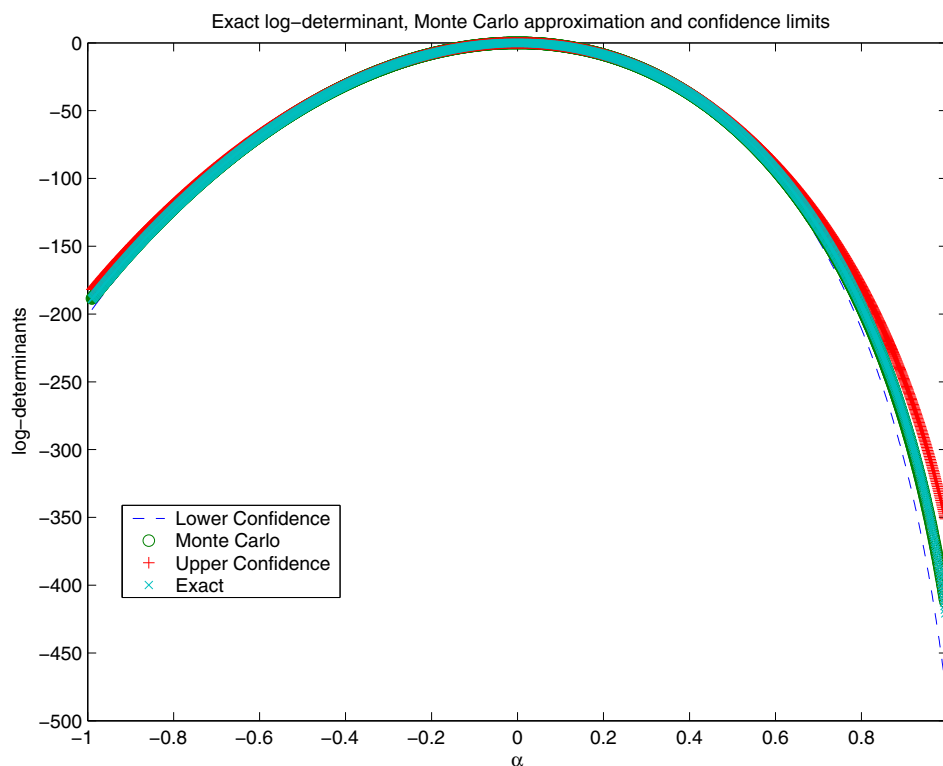


Figure 3.6: Plot of Exact log-determinant with Monte Carlo Approximation and Limits

To see the effects of exact versus approximate log-determinant computations, consider Tables 3.1 and 3.2 using the 3,107 county election data. The estimated autoregressive parameter is only off by 0.01 from using the approximation. The approximate method also uses likelihood dominance inference which results in a lower bound to the signed root deviances. As shown by the tables, the likelihood dominance SRDs are smaller in magnitude than the exact SRDs. However, they can still document statistical significance for many variables, and thus can prove useful in many circumstances.

Variables	Beta Estimates	Signed Root Deviances	PR of Higher SRDS
Voting Pop	-0.7784	-31.1227	0.0000
Education	0.2696	9.3438	0.0000
Home Ownership	0.4530	26.2661	0.0000
Income	0.0071	0.0035	0.9972
Intercept	0.5443	7.8773	0.0000
Alpha	0.7250	32.7719	0.0000

Table 3.1: SAR Estimation Results using Chebyshev Indet approximation and likelihood dominance inference

Variables	Beta Estimates	Signed Root Deviances	PR of Higher SRDS
Voting Pop	-0.7806	-32.1748	0.0000
Education	0.2746	11.9438	0.0000
Home Ownership	0.4525	27.0073	0.0000
Income	0.0047	0.2187	0.8269
Intercept	0.5528	9.4908	0.0000
Alpha	0.7150	34.8648	0.0000

Table 3.2: SAR Estimation results using exact Indet

Some of the routines not only yield the maximum of the likelihood function, but also profile likelihoods in the dependence parameter α by model as shown in Figure 3.7.

