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.
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)
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`
    - Gives a great example of how to translate shapefiles and map data

- Need to create spatial-weighting matrices that parameterize distance
Managing spatial data

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
[Anselin(1988)] used a dataset containing information on property crimes in 49 neighborhoods in Columbus, 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 by the user-written command `spmap` by Maurizio Pisati.

See also

http://econpapers.repec.org/software/bocbocode/s456812.htm
Managing spatial data

Columbus data part II

. use columbusdb, clear
. describe id crime hoval inc

<table>
<thead>
<tr>
<th>variable</th>
<th>type</th>
<th>format</th>
<th>label</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>byte</td>
<td>%12.0g</td>
<td>neighborhood id</td>
</tr>
<tr>
<td>crime</td>
<td>double</td>
<td>%10.0g</td>
<td>residential burglaries and vehicle thefts per 1000 households</td>
</tr>
<tr>
<td>hoval</td>
<td>double</td>
<td>%10.0g</td>
<td>housing value (in $1,000)</td>
</tr>
<tr>
<td>inc</td>
<td>double</td>
<td>%10.0g</td>
<td>household income (in $1,000)</td>
</tr>
</tbody>
</table>

. list id crime hoval inc in 1/5

<table>
<thead>
<tr>
<th>id</th>
<th>crime</th>
<th>hoval</th>
<th>inc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.72598</td>
<td>80.467003</td>
<td>19.531</td>
</tr>
<tr>
<td>2</td>
<td>18.801754</td>
<td>44.567001</td>
<td>21.232</td>
</tr>
<tr>
<td>3</td>
<td>30.626781</td>
<td>26.35</td>
<td>15.956</td>
</tr>
<tr>
<td>4</td>
<td>32.38776</td>
<td>33.200001</td>
<td>4.477</td>
</tr>
<tr>
<td>5</td>
<td>50.73151</td>
<td>23.225</td>
<td>11.252</td>
</tr>
</tbody>
</table>
spmap is an outstanding user-written command for exploring spatial data

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

Property crimes per thousand households
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

\[ y = \lambda W y + X \beta + u \]
\[ u = \rho M u + \epsilon \]

where

- \( y \) is the \( N \times 1 \) vector of observations on the dependent variable
- \( X \) is the \( N \times k \) matrix of observations on the independent variables
- \( W \) and \( M \) are \( N \times N \) spatial-weighting matrices that parameterize the distance between neighborhoods
- \( u \) are spatially correlated residuals and \( \epsilon \) are independent and identically distributed disturbances
- \( \lambda \) and \( \rho \) are scalars that measure, respectively, the dependence of \( y_i \) on nearby \( y \) and the spatial correlation in the errors
Spatial autoregressive models

Cliff-Ord models II

\[ y = \lambda W y + X \beta + u \]
\[ u = \rho M u + \epsilon \]

- Relatively simple, tractable model
- Allows for correlation among unobservables
  - Each \( u_i \) depends on a weighted average of other observations in \( u \)
  - \( M u \) is known as a spatial lag of \( u \)
- Allows for \( y_i \) to depend on nearby \( y \)
  - Each \( y_i \) depends on a weighted average of other observations in \( y \)
  - \( W y \) is known as a spatial lag of \( y \)
- 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.
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 $W$ and $M$.

- 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.
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
In practice, inverse-distance spatial-weighting matrices are usually normalized:

- **Row normalized**, $\tilde{W}$ has element $\tilde{w}_{i,j} = \left(1/\sum_{j=1}^{N} |w_{i,j}|\right) w_{i,j}$
- **Minmax normalized**, $\tilde{W}$ has element $\tilde{w}_{i,j} = (1/f) w_{i,j}$ 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 $W$
- **Spectral normalized**, $\tilde{W}$ has element $\tilde{w}_{i,j} = (1/|\lambda|) w_{i,j}$ where $|\lambda|$ is the modulus of the largest eigenvalue of $W$
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

\[ y = \lambda Wy + X\beta + u \]
\[ u = \rho Mu + \epsilon \]

- The model specifies that a set of \( N \) simultaneous equations for \( y \) and for \( u \)

- Two identification assumptions require that we can solve for \( u \) and \( y \)

- Solving for \( u \) yields

\[ u = (I - \rho M)^{-1}\epsilon \]

- If \( \epsilon \) is IID with finite variance \( \sigma^2 \), the spatial correlation among the errors is given by

\[ \Omega_u = E[uu'] = \sigma^2(I - \rho M^{-1})(I - \rho M')^{-1} \]
Solving for $\mathbf{y}$ yields

