Analyzing spatial autoregressive models using Stata

David M. Drukker

StataCorp

2009 Italian Stata Users Group meeting
November 19, 2009

Part of joint work with Ingmar Prucha of the University of Maryland
Funded in part by NIH grants 1 R43 AG027622-01 and 1 R43 AG027622-02.

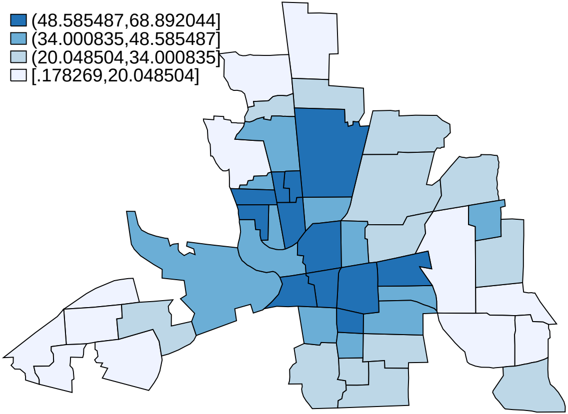
Outline

1. What is spatial data and why is it special?
2. Managing spatial data
3. Spatial autoregressive models

What is spatial data?

- Spatial data contains information on the location of the observations, in addition to the values of the variables

Property crimes per thousand households



Columbus, Ohio 1980 neighborhood data
Source: Anselin (1988)

Modeling Spatial Correlation

- Modeling correlation in unobservable errors
  - Efficiency and consistent standard errors
- Allowing outcome in place *i* to depend on outcomes in nearby places
  - Also known as state dependence or spill-over effects
  - Correction required for consistent point estimates
- Correlation is more complicated than time-series case
  - There is no natural ordering in space as there is in time
  - Space has, at least, two dimensions instead of one
    - Working on random fields complicates large-sample theory
- Models use a-priori parameterizations of distance
  - Spatial-weighting matrices parameterize Tobler's first law of geography Tobler (1970)

    "Everything is related to everything else, but near things are more related than distant things."

Managing spatial data

- Much spatial data comes in the form of shapefiles
  - US Census distributes shapefiles for the US at several resolutions as part of the TIGER project
    - State level, zip-code level, and other resolutions are available
  - Need to translate shapefile data to Stata data
  - User-written (Crow and Gould) `shp2dta` command
- Mapping spatial data
  - Mauricio Pisati wrote `spmap`
  - `http://www.stata.com/support/faqs/graphics/spmap.html` gives a great example of how to translate shapefiles and map data
- Need to create spatial-weighting matrices that parameterize distance

Shapefiles

- Much spatial data comes in the form of ESRI shapefiles
  - Environmental Systems Research Institute (ESRI), Inc. (`http://www.esri.com/`) make geographic information system (GIS) software
  - The ESRI format for spatial data is widely used
    - The format uses three files
    - The .shp and the .shx files contain the map information
    - The .dbf information contains observations on each mapped entity
  - `shp2dta` translates ESRI shapefiles to Stata format
  - Some data is distributed in the MapInfo Interchange Format
    - User-written command (Crow and Gould) `mif2dta` translates MapInfo files to Stata format

The Columbus dataset

- Anselin (1988) used a dataset containing information on property crimes in 49 neighborhoods in Columubus, Ohio in 1980
- Anselin now distributes a version of this dataset in ESRI shapefiles over the web
    - There are three files `columbus.shp`, `columbus.shx`, and `columbus.dbf` in the current working directory
    - To translate this data to Stata I used

```
. shp2dta using columbus, database(columbusdb) coordinates(columbuscoor)  ///
>        genid(id) replace
```

- The above command created `columbusdb.dta` and `columbuscoor.dta`
    - `columbusdb.dta` contains neighborhood-level data
    - `columbuscoor.dta` contains the coordinates for the neighborhoods in the form required spmap the user-written command by Maurizio Pisati
        - See also `http://econpapers.repec.org/software/bocbocode/s456812.htm`