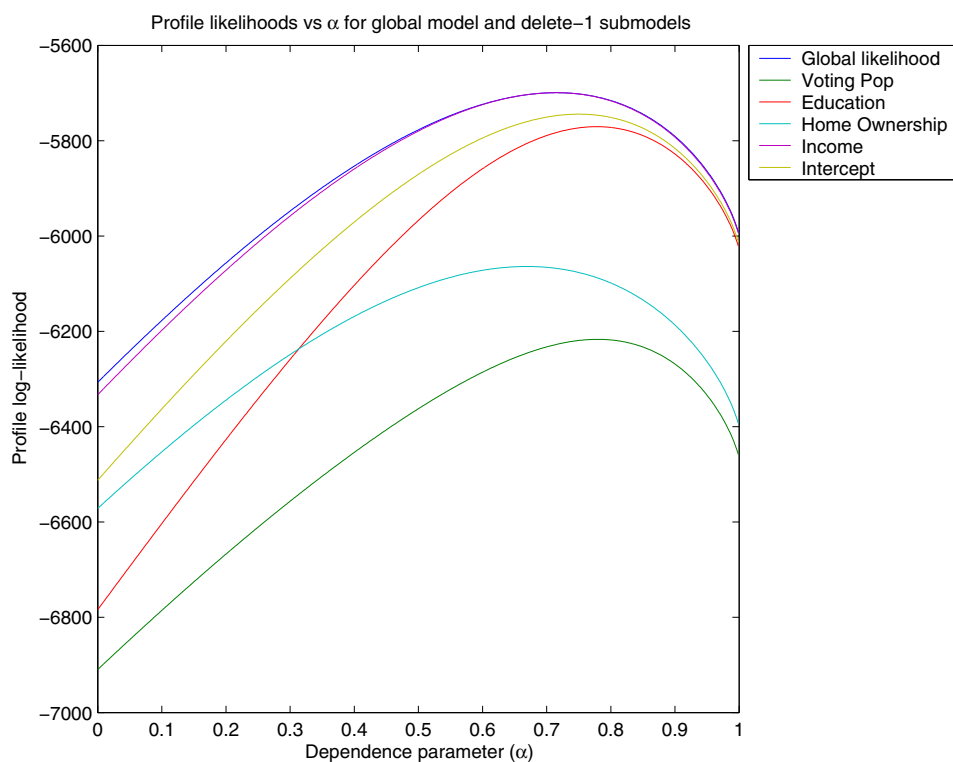


Figure 3.7: SAR Profile Likelihoods by Model

The toolbox includes SAR, CAR, and MESS error models as well as MESS, closest neighbor, and MIX autoregressive models as shown in Table 3.3 and Table 3.4.

Variables	b OLS	b closest	b MESS	b Mix
Voting Pop	-0.8464	-0.7489	-0.7693	-0.7298
Education	0.5167	0.2899	0.1941	0.1818
Home Ownership	0.4291	0.4457	0.4832	0.4580
Income	-0.1439	-0.0332	0.0423	0.0427
Lag Voting Pop	0.0000	0.1878	0.4186	0.4616
Lag Education	0.0000	0.0975	0.1205	0.0450
Lag Home Ownership	0.0000	-0.1569	-0.3249	-0.3299
Lag Income	0.0000	-0.1079	-0.1802	-0.1410
Intercept	0.9814	0.7495	0.5636	0.4205
α	0.0000	0.3352	1.4628	0.6550

Table 3.3: Estimates on Election Data

The closest neighbor approach is intermediate to a non-spatial approach (OLS) and a full spatial approach (MESS) or the approximate mixed routine. Note, the close agreement between MESS and the mixed routine. Note, OLS in this case uses the spatial averages of the basic independent variables as additional independent variables.

None of these operations take long for the election data.

