spivregress — Spatial autoregressive models with endogenous covariates

Description

spivregress is the equivalent of ivregress for spatial data. spivregress fits spatial autoregressive (SAR) models, also known as simultaneous autoregressive models, where the models may contain additional endogenous variables as well as exogenous variables. These models can be used to account for possible dependence between the outcome variable and the unobserved errors.

For models without endogenous regressors, see [SP] spregress.

If you have not read [SP] Intro 1–[SP] Intro 8, you should do so before using spivregress. Your data must be Sp data to use spivregress. See [SP] Intro 3 for instructions on how to prepare your data.

To specify spatial lags, 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.

Quick start

Spatial autoregressive model of \( y_1 \) regressed on \( x_1, x_2 \), endogenous regressor \( y_2 \), which uses \( z_1 \) as an instrument, and a spatial lag for \( y_1 \) specified by the weighting matrix \( W \)

\[
\text{spivregress } y_1 \ x_1 \ x_2 \ (y_2 = z_1), \ dvarlag(W)
\]

Add an autoregressive error term with the lag given by \( M \)

\[
\text{spivregress } y_1 \ x_1 \ x_2 \ (y_2 = z_1), \ dvarlag(W) \ errorlag(M)
\]

Add a spatial lag for the exogenous variable \( x_1 \) based on \( W \)

\[
\text{spivregress } y_1 \ x_1 \ x_2 \ (y_2 = z_1), \ dvarlag(W) \ errorlag(M) \ ivarlag(W: x_1)
\]

Add a second spatial lag for the outcome variable based on the weighting matrix \( M \)

\[
\text{spivregress } y_1 \ x_1 \ x_2 \ (y_2 = z_1), \ dvarlag(W) \ errorlag(M) \ ///
\text{dvarlag(M) ivarlag(W: x_1)}
\]

Add interaction between \( x_1 \) and \( x_2 \) and add categorical instrument \( z_2 \) using factor variable notation

\[
\text{spivregress } y_1 \ x_1 \ x_2 \ c.x1#c.x2 \ (y_2 = z_1 \ i.z2), \ dvarlag(W) \ ///
\text{errorlag(M) dvarlag(M) ivarlag(W: x_1 x2 c.x1#c.x2)}
\]

Menu

Statistics > Spatial autoregressive models
## Syntax

```
spivregress depvar [varlist1] (varlist2 = varlistiv) [if] [in] [, options]
```

- `varlist1` is the list of included exogenous regressors.
- `varlist2` is the list of endogenous regressors.
- `varlistiv` is the list of excluded exogenous regressors used with `varlist1` as instruments for `varlist2`.

### options

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>dvarlag(spmatname)</code></td>
<td>spatially lagged dependent variable; repeatable</td>
</tr>
<tr>
<td><code>errorlag(spmatname)</code></td>
<td>spatially lagged errors; repeatable</td>
</tr>
<tr>
<td><code>ivarlag(spmatname : varlist)</code></td>
<td>spatially lagged exogenous variables from <code>varlist1</code>; repeatable</td>
</tr>
<tr>
<td><code>noconstant</code></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><code>heteroskedastic</code></td>
<td>treat errors as heteroskedastic</td>
</tr>
<tr>
<td><code>force</code></td>
<td>allow estimation when estimation sample is a subset of the sample used to create the spatial weighting matrix</td>
</tr>
<tr>
<td><code>impower(#)</code></td>
<td>order of instrumental-variable approximation</td>
</tr>
</tbody>
</table>

### Reporting

- `level(#)` | set confidence level; default is `level(95)` |
- `display_options` | control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling |

### Optimization

- `optimization_options` | control the optimization process; seldom used |
- `coeflegend` | display legend instead of statistics |

`varlist1`, `varlist2`, `varlistiv`, and `varlist` specified in `ivarlag()` may contain factor variables; see [U] 11.4.3 Factor variables.

`coeflegend` does not appear in the dialog box.

See [U] 20 Estimation and postestimation commands for more capabilities of estimation commands.