Columbus data part II

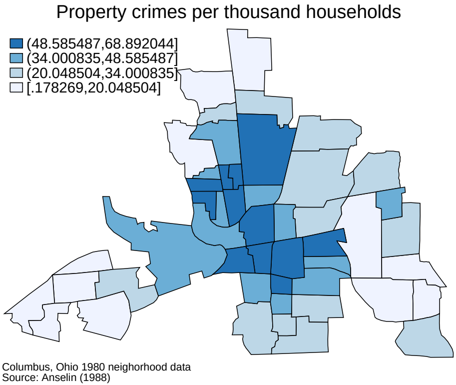
```
. use columbusdb, clear
. describe id crime hoval inc
                 storage  display     value
variable name     type    format      label       variable label
-----------------------------------------------------------------------------
id                byte    %12.0g                  neighborhood id
crime             double  %10.0g                  residential burglaries and
                                                    vehicle thefts per 1000
                                                    households
hoval             double  %10.0g                  housing value (in $1,000)
inc               double  %10.0g                  household income (in $1,000)
. list id crime hoval inc in 1/5

     +-----------------------------------------+
     | id       crime       hoval        inc   |
     |-----------------------------------------|
  1. |  1    15.72598   80.467003     19.531   |
  2. |  2   18.801754   44.567001     21.232   |
  3. |  3   30.626781       26.35     15.956   |
  4. |  4    32.38776   33.200001      4.477   |
  5. |  5    50.73151      23.225     11.252   |
     +-----------------------------------------+
```

Visualizing spatial data

- spmap is an outstanding user-written command for exploring spatial data

```
. spmap crime using columbuscoor, id(id) legend(size(medium)      ///
>         position(11)) fcolor(Blues)                             ///
>         title("Property crimes per thousand households")        ///
>         note("Columbus, Ohio 1980 neighorhood data" "Source: Anselin (1988)")
```



Modeling spatial data

- Cliff-Ord type models are used in many social-sciences
  - So named for Cliff and Ord (1973, 1981); Ord (1975)
  - The model is given by

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$
$$\mathbf{u} = \rho \mathbf{M}\mathbf{u} + \boldsymbol{\epsilon}$$

  where

    - $\mathbf{y}$ is the $N \times 1$ vector of observations on the dependent variable
    - $\mathbf{X}$ is the $N \times k$ matrix of observations on the independent variables
    - $\mathbf{W}$ and $\mathbf{M}$ are $N \times N$ spatial-weighting matrices that parameterize the distance between neighborhoods
    - $\mathbf{u}$ are spatially correlated residuals and ϵ are independent and identically distributed disturbances
    - λ and ρ are scalars that measure, respectively, the dependence of y_i on nearby y and the spatial correlation in the errors

Cliff-Ord models II

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{u}$$
$$\mathbf{u} = \rho \mathbf{M}\mathbf{u} + \epsilon$$

- Relatively simple, tractable model
- Allows for correlation among unobservables
  - Each u_i depends on a weighted average of other observations in $\mathbf{u}$
  - $\mathbf{Mu}$ is known as a spatial lag of $\mathbf{u}$
- Allows for y_i to depend on nearby y
  - Each y_i depends on a weighted average of other observations in $\mathbf{y}$
  - $\mathbf{Wy}$ is known as a spatial lag of $\mathbf{y}$
- Growing amount of statistical theory for variations of this model

Spatial-weighting matrices

- Spatial-weighting matrices parameterize Tobler's first law of geography Tobler (1970)
  "Everything is related to everything else, but near things are more related than distant things."
- Inverse-distance matrices and contiguity matrices are common parameterizations for the spatial-weighting matrix
  - In an inverse-distance matrix W, $w_{ij} = 1/D(i,j)$ where $D(i,j)$ is the distance between places i and j
  - In a contiguity matrix W,

$$w_{i,j} = \begin{cases} d_{i,j} & \text{if } i \text{ and } j \text{ are neighbors} \\ 0 & \text{otherwise} \end{cases}$$

where $d_{i,j}$ is a weight

Parameterizing spatial-weighting matrices

- Restricting the number of neighbors that affect any given place reduces dependence
- Restricting the extent to which neighbors affect any given place reduces dependence
- Contiguity matrices only allow contiguous neighbors to affect each other
  - This structure naturally yields spatial-weighting matrices with limited dependence
- Inverse-distance matrices sometimes allow for all places to affect each other
  - These matrices are normalized to limit dependence
  - Sometimes places outside a given radius are specified to have zero affect, which naturally limits dependence