Variables	b OLS	b closest	b MESS	b Mix
Voting Pop	-34.7211	-30.5020	-28.8899	-29.4018
Education	30.8928	13.4922	7.6654	7.7169
Home Ownership	23.0066	25.4083	26.8836	27.3515
Income	-7.2373	-1.5742	1.7478	1.8952
Lag Voting Pop	0.0000	7.7600	10.7656	12.5507
Lag Education	0.0000	4.5090	3.9785	1.5705
Lag Home Ownership	0.0000	-9.4245	-11.1259	-12.1200
Lag Income	0.0000	-5.1399	-5.5551	-4.6603
Intercept	20.2782	16.0533	10.7480	8.4626
α	0.0000	24.2762	31.5948	33.7097

Table 3.4: Signed Root Deviances using Election Data

Operation	Timings in seconds
OLS	0.0150
Closest AR	0.0470
MESS	0.0940
Exact Log-det	0.8750
Mix	0.0470
Doubly Stochastic Scaling	0.2660

Table 3.5: Timing for operations to Election data (n=3,107)

In addition to global models, the toolbox has spatial autoregressive local estimation (SALE). The user chooses the bandwidth (subsample size) by examining cross-validation error at the fringe observations as in Figure 3.9 or each observation as in Figure 3.8.

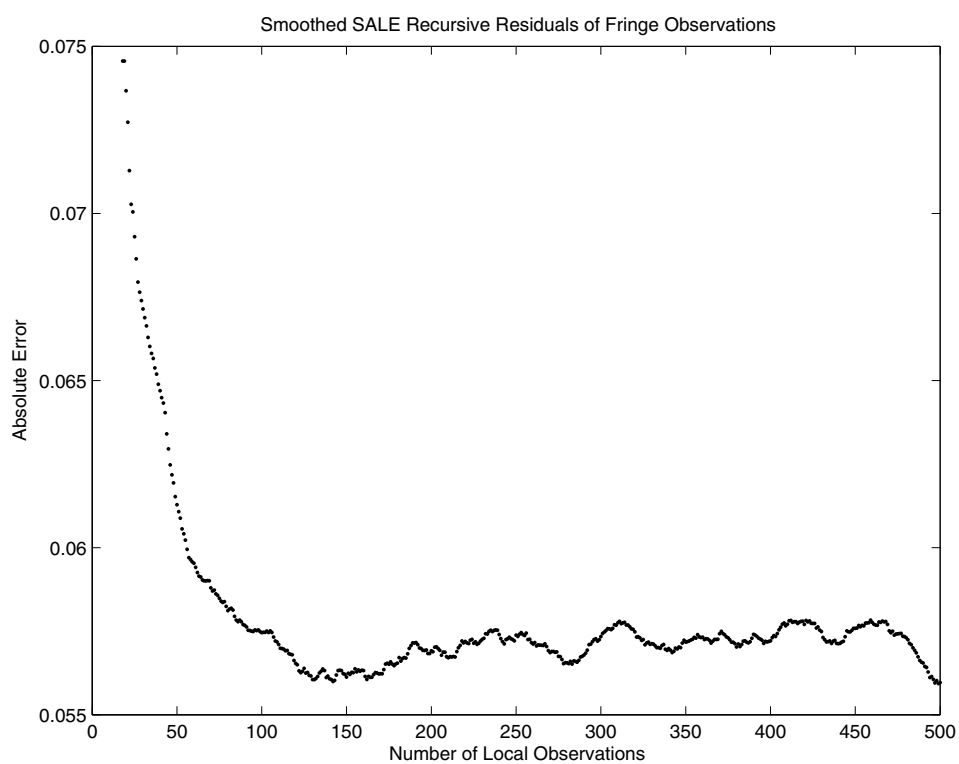


Figure 3.8: Plot of Fringe prediction error versus subsample size

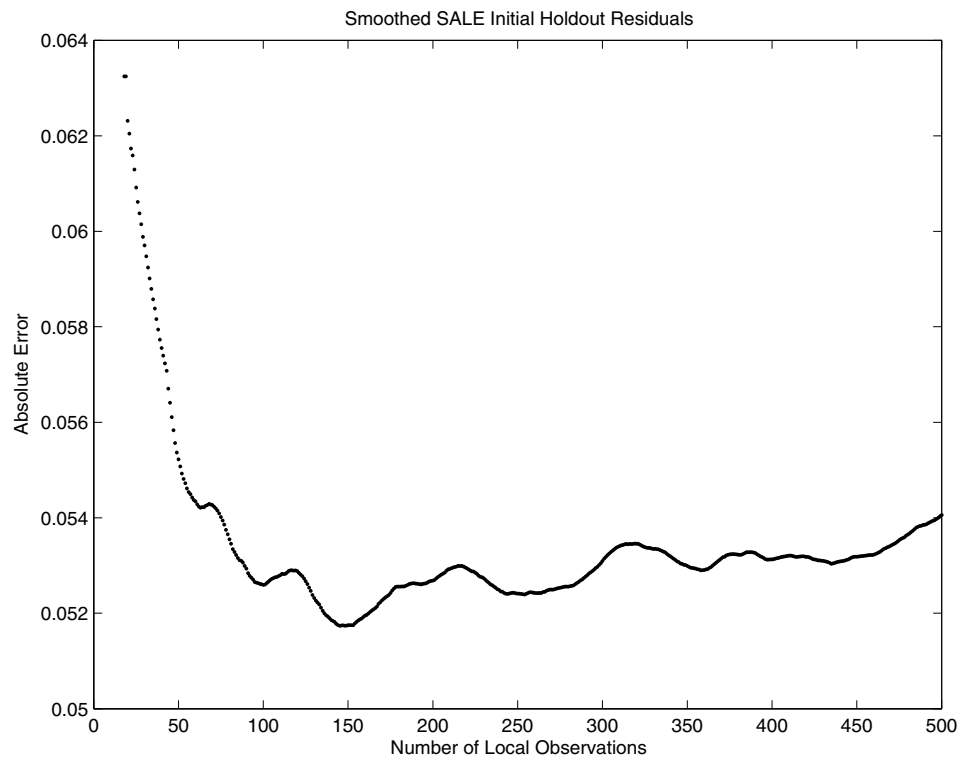


Figure 3.9: Plot of prediction error on center of area versus subsample size

Usually there is spatial dependence even in small subsamples as shown in Figure [3.10](#).