### Options

- `dvarlag(spmatname)` specifies a spatial weighting matrix that defines a spatial lag of the dependent variable. This option is repeatable to allow higher-order models. By default, no spatial lags of the dependent variable are included.
- `errorlag(spmatname)` specifies a spatial weighting matrix that defines a spatially lagged error. This option is repeatable to allow higher-order models. By default, no spatially lagged errors are included.
- `ivarlag(spmatname : varlist)` specifies a spatial weighting matrix and a list of exogenous variables that define spatial lags of the variables. The variables in `varlist` must be a subset of the exogenous variables in `varlist1`. This option is repeatable to allow spatial lags created from different matrices. By default, no spatial lags of the exogenous variables are included.
noconstant; see [R] Estimation options.

heteroskedastic specifies that the estimator treat the errors as heteroskedastic instead of homoskedastic, which is the default; see Methods and formulas in [SP] spregress.

force requests that estimation be done when the estimation sample is a proper subset of the sample used to create the spatial weighting matrices. The default is to refuse to fit the model. Weighting matrices potentially connect all the spatial units. When the estimation sample is a subset of this space, the spatial connections differ and spillover effects can be altered. In addition, the normalization of the weighting matrix differs from what it would have been had the matrix been normalized over the estimation sample. The better alternative to force is first to understand the spatial space of the estimation sample and, if it is sensible, then create new weighting matrices for it. See [SP] spmatrix and Missing values, dropped observations, and the W matrix in [SP] Intro 2.

impower(#) specifies the order of an instrumental-variable approximation used in fitting the model. The derivation of the estimator involves a product of # matrices. Increasing # may improve the precision of the estimation and will not cause harm, but will require more computer time. The default is impower(2). See Methods and formulas for additional details on impower(#).

level(#); see [R] Estimation options.

display_options: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabels, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] Estimation options.

optimization_options: iterate(#), no log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), and nonrtolerance; see [M-5] optimize().

The following option is available with spivregress but is not shown in the dialog box:

coeflegend; see [R] Estimation options.

Remarks and examples

See [SP] Intro for an overview of SAR models.

spivregress fits spatial autoregressive models that include endogenous regressors. The spivregress command is for use with cross-sectional data. It requires each observation to represent one unique spatial unit. See [SP] Intro 3 and the introductory sections that follow for instructions with examples on how to prepare your data for analysis with spivregress.

spivregress fits models like the following:

```
spivregress y1 x1 x2 (y2 y3 = z1 z2 z3), dvarlag(W) errorlag(M) ///
   ivarlag(W: x1)
```

dvarlag(W) specifies a spatial lag of the dependent variable y1, with the formulation of the lag given by the spatial weighting matrix W. You can include multiple dvarlag() options, each with different weighting matrices, to model higher-order spatial lags of the dependent variable.

errorlag(M) specifies an autoregressive error term based on the weighting matrix M. You can include multiple errorlag() options.
ivarlags(W: x1) specifies a spatial lag of the exogenous variable x1. You cannot include in the model spatial lags of the endogenous regressors y2 and y3 or spatial lags of the excluded exogenous regressors z1, z2, and z3.

spivregress uses a generalized method of moments estimator known as generalized spatial two-stage least squares (GS2SLS), the same estimator used by spregress, gs2sls. See Methods and formulas. Also see Choosing weighting matrices and their normalization in [SP] spregress for details about the GS2SLS estimator.

Example 1: SAR models with endogenous regressors

Suppose we want to know whether prohibiting alcohol sales in a county decreases the rate of arrests for driving under the influence (DUI). We use the artificial dataset dui_southern.dta, containing DUI rates in counties in southern states of the United States.

Because the analysis dataset and the Stata-formatted shapefile must be in our working directory to spset the data, we first save both dui_southern.dta and dui_southern.shp.dta to our working directory by using the copy command. We then load the data and type spset to see the Sp settings.

