estat moran — Moran’s test of residual correlation with nearby residuals

Description

estat moran is a postestimation test that can be run after fitting a model using regress with spatial data. It performs the Moran test for spatial correlation among the residuals.

Quick start

Linear regression of y on x1 and x2, then testing for spatial correlation among the residuals using the spatial weighting matrix W

regress y x1 x2
estat moran, errorlag(W)

After the same regress command, add another spatial weighting matrix

estat moran, errorlag(W) errorlag(M)

After regress with no independent variables

regress y
estat moran, errorlag(W)

Menu for estat

Statistics > Postestimation

Syntax

estat moran, errorlag(spmatname) [errorlag(spmatname) ...]

Option

errorlag(spmatname) specifies a spatial weighting matrix that defines the error spatial lag that will be tested. errorlag() is required. This option is repeatable to allow testing of higher-order error lags.

Remarks and examples

If you have not read [SP] Intro 1–[SP] Intro 8, you should do so before using estat moran.

To use estat moran, your data must be cross-sectional Sp data. See [SP] Intro 3 for instructions on how to prepare your data.
To specify the form of the spatial correlation to be tested, you will need to have one or more spatial weighting matrices. See [SP] Intro 2 and [SP] spmatrix for an explanation of the types of weighting matrices and how to create them.

Before fitting a spatial autoregressive (SAR) model with spregress, you may want to fit the model with regress and then run estat moran. If the Moran test is significant, you will likely want to fit the model with spregress. If the test is not significant, you may question the need to fit a SAR model.

regress can be used with a single variable before running estat moran. This is a test of the spatial correlation of the variable.

Example 1: A test for spatial correlation

We have data on the homicide rate in counties in southern states of the U.S. homicide1990.dta contains hrate, the county-level homicide rate per year per 100,000 persons; ln_population, the logarithm of the county population; ln_pdensity, the logarithm of the population density; and gini, the Gini coefficient for the county, a measure of income inequality where larger values represent more inequality (Gini 1909). The data are an extract of the data originally used by Messner et al. (2000); see Britt (1994) for a literature review of the topic. This dataset is also used for the examples in [SP] spregress.

We used spshape2dta in the usual way to create the datasets homicide1990.dta and homicide1990_shp.dta. The latter file contains the boundary coordinates for U.S. southern counties. See [SP] Intro 4, [SP] Intro 7, [SP] spshape2dta, and [SP] spset.

Because the analysis dataset and the Stata-formatted shapefile must be in our working directory to spset the data, we first save both homicide1990.dta and homicide1990_shp.dta to our working directory by using the copy command. We then load the data and type spset to display the Sp attributes of the data.

```
. use homicide1990
(S.Messner et al.(2000), U.S southern county homicide rates in 1990)
. spset
   Sp dataset homicide1990.dta
      data: cross sectional
      spatial-unit id: _ID
      coordinates: _CX, _CY (planar)
      linked shapefile: homicide1990_shp.dta
```
We plot the homicide rate on a map of the counties by using the `grmap` command; see [SP] `grmap`. Figure 1 is the result.

```
. grmap hrate
```

Figure 1: Homicide rate in 1990 for southern U.S. counties

The homicide rate appears to be spatially dependent because the high homicide-rate counties appear to be clustered together.

To conduct the Moran test, we need a spatial weighting matrix. We will create a contiguity matrix and use the default spectral normalization for this matrix. See [SP] `Intro 2` and [SP] `spmatrix create` for details. We type

```
. spmatrix create contiguity W
```

Now, we run `regress` and then `estat moran`:

```
. regress hrate

Source | SS    | df  | MS
------ |------ |---- |---
Model  | 0     | 0   | .
Residual | 69908.59 | 1,411 | 49.5454217
Total  | 69908.59 | 1,411 | 49.5454217

hrate

| Coef.  | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|--------|-----------|------|-----|----------------------|
| _cons  | 9.549293  | .1873201 | 50.98 | 0.000 | 9.181837 | 9.916749 |

. estat moran, errorlag(W)

Moran test for spatial dependence
Ho: error is i.i.d.
Errorlags: W
chi2(1) = 265.84
Prob > chi2 = 0.0000
```

The test reports that we can reject that the errors are i.i.d. This is not surprising based on our visual appraisal of the data.

`estat moran` can be used with more than one weighting matrix. In this case, it produces a joint test of whether any of the weighting matrices specify a spatial dependence.
. spmatrix create idistance M
. estat moran, errorlag(W) errorlag(M)

Moran test for spatial dependence
Ho: error is i.i.d.
Errorlags: W M
   chi2(2) = 898.62
   Prob > chi2 = 0.0000

We can also use estat moran after a linear regression with independent variables:

. regress hrate ln_population ln_pdensity gini

```
               Source |       SS      df     MS
----------------+------------------
        Model   |   11950.8309     3   3983.61032
        Residual |   57957.7591    1,408   41.1631812
       Total     |   69908.59    1,411  49.5454217

                   F(3, 1408) = 96.78  Prob > F = 0.0000
                   Adj R-squared = 0.1709
                   R-squared = 0.1709
                   Root MSE = 6.4159

                      Coef.    Std. Err.    t    P>|t|    [95% Conf. Interval]
----------------+------------------+---------+---------+------------------+
         hrate   |   0.5559273    0.2574637    2.16   0.031        .0508736    1.060981
        ln_population |   0.8231517    0.2304413    3.57   0.000        .3711065    1.275197
        ln_pdensity |   84.33136     5.169489    16.31   0.000       74.19063    94.47209
          gini    |   84.33136    5.169489    16.31   0.000       74.19063    94.47209
         _cons   |  -32.46353    2.891056   -11.23   0.000      -38.13477    -26.79229
----------------+------------------+---------+---------+------------------+
```

. estat moran, errorlag(W)

Moran test for spatial dependence
Ho: error is i.i.d.
Errorlags: W
   chi2(1) = 186.72
   Prob > chi2 = 0.0000
The Moran test is significant. We fit a SAR model using `spregress`, `gs2sls`:

```
. spregress hrate ln_population ln_pdensity gini, gs2sls errorlag(W)
(1412 observations)
(1412 observations (places) used)
(weighting matrix defines 1412 places)
(output omitted)
```

**Spatial autoregressive model**

| Coef. | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|-------|-----------|-----|------|----------------------|
| hrate | 0.3184462 | 0.2664379 | 1.20 | 0.232 | -0.2037625 to 0.8406549 |
| ln_population | 0.8156068 | 0.2469074 | 3.30 | 0.001 | 0.3316771 to 1.299537 |
| ln_pdensity | 88.44808 | 5.925536 | 14.93 | 0.000 | 100.0619 | 76.83425 |
| gini | -31.81189 | 3.115188 | -10.21 | 0.000 | -25.70624 to -25.70624 |

W
```
e.hrate 0.5250879 0.0326974 16.06 0.000 0.4610021 to 0.5891736
```

Wald test of spatial terms: chi2(1) = 257.89 Prob > chi2 = 0.0000

See `[SP] spregress`.

**Stored results**

`estat moran` stores the following in `r()`:

Scalars
- `r(chi2)` = \( \chi^2 \)  
- `r(df)` = degrees of freedom of \( \chi^2 \)  
- `r(p)` = \( p \)-value for model test

Macros
- `r(elmat)` = weighting matrices used to specify error lag

**Methods and formulas**

Consider the model

\[ y = X\beta + u \]

where \( y \) is the \( n \times 1 \) dependent-variable vector, \( X \) is the \( n \times K \) matrix of covariates, \( \beta \) is the \( K \times 1 \) vector of regression parameters, and \( u \) is the \( n \times 1 \) vector of disturbances. We assume that \( u_i \) are identically distributed with \( E(u_i) = 0 \) and \( E(u_i^2) = \sigma^2 \). We want to test the hypothesis that \( u_i \) are uncorrelated; that is, we want to test

\[ H_0 : E(uu') = \sigma^2 I \]

Consider the case where the researcher believes that the spatial weighting matrix \( W_1 \) gives a proper representation of spatial links for the disturbances \( u \). In this case, the researcher could test \( H_0 \) using the standard Moran \( I \) test statistic (Moran 1950),

\[ I = \frac{\hat{u}'W_1\hat{u}}{\hat{\sigma}^2 \Big[ \text{tr} \{(W_1' + W_1)W_1\} \Big]^{1/2}} \]
where \( \hat{u} = y - X\hat{\beta} \) are the estimated residuals and \( \hat{\sigma}^2 = \hat{u}'\hat{u}/n \) is the corresponding estimator for \( \sigma^2 \). Under appropriate assumptions, it follows from Kelejian and Prucha (2001) that \( I \sim N(0,1) \) and \( I^2 \sim \chi^2(1) \).

Next, consider the case where the researcher is not sure whether any of the weighting matrices \( W_1, W_2, \ldots, W_q \) properly model the spatial interdependence between \( u_i \). In this case, the researcher can test \( H_0 \) using the \( I(q)^2 \) test statistic:

\[
I(q)^2 = \left[ \begin{array}{c}
\hat{u}'W_1\hat{u}/\hat{\sigma}^2 \\
\vdots \\
\hat{u}'W_q\hat{u}/\hat{\sigma}^2 
\end{array} \right] \Phi^{-1} \left[ \begin{array}{c}
\hat{u}'W_1\hat{u}/\hat{\sigma}^2 \\
\vdots \\
\hat{u}'W_q\hat{u}/\hat{\sigma}^2 
\end{array} \right]'
\]

where \( \Phi = (\phi_{rs}) \) and \( r, s = 1, \ldots, q \):

\[
\phi_{rs} = \frac{1}{2} \text{tr} \{ (W_r + W_r') (W_s + W_s') \}
\]

It follows from Kelejian and Prucha (2001) and Drukker and Prucha (2013) that \( I(q)^2 \sim \chi^2(q) \) under \( H_0 \).

References


Also see

[SP] Intro — Introduction to spatial data and SAR models

[SP] spmatrix create — Create standard weighting matrices

[SP] spregress — Spatial autoregressive models

[R] regress — Linear regression