Spatial-weighting matrices parameterize dependence

- The spatial-weighting matrices parameterize the spatial dependence, up to estimable scalars
- If there is too much dependence, existing statistical theory is not applicable
- Older literature used a version of “stationarity”, newer literature uses easier to interpret restrictions on $\mathbf{W}$ and $\mathbf{M}$
  1. All the diagonal elements of $\mathbf{W}$ and $\mathbf{M}$ are zero
  2. The matrices $(\mathbf{I} - \lambda\mathbf{W})$ and $(\mathbf{I} - \rho\mathbf{M})$ are nonsingular for the λ and ρ in specified intervals
  3. The row and column sums of $\mathbf{W}$, $\mathbf{M}$, $(\mathbf{I} - \lambda\mathbf{W})$, and $(\mathbf{I} - \rho\mathbf{M})$ are bounded uniformly in absolute value
- Restriction 1 is just a normalization rule

Intuition for these restrictions on spatial-weighting matrices

- The model is a pair of simultaneous equation systems

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{u}$$
$$\mathbf{u} = \rho \mathbf{M}\mathbf{u} + \epsilon$$

- To work with this model, we must be able solve these equations

$$\mathbf{y} = (\mathbf{I} - \lambda \mathbf{W})^{-1}\mathbf{X}\beta + (\mathbf{I} - \lambda \mathbf{W})^{-1}\mathbf{u}$$
$$\mathbf{u} = (\mathbf{I} - \rho \mathbf{M})^{-1}\epsilon$$

  which clearly requires that $\mathbf{I} - \lambda \mathbf{W}$ and $\mathbf{I} - \rho \mathbf{M}$ be nonsingular
- The restrictions on the row and column sums of the matrices ensures that products of these matrices are finite

Normalization the spatial-weighting matrices

- Normalizing the spatial-weighting matrices by a scalar fixes the scale of λ and ρ
- Normalizing by a vector, say a vector of row sums, changes more than the scale of the parameters
  - In row-sum normalization, $w_{ij} = (1/s_i)w_{ij}^*$, where $s_i = \sum_{j=1}^n |w_{ij}^*|$
  - Each row is normalized by a different scalar, s_i
- Spectral or min-max normalizations may be easier to interpret than the traditional row normaliztion
  - Spectral normalization set $w[i,i] = (1/\tau)w^*[i,j]$ where τ is the largest of the moduli of the eigenvalues of the unnormalized spatial-weighting matrix $\mathbf{W}^*$
  - Min-max normalization approximates the largest modulus

$$\tau = \min\left\{ \max_{1\leq i\leq n} \sum_j^n |w_{ij}^*|, \max_{1\leq j\leq n} \sum_j^n |w_{ij}^*| \right\}$$

The no-uniform-weights condition

- Kelejian and Prucha (2002) show that the spatial-weighting matrix cannot be a uniform weight matrices in which $w_{ij} = c$
- A uniform-weight matrix yields a spatial lag of y that is collinear with the constant term
  - If $w_{ij} = c$, $\mathbf{Wy} = \begin{pmatrix} nc\sum_{i=1}^{n} y_i \\ \vdots \\ nc\sum_{i=1}^{n} y_i \end{pmatrix}$ which is perfectly collinear with the constant term
- In practice, the result indicates that there must be sufficient variation in the elements of $\mathbf{W}$ to ensure sufficience variation in $\mathbf{Wy}$

Creating and Managing spatial weighting matrices in Stata

- There is a forthcoming user-written command by David Drukker, Hua Peng, and Rafal Raciborski called `spmat` for creating spatial weighting matrices
  - `spmat` uses variables in the dataset to create a spatial-weighting matrix
  - `spmat` can create inverse-distance spatial-weighting matrices and contiguity spatial-weighting matrices
  - `spmat` can also save spatial-weighting matrices to disk and read them in again
  - `spmat` can also import spatial-weighting matrices from text files
  - `spmat` can provide intensity plots and summary statistics of spatial-weighting matrices