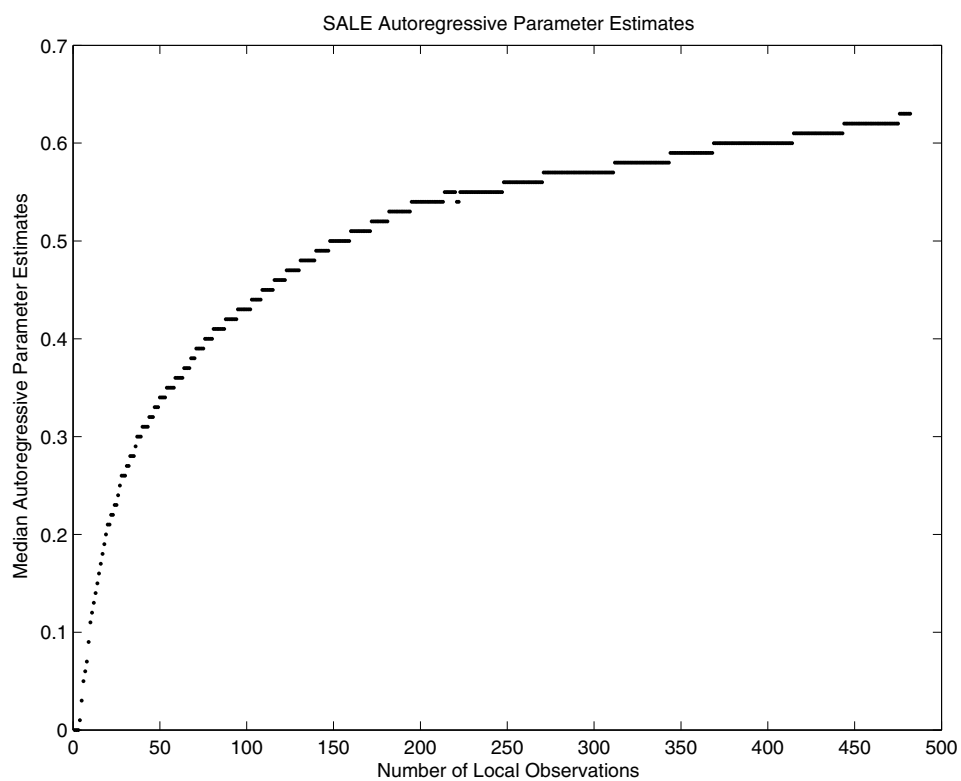


Figure 3.10: Spatial dependence parameter estimate versus subsample size

Local estimation leads to spatially varying parameter estimates, such as those shown in [Figure 3.11](#).

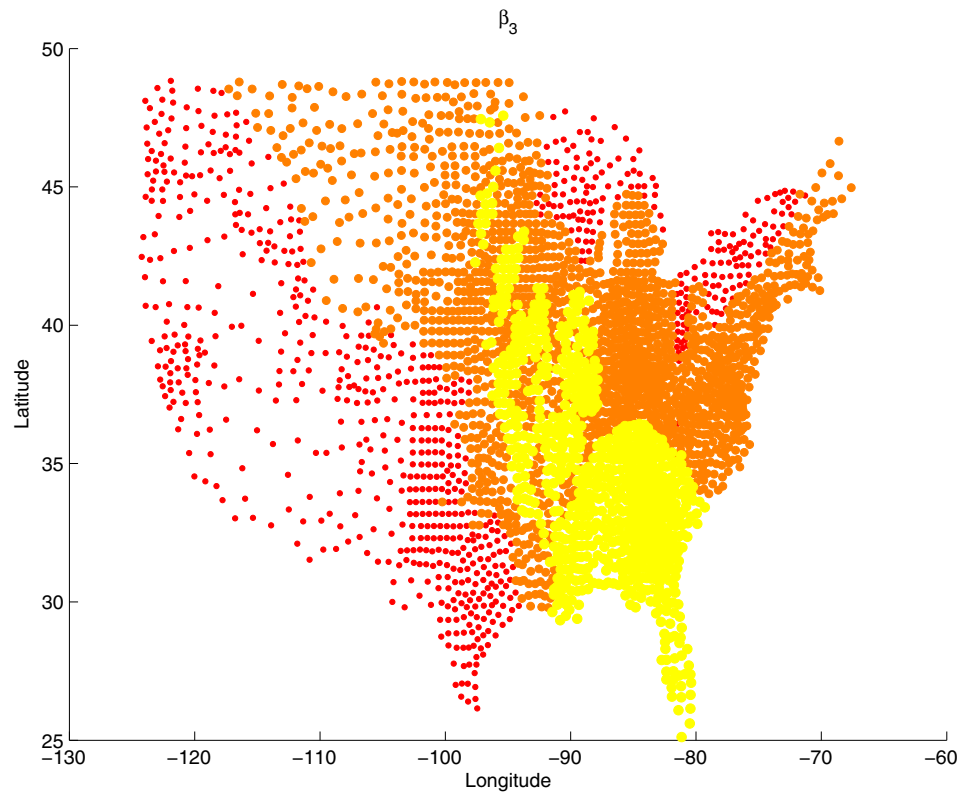


Figure 3.11: Map of influence of homeownership on voting

In addition, a user can obtain an idea of the sensitivity of parameter estimates to spatial variation such as summarized in Table 3.6.

Percentiles	0	10	25	50	75	90	100
α	0.3100	0.4300	0.5500	0.6300	0.6700	0.7100	0.7800
Voting Pop	-1.0079	-0.8875	-0.8075	-0.6962	-0.5967	-0.5299	-0.4066
Education	-0.1871	0.0517	0.1142	0.1788	0.3135	0.4923	0.7118
Home Ownership	0.1530	0.2964	0.3419	0.4253	0.5161	0.7660	0.8461
Income	-0.2576	-0.1644	-0.1072	-0.0473	0.0628	0.1448	0.2325
Lag Voting Pop	-0.1525	0.0655	0.2868	0.4527	0.5490	0.6523	0.8806
Lag Education	-0.5734	-0.3666	-0.1812	-0.0299	0.0677	0.1322	0.3318
Lag Home Ownership	-0.7010	-0.4499	-0.3278	-0.2337	-0.1100	-0.0382	0.1880
Lag Income	-0.4764	-0.2374	-0.1733	-0.0897	0.0216	0.0862	0.2786
Intercept	-0.1605	0.0473	0.1463	0.2847	0.6612	1.0772	1.8902

