

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

Description	Quick start	Menu for estat	Syntax
Option	Remarks and examples	Stored results	Methods and formulas
References	Also see		

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 x_1 and x_2 , 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) ...]
```

`collect` is allowed; see [U] 11.1.10 [Prefix commands](#).

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.

```
. copy https://www.stata-press.com/data/r17/homicide1990.dta .
. copy https://www.stata-press.com/data/r17/homicide1990_shp.dta .
. use homicide1990
(S.Messner et al.(2000), U.S southern county homicide rates in 1990)
. spset
      Sp dataset: homicide1990.dta
Linked shapefile: homicide1990_shp.dta
      Data: Cross sectional
Spatial-unit ID:  _ID
Coordinates:  _CX, _CY (planar)
```

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
```

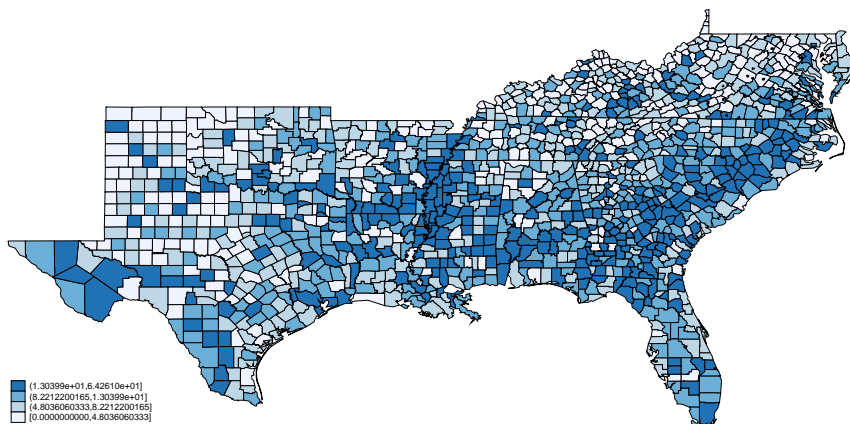


Figure 1: Homicide rate in 1990 for southern U.S. counties

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	Number of obs	=	1,412
Model	0	0	.	F(0, 1411)	=	0.00
Residual	69908.59	1,411	49.5454217	Prob > F	=	.
Total	69908.59	1,411	49.5454217	R-squared	=	0.0000
				Adj R-squared	=	0.0000
				Root MSE	=	7.0389

hrate	Coefficient	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
H0: Error terms are 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
H0: Error terms are 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	Number of obs	=	1,412
Model	11950.8309	3	3983.61032	F(3, 1408)	=	96.78
Residual	57957.7591	1,408	41.1631812	Prob > F	=	0.0000
				R-squared	=	0.1709
				Adj R-squared	=	0.1692
Total	69908.59	1,411	49.5454217	Root MSE	=	6.4159

hrate	Coefficient	Std. err.	t	P> t	[95% conf. interval]
ln_populat ⁿ	.5559273	.2574637	2.16	0.031	.0508736 1.060981
ln_pdensity	.8231517	.2304413	3.57	0.000	.3711065 1.275197
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
H0: Error terms are 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
GS2SLS estimates
Number of obs = 1,412
Wald chi2(3) = 243.84
Prob > chi2 = 0.0000
Pseudo R2 = 0.1686
```

hrate	Coefficient	Std. err.	z	P> z	[95% conf. interval]
hrate					
ln_populat ⁿ	.3184462	.2664379	1.20	0.232	-.2037625 .8406549
ln_pdensity	.8156068	.2469074	3.30	0.001	.3316771 1.299537
gini	88.44808	5.925536	14.93	0.000	76.83425 100.0619
_cons	-31.81189	3.115188	-10.21	0.000	-37.91755 -25.70624

W	e.hrate	Coefficient	Std. err.	z	P> z	[95% conf. interval]
		.5250879	.0326974	16.06	0.000	.4610021 .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

<code>r(chi2)</code>	χ^2
<code>r(df)</code>	degrees of freedom of χ^2
<code>r(p)</code>	p -value for model test

Macros

<code>r(emat)</code>	weighting matrices used to specify error lag
----------------------	--

Methods and formulas

Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

where \mathbf{y} is the $n \times 1$ dependent-variable vector, \mathbf{X} is the $n \times K$ matrix of covariates, $\boldsymbol{\beta}$ is the $K \times 1$ vector of regression parameters, and \mathbf{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

$$H_0 : E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}$$

Consider the case where the researcher believes that the spatial weighting matrix \mathbf{W}_1 gives a proper representation of spatial links for the disturbances \mathbf{u} . In this case, the researcher could test H_0 using the standard Moran I test statistic (Moran 1950),

$$I = \frac{\hat{\mathbf{u}}'\mathbf{W}_1\hat{\mathbf{u}}}{\hat{\sigma}^2 [\text{tr}\{(\mathbf{W}'_1 + \mathbf{W}_1)\mathbf{W}_1\}]^{1/2}}$$

where $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ are the estimated residuals and $\hat{\sigma}^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/n$ is the corresponding estimator for σ^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 $\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_q$ properly model the spatial interdependence between u_i . In this case, the researcher can test H_0 using the $I(q)^2$ test statistic:

$$I(q)^2 = \begin{bmatrix} \hat{\mathbf{u}}'\mathbf{W}_1\hat{\mathbf{u}}/\hat{\sigma}^2 \\ \vdots \\ \hat{\mathbf{u}}'\mathbf{W}_q\hat{\mathbf{u}}/\hat{\sigma}^2 \end{bmatrix}' \Phi^{-1} \begin{bmatrix} \hat{\mathbf{u}}'\mathbf{W}_1\hat{\mathbf{u}}/\hat{\sigma}^2 \\ \vdots \\ \hat{\mathbf{u}}'\mathbf{W}_q\hat{\mathbf{u}}/\hat{\sigma}^2 \end{bmatrix}$$

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

$$\phi_{rs} = \frac{1}{2} \text{tr}\{(\mathbf{W}_r + \mathbf{W}'_r)(\mathbf{W}_s + \mathbf{W}'_s)\}$$

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

References

- Britt, C. L. 1994. Crime and unemployment among youths in the United States, 1958–1990: A time series analysis. *American Journal of Economics and Sociology* 53: 99–109. <https://doi.org/10.1111/j.1536-7150.1994.tb02680.x>.
- Drukker, D. M., and I. R. Prucha. 2013. On the $I^2(q)$ test statistic for spatial dependence: Finite sample standardization and properties. *Spatial Economic Analysis* 8: 271–292. <https://doi.org/10.1080/17421772.2013.804630>.
- Gini, C. 1909. Concentration and dependency ratios (in Italian). English translation in *Rivista di Politica Economica* 1997 87: 769–789.
- Kelejian, H. H., and I. R. Prucha. 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104: 219–257. [https://doi.org/10.1016/S0304-4076\(01\)00064-1](https://doi.org/10.1016/S0304-4076(01)00064-1).
- Messner, S. F., L. Anselin, D. F. Hawkins, G. Deane, S. E. Tolnay, and R. D. Baller. 2000. An Atlas of the Spatial Patterning of County-Level Homicide, 1960–1990. Pittsburgh: National Consortium on Violence Research.
- Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37: 17–23. <https://doi.org/10.2307/2332142>.

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