Creating and Managing spatial weighting matrices in Stata

- In the examples below, we create a contiguity matrix and two inverse-distance matrices that differ only in the normalization

```
. spmat contiguity idmat_c using columbuscoor, id(id)
. spmat idistance  idmat_mmax, id(id) coordinates(x y) normalize(row)
. spmat idistance  idmat_spec, id(id) coordinates(x y) normalize(spectral)
```

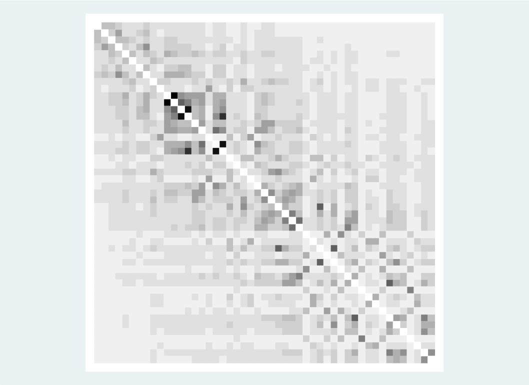
Intensity plot

- An intensity plot displays the intensity of the elements of a matrix in a two-dimensional graph
- The y-axis corresponds to the rows and the x-axis corresponds to the columns
    - An x-y point identifies an element in the matrix
    - The intensity of the gray-scale color describes the size of the matrix element
    - The intensity of the (1,1) element of the matrix is in the top-left of the graph
    - The intensity of the (n,n) element of the matrix is in the bottom-right of the graph

Intensity plot

- Assign the matrix values to B bins
  - Zero values get their own bin , coded as white
  - The nonzero values are spread uniformly over the remaining bins, higher values are assigned to darker colors

```
. spmat graph idmat_mmax, name(mmax)
```



Summarizing a spatial-weighting matrix

```
. spmat summarize idmat_mmax
Summary of spatial-weighting object idmat_mmax
```

Matrix	Description
Dimensions	49 x 49
# of zeros	49
Minimum	0
Maximum	.1758263
Mean	.01881
Median	.013461
Symmetric	yes

Sorting induces banded structure

- Most spatial-weighting matrices should be banded
- Drukker et al. (2009b) show that sorting the data on the distance from one place before creating the spatial-weighting matrix will cause many spatial-wieghting matrices to have a banded structure
- `spmat` will be able to store the matrix as banded
  - Reduces memory from $N * N$ elements to $N * (b_u + b_L + 1)$, where b_U and b_L are the upper and lower bandwidths
  - Faster computation
  - You do not need sparse-matrices to do spatial statistics with many places,
    - banded matrices solve storage problem
    - Computation with banded matrices is faster than with sparse matrices

Dense and banded matrices

Dense matrix

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

Banded matrix

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

- Upper bandwidth of banded matrix is 1, lower bandwidth is 2

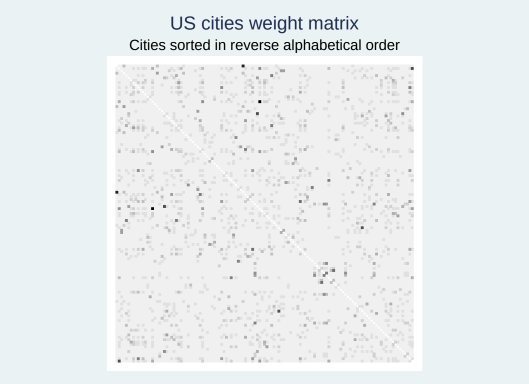
Example with US cities data

- We have data on the distance between 125 US cities
- This data is distributed with the cities in reverse alphabetical order
- Making an inverse-distance spatial-weighting matrix from the data in this order yields a matrix without any structure

```
. use us125
. spmat idistance C1 , id(pid) coordinates(x y) miles
. spmat graph C1, name(C1) title(US cities weight matrix)        ///
>         subtitle(Cities sorted in reverse alphabetical order)
```

