Analyzing spatial autoregressive models using Stata

David M. Drukker

StataCorp

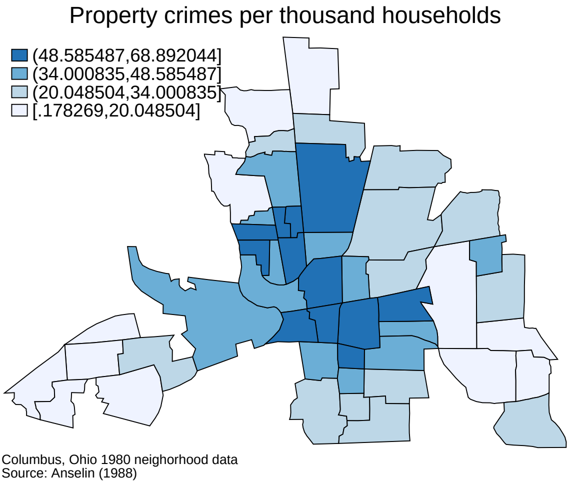
Summer North American Stata Users Group meeting
July 24-25, 2008

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

Outline

What is spatial data?

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



Correcting for Spatial Correlation

- Correct for correlation in unobservable errors
  - Efficiency and consistent standard errors
- Correct for outcome in place *i* depending 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

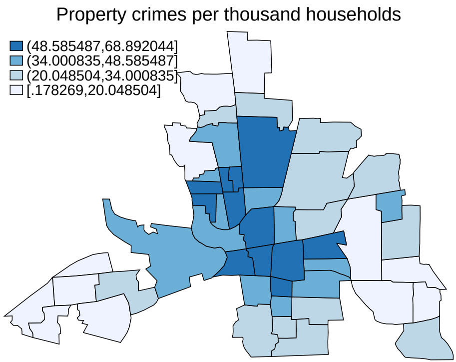
```
. use columbusdb, clear
. describe id crime hoval inc
                     storage  display     value
variable name        type     format      label      variable label
--------------------------------------------------------------------------------
id                   byte     %12.0g                 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)")
```



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

Modeling spatial data

- Cliff-Ord type models used in many social-sciences
  - So named for [Cliff and Ord(1973), Cliff and Ord(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

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 $\mathbf{M}$
    - The row and column sums must be finite, as the number of places grows to infinity
- Restricting the number of neighbors that affect any given place reduces dependence
- Restricting the extent to which neighbors affect any given place reduces dependence

Spatial-weighting matrices parameterize dependence II

- 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

Normalizing spatial-weighting matrices

- In practice, inverse-distance spatial-weighting matrices are usually normalized
  - Row normalized, $\widetilde{\mathbf{W}}$ has element $\widetilde{w}_{i,j} = (1/\sum_{j=1}^{N}|w_{i,j}|)w_{i,j}$
  - Minmax normalized, $\widetilde{\mathbf{W}}$ has element $\widetilde{w}_{i,j} = (1/f)w_{i,j}$ where f is $\min(s_r,\, s_c)$ and s_r is the largest row sum and s_c is the largest column sum of $\mathbf{W}$
  - Spectral normalized, $\widetilde{\mathbf{W}}$ has element $\widetilde{w}_{i,j} = (1/|v_|)w_{i,j}$ where $|v|$ is the modulus of the largest eigenvalue of $\mathbf{W}$

Creating and Managing spatial weighting matrices in Stata

- There is a forthcoming user-written command by David Drukker 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
- 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, p(id)
. spmat idistance  idmat_row, p(id) pinformation(x y) normalize(row)
. spmat idistance  idmat_mmax, p(id) pinformation(x y) normalize(spectral)
```

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 a set of N simultaneous equations for $\mathbf{y}$ and for $\mathbf{u}$
- Two identification assumptions require 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

- Pluggin 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 (2008) provide simulation evidence showing that the ML estimator produces inconsistent estimates when the errors are heteroskedastic

sarml command

- Forthcoming user-written Stata command `sarml` estimates the parameters of Cliff-Ord models by ML

```
. sarml crime hoval inc, armat(idmat_c) ecmat(idmat_c) nolog
Spatial autoregressive model                       Number of obs   =        49
(Maximum likelihood estimates)                     Wald chi2(2)    =   43.6207
                                                   Prob > chi2     =    0.0000
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
crime						
hoval	-.2806984	.0972479	-2.89	0.004	-.4713008	-.090096
inc	-1.201205	.3363972	-3.57	0.000	-1.860532	-.5418786
_cons	56.79677	6.19428	9.17	0.000	44.6562	68.93733
lambda						
_cons	.042729	.0286644	1.49	0.136	-.0134522	.0989103
rho						
_cons	.0639603	.0768638	0.83	0.405	-.0866899	.2146105
sigma						
_cons	9.78293	1.005501	9.73	0.000	7.812185	11.75367

```
. estimates store sarml
```

Generalized spatial Two-stage least squares (GS2SLS)

- Kelejian and Prucha [Kelejian and Prucha(1999), Kelejian and Prucha(1998), Kelejian and Prucha(2004)] along with coauthors [Arraiz et al.(2008)Arraiz, Drukker, Kelejian, and Prucha] 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 et al.(2008)Arraiz, Drukker, Kelejian, and Prucha] 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 three steps
  1. Consistent estimates of β 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,
  2. Estimate ρ and σ by GMM using sample constructed from functions of the residuals
     - The moment conditions explicitly allow for heteroskedastic innovations.
  3. Use the estimates of ρ and σ to perform a spatial Cochrane-Orcut transformation of the data and obtain more efficient estimates of β and λ
- The authors derive the joint large-sample distribution of the estimators

g2sls command

- Forthcoming user-written command g2sls implements the [Arraiz et al.(2008)Arraiz, Drukker, Kelejian, and Prucha] estimator

```
. gs2sls crime hoval inc, armat(idmat_c) ecmat(idmat_c) nolog
GS2SL regression                                    Number of obs =      49

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
crime        |
       hoval |  -.2561395   .1621328   -1.58   0.114     -.573914     .061635
         inc |  -1.186221   .5223525   -2.27   0.023    -2.210013   -.1624286
       _cons |   52.28828   9.441074    5.54   0.000     33.78411    70.79244
-------------+----------------------------------------------------------------
lambda       |
       _cons |   .0651557   .0273606    2.38   0.017     .0115299    .1187816
-------------+----------------------------------------------------------------
rho          |
       _cons |   .0224828   .0801967    0.28   0.779    -.1346998    .1796654
------------------------------------------------------------------------------

. estimates store gs2sls
```

g2sls command II

```
. estimates table gs2sls sarml, b se
```

Variable	gs2sls	sarml
crime		
hoval	-.2561395	-.28069841
	.16213284	.09724792
inc	-1.1862206	-1.2012051
	.52235246	.33639724
_cons	52.28828	56.796767
	9.4410739	6.1942805
lambda		
_cons	.06515574	.04272904
	.02736064	.02866443
rho		
_cons	.0224828	.0639603
	.08019667	.07686377
sigma		
_cons		9.7829296
		1.0055006

legend: b/se

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, I., D. M. Drukker, H. H. Kelejian, and I. R. Prucha. 2008.
A Spatial Cliff-Ord-type Model with Heteroskedastic Innovations: Small and Large Sample Results.
Technical report, Department of Economics, University of Maryland.
http://www.econ.umd.edu/ prucha/Papers/WP_Hetero_GM_IV_MC.pdf
.

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

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

Kelejian, H. H., and I. 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.

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

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.