Table 3.6: Distribution of local estimates

Method	Time in seconds
OLS	0.1560
Closest AR	0.1090
MESS	0.6090
Approximate Mix	0.3280
Doubly Stochastic Scaling	7.1880
Delaunay Weight Matrix	5.4220

Table 3.7: Times for Different Methods for 57,647 Observations

To provide an idea about the performance of the techniques for a larger problem, we estimated a simple hedonic regression over US census tracts. This resulted in 57,647 observations. Table 3.7 shows the timings for some of the various operations. All of these seem quite fast. A user can find a Delaunay weight matrix and estimate a spatial autoregression in under 10 seconds on desktop machines.

Just selecting a particular weight matrix seems arbitrary. Here we take 30 nearest neighbors and weight these geometrically. A ρ of 1 indicates no decline in the weight given to further neighbors relative to closer ones, while a ρ of 0.5 would give half the weight to the second nearest neighbor as it would to the first nearest neighbor. Thus, ρ allows changes in the effective number of neighbors used without actually varying the number of neighbors. It often makes sense in this approach to set the number of neighbors to a fairly high level (such as 30). Table 3.8 shows the effect of varying ρ on the profile log-likelihood. Small changes in ρ make large changes in the profile log-likelihood, evidence of the importance of this parameter.

It did not take overly long to find the nearest neighbors or the optimal ρ , even with doubly stochastic scalings of the weight matrix as shown by Table 3.9.

These operations lead to a table of profile log-likelihoods (Table 3.10) across weight matrices. Examining the MESS loglikelihoods over ρ and Delaunay and contrasting it with the loglikelihood from applying OLS to the

ρ	log-likelihood
0.8000	-231,895.0426
0.8500	-230,968.8487
0.9000	-230,828.6413
0.9500	-231,699.1089
1.0000	-233,434.5371

Table 3.8: Likelihoods across ρ for Doubly Stochastic Scaling

Operation	Time in seconds
NN computation	31.4060
RS Time to find optimum ρ	56.7810
DS Time to find optimum ρ	94.8750

Table 3.9: Times for Optimizing the Likelihood over ρ for Both Scalings

basic non-spatial independent variables demonstrates that even a suboptimal choice of ρ or Delaunay still dominates the use of an aspatial model in this case and that optimizing over ρ dominates an arbitrary choice of weight matrices (Table 3.10). Moreover, the doubly stochastic (DS) scaling helped greatly for this example over the regular scaling (RS). In addition, inspection of aspatial OLS versus MESS with an optimal selection of ρ in Table 3.11 shows clear differences among the approaches. Note, the land area variable became insignificant after modeling space.

It is not difficult to estimate a spatial autoregression with over one million observations. In fact, the toolbox provides an example (big_one subdirectory) under the dataset directory whereby a one million observation spatial autoregression is estimated in just under 20 seconds. It took 130.63 seconds to find the weight matrix, 60.24 seconds to simulate the dependent variable, and 19.42 seconds to estimate the autoregression.