Plot from default sort

- There is no structure in this spatial-weighting matrix
- The dark points near the north-east and south-west corners indicate that the minimum bandwidth is about the same as the matrix dimension



Value truncation

- With large spatial-weighting matrices, we sometimes impose the condition that distant places have zero effect on each other
- This restriction changes the spatial-weighting matrices and the model parameters
  - For example, we can impose the condition that US cities which are more than 500 miles apart have zero effect on each other (instead of .002 or smaller)

Bandwidths from default sort

```
. spmat summarize C1, vtruncate(.002)
Summary of spatial-weighting object C1
```

	Current matrix	Truncated matrix
Dimensions	125 x 125	125 x 125
# of zeros	125	12451
Minimum	0	0
Maximum	.0482526	.0482526
Mean	.0016527	.0008975
Median	.001044	0
Symmetric	yes	yes
Banded	no	no

Truncation scenario summary

	Lower band	Upper band
Best	123	123
75%	87	79
Mean	56.584	55.264
Median	56	55
Tukey value	178.5	160
> Tukey value	0	0

The Worcester sort

```
. gen double dww = sqrt( (x-x[5])^2 + (y-y[5])^2 )
. sort dww
. spmat idistance C2 , id(pid) coordinates(x y) miles
. spmat graph C2, name(C2) title(US cities weight matrix)          ///
>         subtitle(Cities sorted by distance from Worcester, MA )
```

Plot from Worcester sort

- The banded structure is clearly evident
- We could save this spatial-weighting matrix as a banded matrix, use less memory and perform the computations faster

Bandwidths from Worcester sort

```
. spmat summarize C2, vtruncate(.002)
Summary of spatial-weighting object C2
```

	Current matrix	Truncated matrix
Dimensions	125 x 125	125 x 125
# of zeros	125	12451
Minimum	0	0
Maximum	.0482526	.0482526
Mean	.0016527	.0008975
Median	.001044	0
Symmetric	yes	yes
Banded	no	no

```
Truncation scenario summary
```

	Lower band	Upper band
Best	33	33
75%	28	28
Mean	19.208	19.816
Median	20	21
Tukey value	53.5	50.5
> Tukey value	0	0

US county data (unsorted)

```
. use county2, clear
. spmat contiguity C1 using countyxy, id(id) replace
. spmat summarize C1, vtruncate(.5)
Summary of spatial-weighting object C1
```

	Current matrix	Truncated matrix
Dimensions	3109 x 3109	3109 x 3109
# of zeros	9648149	9648149
Minimum	0	0
Maximum	1	1
Mean	.0018345	.0018345
Median	0	0
Symmetric	yes	yes
Banded	no	no

Truncation scenario summary

	Lower band	Upper band
Best	3082	3082
75%	1774	1843
Mean	1041.577	1048.394
Median	929	959
Tukey value	4222	4484.5
> Tukey value	0	0

```
.                                  // observation 1425 is San Juan County, WA
. generate d0 = sqrt( (x- x[1425])^2 + (y - y[1425])^2 )
. sort d0                          // d0 is distance from San Juan County, WA
. spmat contiguity C2 using countyxy, id(id) replace
. spmat summarize C2, vtruncate(.5)
Summary of spatial-weighting object C2
```

	Current matrix	Truncated matrix
Dimensions	3109 x 3109	3109 x 3109
# of zeros	9648149	9648149
Minimum	0	0
Maximum	1	1
Mean	.0018345	.0018345
Median	0	0
Symmetric	yes	yes
Banded	no	no

Truncation scenario summary

	Lower band	Upper band
Best	356	356
75%	91	90
Mean	74.73046	74.93052
Median	65	65
Tukey value	160	156
> Tukey value	207	204

Some underlying statistical theory

- Recall the model

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{u}$$
$$\mathbf{u} = \rho \mathbf{M}\mathbf{u} + \epsilon$$