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

$\mathbf{W} \mathbf{y}$ is not an exogenous variable

Using the above solution for $\mathbf{y}$ we can see that

$$E[(\mathbf{W} \mathbf{y})' \mathbf{u}] = \mathbf{W} (I - \lambda \mathbf{W})^{-1} \Omega_{\mathbf{u}} \neq 0$$
The above solution for $y$ permits the derivation of the log-likelihood function.
In practice, we use the concentrated log-likelihood function

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

where

$$\hat{\sigma}^2(\lambda, \rho) = \frac{1}{n} y_*^*(\lambda, \rho)' \left[ I - X_*(\rho) [X_*(\rho)'X_*(\rho)]^{-1} X_*(\rho)' \right] y_*^*(\lambda, \rho)$$

$$y^*(\lambda) = (I - \lambda W)y,$$

$$y_*^*(\lambda, \rho) = (I - \rho M)y^*(\lambda) = (I - \rho M)(I - \lambda W)y,$$

$$X_*(\rho) = (I - \rho M)X.$$
Plugging the values $\hat{\lambda}$ and $\hat{\rho}$ that maximize the above concentrated log-likelihood function into

$$
\hat{\beta}(\lambda, \rho) = \left[ X_*(\rho)'X_*(\rho) \right]^{-1} X_*(\rho)'y_*(\lambda, \rho)
$$

produces the ML estimate of $\beta$. 
Three types problems remain

- Numerical
- Lack of general statistical theory
- Quasi-maximum likelihood theory does not apply
The ML estimator requires computing the determinants $|I - \lambda W|$ and $|I - \rho M|$ for each iteration.

[Ord(1975)] showed $|I - \rho W| = \prod_{i=1}^{n} (1 - \rho v_i)$ where $(v_1, v_2, ..., v_n)$ are the eigenvalues of $W$.

- 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.
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.
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
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

| Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|-----------|-------|------|----------------------|
| crime  |           |       |      |                      |
| hoval  | -.2806984 | .0972479 | -2.89 | 0.004 | -.4713008 to -.090096 |
| inc    | -1.201205 | .3363972 | -3.57 | 0.000 | -1.860532 to -.5418786 |
| _cons  | 56.79677  | 6.19428  | 9.17  | 0.000 | 44.6562 to 68.93733   |
| lambda |           |       |      |                      |
| _cons  | .042729   | .0286644 | 1.49  | 0.136 | -.0134522 to .0989103 |
| rho    |           |       |      |                      |
| _cons  | .0639603  | .0768638 | .83   | 0.405 | -.0866899 to .2146105 |
| sigma  |           |       |      |                      |
| _cons  | 9.78293   | 1.005501 | 9.73  | 0.000 | 7.812185 to 11.75367  |

```
. estimates store sarml
```

[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
The estimator is produced in three steps

1. Consistent estimates of $\beta$ and $\lambda$ are obtained by instrumental variables
   - Following [Kelejian and Prucha(1998)]
     \[ X, WX, W^2X, \ldots MX, MWX, MW^2X, \ldots \] are valid instruments,

2. Estimate $\rho$ and $\sigma$ by GMM using sample constructed from functions of the residuals
   - The moment conditions explicitly allow for heteroskedastic innovations.

3. Use the estimates of $\rho$ and $\sigma$ to perform a spatial Cochrane-Orcut transformation of the data and obtain more efficient estimates of $\beta$ and $\lambda$

The authors derive the joint large-sample distribution of the estimators
Forthcoming user-written command g2s1s implements the [Arraiz et al.(2008)Arraiz, Drukker, Kelejian, and Prucha] estimator

```
    . g2s1s crime hoval inc, armat(idmat_c) ecmat(idmat_c) nolog
    GS2SL regression

                  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 g2s1s
### g2s1s command II

```
. estimates table g2s1s sarml, b se
```

<table>
<thead>
<tr>
<th>Variable</th>
<th>g2s1s</th>
<th>sarml</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>crime</td>
<td></td>
</tr>
<tr>
<td></td>
<td>hoval</td>
<td>-.2561395</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.16213284</td>
</tr>
<tr>
<td></td>
<td>inc</td>
<td>-1.1862206</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.52235246</td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>52.28828</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.4410739</td>
</tr>
<tr>
<td></td>
<td>lambda</td>
<td></td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>.06515574</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.02736064</td>
</tr>
<tr>
<td></td>
<td>rho</td>
<td></td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>.0224828</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.08019667</td>
</tr>
<tr>
<td></td>
<td>sigma</td>
<td></td>
</tr>
<tr>
<td></td>
<td>_cons</td>
<td>9.7829296</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.0055006</td>
</tr>
</tbody>
</table>

Legend: b/se
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


A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model.  
Estimation of simultaneous systems of spatially interrelated cross sectional equations.  
Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models.  
Estimation Methods for Spatial Interaction. 

A computer movie simulating urban growth in the Detroit region. 