Variable	Value
Aspatial likelihood (OLS)	-266,505.1663
Closest Neighbor	-243,986.1676
RS Delaunay maximum likelihood	-244,509.8257
RS Maximum likelihood across ρ	-254,216.3172
RS Optimum ρ	0.9000
DS Delaunay maximum likelihood	-235,626.6213
DS Maximum likelihood across ρ	-230,828.6413
DS Optimum ρ	0.9000

Table 3.10: Likelihoods Across Doubly Stochastic and Regular Scalings

Variables	OLS	OLS SRD	b MESS	SRDS MESS
Land area	-0.0850	-96.8455	-0.0008	-0.7159
Pop	0.1146	36.8592	0.0239	12.7630
Per cap Income	1.0837	208.2192	0.6786	152.7645
Age	-0.1269	-34.2809	-0.1384	-45.0489
Lag Land area			-0.0178	-14.5650
Lag Pop			0.0165	5.2437
Per cap Income			-0.3702	-65.2683
Lag Age			0.1088	27.5807
Intercept	1.2236	22.9837	-0.5988	-15.2122
α			3.0986	253.2835
ρ (relative to $\rho = 1$)			0.90	72.1927
Nearest Neighbors			30	
parameters	5		11	

Table 3.11: OLS versus MESS Results Using Optimal ρ for Doubly Stochastic Scaling

Chapter 4

References

If you need to know more about spatial statistics or about some of the specific routines, you may wish to examine:

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

Barry, Ronald, and R. Kelley Pace, "A Monte Carlo Estimator of the Log Determinant of Large Sparse Matrices," *Linear Algebra and its Applications*, Volume 289, Number 1-3, 1999, p. 41-54.

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*, Volume 91, Number 435, p. 993-998.

Christensen, Ronald. (1991) *Linear Models for Multivariate, Time Series, and Spatial Data*. New York: Springer-Verlag.

Cressie, Noel A.C. (1993) *Statistics for Spatial Data*, Revised ed. New York. John Wiley.

Dubin, Robin A. (1988) "Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms," *Review of Economics and Statistics* 70, 466-474.

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

- LeSage, James and R. Kelley Pace, "Spatial Dependence in Data Mining," *Data Mining for Scientific and Engineering Applications*, Edited by Robert L. Grossman, Chandrika Kamath, Philip Kegelmeyer, Vipin Kumar, and Raju R. Namburu, Kluwer Academic Publishing, 2001.
- LeSage, James and R. Kelley Pace, "Spatial Probit and Tobit," *Spatial Statistics and Spatial Econometrics*, Edited by Art Getis, Palgrave, 2003.
- Li, Bin. (1995) "Implementing Spatial Statistics on Parallel Computers," in: Arlinghaus, S., ed. *Practical Handbook of Spatial Statistics* (CRC Press, Boca Raton), pp. 107-148.
- Ord, J.K. (1975). "Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association* 70, 120-126.
- Pace, R. Kelley, and Ronald Barry. (1997) "Fast CARs," *Journal of Statistical Computation and Simulation* 59, p. 123-147.
- Pace, R. Kelley, and Ronald Barry. (1997) "Quick Computation of Regressions with a Spatially Autoregressive Dependent Variable," *Geographical Analysis* 29, 232-247.
- Pace, R. Kelley, and Dongya Zou, "Closed-Form Maximum Likelihood Estimates of Nearest Neighbor Spatial Dependence," *Geographical Analysis*, Volume 32, Number 2, April 2000, p. 154-172.
- Pace, R. Kelley and Ronald Barry, O.W. Gilley, C.F. Sirmans, "A Method for Spatial-temporal Forecasting with an Application to Real Estate Prices," *International Journal of Forecasting*, Volume 16, Number 2, April-June 2000, p. 229-246.
- Pace, R. Kelley, and James P. LeSage, "Semiparametric Maximum Likelihood Estimates of Spatial Dependence," *Geographical Analysis*, Volume 34, Number 1, January 2002, p. 75-90.
- Pace, R. Kelley, and James LeSage, "Likelihood Dominance Spatial Inference," forthcoming *Geographical Analysis* in January 2003.
- Pace, R. Kelley, and James LeSage, "Spatial Autoregressive Local Estimation," *Spatial Statistics and Spatial Econometrics*, Edited by Art Getis, Palgrave, 2003.
- Pace, R. Kelley, and James LeSage, "Chebyshev Approximation of Log-determinants of Spatial Weight Matrices," forthcoming in *Computational Statistics and Data Analysis*.
- Ripley, Brian D. (1981) *Spatial Statistics*. New York. John Wiley.