- The model specifies that set of N simultaneous equations for $\mathbf{y}$ and for $\mathbf{u}$
- The identification assumptions ensure that we can solve for $\mathbf{u}$ and $\mathbf{y}$
- Solving for $\mathbf{u}$ yields

$$\mathbf{u} = (\mathbf{I} - \rho \mathbf{M})^{-1}\epsilon$$

- If ϵ is IID with finite variance σ^2, the spatial correlation among the errors is given by

$$\mathbf{\Omega}_u = E[\mathbf{u}\mathbf{u}'] = \sigma^2(I - \rho\mathbf{M})^{-1}(I - \rho\mathbf{M}')^{-1}$$

Some underlying statistical theory II

- Solving for $\mathbf{y}$ yields

$$\mathbf{y} = (I - \lambda \mathbf{W})^{-1} \mathbf{X}\beta + (I - \lambda \mathbf{W})^{-1}(I - \rho \mathbf{M})^{-1} \boldsymbol{\epsilon}$$

- **Wy** is not an exogenous variable
  Using the above solution for $\mathbf{y}$ we can see that

$$E[(\mathbf{W}\mathbf{y})\mathbf{u}'] = \mathbf{W}(I - \lambda \mathbf{W})^{-1} \boldsymbol{\Omega}_u \neq 0$$

Maximum likelihood estimator

- The above solution for **y** permits the derivation of the log-likelihood function
- In practice, we use the concentrated log-likelihood function

$$\ln L_2^*(\lambda,\rho) = -\frac{n}{2}\left(\ln(2\pi) + 1 + \ln \widehat{\sigma}^2(\lambda,\rho)\right) + \ln ||\mathbf{I} - \lambda \mathbf{W}|| + \ln ||\mathbf{I} - \rho \mathbf{M}||$$

where

$$\begin{aligned}
\widehat{\sigma}^2(\lambda,\rho) &= \frac{1}{n}\mathbf{y}_*^*(\lambda,\rho)' \left[\mathbf{I} - \mathbf{X}_*(\rho)\left[\mathbf{X}_*(\rho)'\mathbf{X}_*(\rho)\right]^{-1}\mathbf{X}_*(\rho)'\right]\mathbf{y}_*^*(\lambda,\rho) \\
\mathbf{y}^*(\lambda) &= (\mathbf{I} - \lambda\mathbf{W})\mathbf{y}, \\
\mathbf{y}_*^*(\lambda,\rho) &= (\mathbf{I} - \rho\mathbf{M})\mathbf{y}^*(\lambda) = (\mathbf{I} - \rho\mathbf{M})(\mathbf{I} - \lambda\mathbf{W})\mathbf{y}, \\
\mathbf{X}_*(\rho) &= (\mathbf{I} - \rho\mathbf{M})\mathbf{X},
\end{aligned}$$

Pluggin the values $\hat{\lambda}$ and $\hat{\rho}$ that maximize the above concentrated log-likelihood function into equation $\widehat{\sigma}^2(\lambda,\rho)$ produces the ML estimate of σ^2.

Maximum likelihood estimator II

- Substituting the values $\hat{\lambda}$ and $\hat{\rho}$ that maximize the above concentrated log-likelihood function into

$$\widehat{\boldsymbol{\beta}}(\lambda, \rho) = \left[\mathbf{X}_*(\rho)'\mathbf{X}_*(\rho)\right]^{-1}\mathbf{X}_*(\rho)'\mathbf{y}_*^*(\lambda, \rho)$$

produces the ML estimate of $\boldsymbol{\beta}$.

Maximum likelihood estimator III

- Three types problems remain
  - Numerical
  - Lack of general statistical theory
  - Quasi-maximum likelihood theory does not apply

Numerical problems with ML estimator

- The ML estimator requires computing the determinants $|\mathbf{I} - \lambda \mathbf{W}|$ and $|\mathbf{I} - \rho \mathbf{M}|$ for each iteration
- Ord (1975) showed $|I - \rho \mathbf{W}| = \prod_{i=1}^{n}(1 - \rho v_i)$ where $(v_1, v_2, ..., v_n)$ are the eigenvalues of $\mathbf{W}$
  - This reduces, but does not remove, the problem
  - For instance, with zip-code-level data, this would require obtaining the eigenvalues of a 32,000 by 32,000 square matrix

Lack of general statistical theory

- There is still no large-sample theory for the distribution of the ML for the Cliff-Ord model
- Special cases covered by Lee (2004)
  - Allows for spatially correlated errors, but no spatially lagged dependent variable
- This estimator is frequently used, even though there is no large-sample theory for the distribution of the estimator

Quasi-maximum likelihood theory does not apply

- Simple deviations from Normal IID can cause the ML estimator to produce inconsistent estimates
    - Arraiz, Drukker, Kelejian, and Prucha (2009) provide simulation evidence that the ML estimator produces inconsistent estimates when the errors are heteroskedastic

spreg ml command

- Forthcoming user-written Stata command spreg ml estimates the parameters of Cliff-Ord models by ML

```
. spreg ml y lwage police , elmat(chess) dlmat(chess) pid(pid)
Iteration 0:   log likelihood = -4120.2131
  (output omitted)
Spatial autoregressive model                    Number of obs   =        625
(Maximum likelihood estimates)                  Wald chi2(2)    =    1224.99
                                                Prob > chi2     =     0.0000
```

y	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
y						
lwage	.9566215	.0314487	30.42	0.000	.8949833	1.01826
police	1.153248	.0709958	16.24	0.000	1.014099	1.292397
_cons	.6261475	.0734968	8.52	0.000	.4820965	.7701985
lambda						
_cons	.7340299	.0036412	201.59	0.000	.7268932	.7411666
rho						
_cons	.7685249	.0030415	252.68	0.000	.7625637	.7744861
sigma						
_cons	1.956339	.0562829	34.76	0.000	1.846026	2.066651

Generalized spatial Two-stage least squares (GS2SLS)

- Kelejian and Prucha (1999, 1998, 2004, 2009) along with coauthors Arraiz, Drukker, Kelejian, and Prucha (2009) derived an estimator that uses instrumental variables and the generalized-method-of-moments (GMM) to estimate the parameters of cross-sectional Cliff-Ord models
- Arraiz, Drukker, Kelejian, and Prucha (2009) show that the estimator produces consistent estimates when the disturbances are heteroskedastic and give simulation evidence that the ML estimator produces inconsistent estimates in the case

GS2SLS II

- The estimator is produced in four steps
  1. Consistent estimates of $\boldsymbol{\beta}$ and λ are obtained by instrumental variables
     - Following Kelejian and Prucha (1998) $\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}, \ldots \mathbf{MX}, \mathbf{MWX}, \mathbf{MW}^2\mathbf{X}, \ldots$ are valid instruments,
     - By default, we use $\mathbf{H} = \mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X})$
  2. Estimate ρ and σ by GMM using sample constructed from functions of the residuals
     - The moment conditions explicitly allow for heteroskedastic innovations
     - Drukker, Egger, and Prucha (2009a) work out the details for homoskedastic case
  3. Use the estimates of ρ and σ to perform a spatial Cochrane-Orcut transformation of the data and obtain more efficient estimates of $\boldsymbol{\beta}$ and λ
  4. Use the efficient estimates of $\boldsymbol{\beta}$ and λ to obtain an efficient GMM estimator of ρ
- The authors derive the joint large-sample distribution of the estimators

spreg g2sls command

- Forthcoming user-written command `spreg g2sls` implements the Arraiz et al. (2009) and the Drukker, Egger, and Prucha (2009a) estimators

```
. spreg gs2sls y lwage police , dlmat(chess) elmat(chess) pid(id)
Estimating rho by GMM
Iteration 1:   SSR = 14819.059
  (output omitted)
GS2SLS regression                                    Number of obs =      625
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
y						
lwage	1.164297	.113891	10.22	0.000	.9410746	1.387519
police	1.400984	.1775355	7.89	0.000	1.053021	1.748947
_cons	1.510546	.4212849	3.59	0.000	.6848427	2.336249
lambda						
_cons	.9836465	.1101239	8.93	0.000	.7678076	1.199485
rho						
_cons	.7212283	.0188375	38.29	0.000	.6843075	.7581491

g2sls command II

```
. estimates table ml gs2sls
```

Variable	ml	gs2sls
y		
lwage	.95662152	1.164297
police	1.1532476	1.4009842
_cons	.62614744	1.5105459
lambda		
_cons	.73402989	.98364646
rho		
_cons	.76852488	.72122828
sigma		
_cons	1.9563386	

Allowing for endogenous covariates

- Kelejian and Prucha (2004); Drukker, Egger, and Prucha (2009a) extend the estimation technique to allow for endogenous covariates
- The model is now

$$\mathbf{y} = \lambda \mathbf{W}\mathbf{y} + \mathbf{X}_x \beta + \mathbf{X}_n \gamma + \mathbf{u}$$
$$\mathbf{u} = \rho \mathbf{M}\mathbf{u} + \epsilon$$

  where $\mathbf{X}_x$ contains exogenous covariates and $\mathbf{X}_n$ contains endogenous covariates
- We assume that we have additional instruments $\mathbf{Z}$
- The only important change in the estimation technique is to use instruments
  $\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}, \ldots \mathbf{MX}, \mathbf{MWX}, \mathbf{MW}^2\mathbf{X}, \ldots$
  where $\mathbf{X} = [\mathbf{X}_x, \mathbf{Z}]$

spivreg

- The forthcoming user-written command spivreg implements this estimator

```
. spivreg y lwage (police = convict arrest) , dlmat(chess) elmat(chess) pid(id)
Estimating rho using 2SLS residuals
Iteration 0:   GMM criterion =  145822.98
```

(*output omitted*)

```
Spatial regression with endogenous variables      Number of obs    =        625
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
y						
police	.9782551	.0984056	9.94	0.000	.7853838	1.171127
lwage	.9375394	.0533136	17.59	0.000	.8330467	1.042032
_cons	.639005	.1981524	3.22	0.001	.2506335	1.027376
lambda						
_cons	.7078255	.052859	13.39	0.000	.6042238	.8114272
rho						
_cons	.7951642	.0676218	11.76	0.000	.662628	.9277004

Summary and further research

- An increasing number of datasets contain spatial information
- Modeling the spatial processes in a dataset can improve efficiency, or be essential for consistency
- The Cliff-Ord type models provide a useful parametric approach to spatial data
- There is reasonably general statistical theory for the GS2SLS estimator for the parameters of cross-sectional Cliff-Ord type models
- We are now working on extending the GS2SLS to panel-data Cliff-Ord type models with large N and fixed T

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*, Boston: Kluwer Academic Publishers.

Arraiz, Irani, David M. Drukker, Harry H. Kelejian, and Ingmar R. Prucha. 2009. "A Spatial Cliff-Ord-type Model with Heteroskedastic Innovations: Small and Large Sample Results," *Journal of Regional Science*, forthcoming.

Cliff, A. D. and J. K. Ord. 1973. *Spatial Autocorrelation*, London: Pion.

———. 1981. *Spatial Processes, Models and Applications*, London: Pion.

Drukker, David M., Peter Egger, and Ingmar R. Prucha. 2009a. "On Single Equation GMM estimation of a Spatial Autoregressive Model with Spatially Autoregressive Disturbance," Tech. rep., Department of Economics
University of Maryland.

Drukker, David M., Hua Peng, Ingmar R. Prucha, and Rafal Raciborski. 2009b. "Sorting induces a banded stucture in spatial-weighting matrices," Working paper, Department of Economics, University of Maryland.

Kelejian, Harry H. and Ingmar R. Prucha. 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17(1), 99–121.

———. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40(2), 509–533.

———. 2002. "2SLS and OLS in a spatial autoregressive model with equal spatial weights," *Regional Science and Urban Economics*, (32), 691–707.

———. 2004. "Estimation of simultaneous systems of spatially interrelated cross sectional equations," *Journal of Econometrics*, 118, 27–50.

———. 2009. "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," *Journal of Econometrics*, forthcoming.

Lee, L. F. 2004. "Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models," *Econometrica*, (72), 1899–1925.

Ord, J. K. 1975. "Estimation Methods for Spatial Interaction," *Journal of the American Statistical Association*, 70, 120–126.

Tobler, W. R. 1970. "A computer movie simulating urban growth in the Detroit region," *Economic Geography*, 46, 234–40.