. use dui_southern
. spset

Sp dataset dui_southern.dta
data: cross sectional
spatial-unit id: _ID
coordinates: _CX, _CY (planar)
linked shapefile: dui_southern_shp.dta

The outcome of interest is dui, which is the alcohol-related arrest rate per 100,000 daily vehicle miles traveled (DVMT). Explanatory variables include police, the number of sworn officers per 100,000 DVMT; nondui, the nonalcohol-related arrest rate per 100,000 DVMT; vehicles, the number of registered vehicles per 1,000 residents; and dry, a variable that indicates whether a county prohibits the sale of alcohol within its borders.

Because the size of the police force may be a function of dui and nondui arrest rates, we treat police as endogenous. We assume the variable election is a valid instrument, where election is 1 if the county government faces an election and is 0 otherwise.

We believe the DUI arrest rate to be spatially correlated, with the rate in a county affecting the rates in neighboring counties. Formally, the model we want to fit is

\[
\text{dui} = \beta_0 + \beta_1 \times \text{nondui} + \beta_2 \times \text{dry} + \beta_3 \times \text{vehicles} + \pi_1 \times \text{police} + \lambda W \times \text{dui} + u
\]

The term \(W \times \text{dui}\) defines a spatial lag of dui. See [SP] Intro 2 for an explanation of how spatial lags are defined by weighting matrices, and see Choosing weighting matrices and their normalization in [SP] spregress. The equation for \(u\) gives the error an autoregressive form also specified by the weighting matrix \(W\). The variable police is endogenous and may be correlated with the error \(u\). We instrument it with the variable election. See Methods and formulas for how the endogeneity of police is handled by the estimator.

Before we can fit the model, we must create the weighting matrix \(W\). We will create one that puts the same positive weight on contiguous counties and a 0 weight on all other counties—a matrix known as a contiguity matrix. We will use the default spectral normalization for the matrix. See [SP] Intro 2 and [SP] spmatrix create for details. We type
We fit the model by typing

```
.spmatrix create contiguity W
```

(1422 observations)
(1422 observations (places) used)
(weighting matrix defines 1422 places)

### Estimating rho using 2SLS residuals:

- **initial:** GMM criterion = 0.00254902
- **alternative:** GMM criterion = 0.00377532
- **rescale:** GMM criterion = 0.00009468
- **Iteration 0:** GMM criterion = 0.00009468
- **Iteration 1:** GMM criterion = 0.0001513
- **Iteration 2:** GMM criterion = 0.0001512

### Estimating rho using GS2SLS residuals:

- **Iteration 0:** GMM criterion = 0.00086665
- **Iteration 1:** GMM criterion = 0.00085487
- **Iteration 2:** GMM criterion = 0.00085486

<table>
<thead>
<tr>
<th>Spatial autoregressive model</th>
<th>Number of obs</th>
<th>1,422</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS2SLS estimates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wald chi2(5)</td>
<td>4393.21</td>
<td></td>
</tr>
<tr>
<td>Prob &gt; chi2</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Pseudo R2</td>
<td>0.7378</td>
<td></td>
</tr>
</tbody>
</table>

| dui | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-----|--------|-----------|-------|--------|---------------------|
| dui |        |           |       |        |                     |
| police | -1.283189 | 0.1138994 | -11.27 | 0.000  | -1.506428 -1.059951 |
| nondui | -0.001833 | 0.0025467 | -0.72  | 0.472  | -.0068245 .0031585  |
| vehicles | 0.0906069  | 0.0045059 | 20.11  | 0.000  | .0817755 .0994384  |
| dry |        |           |       |        |                     |
| Yes | 0.4631025 | 0.076754  | 6.03   | 0.000  | .3126674 .6135377  |
| _cons | 8.714745  | 1.060428  | 8.22   | 0.000  | 6.636345 10.79315 |

| W  | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----|--------|-----------|-------|--------|---------------------|
| dui | .3859225 | 0.0194397 | 19.85 | 0.000  | .3478214 .4240235  |
| e.dui | .2169234 | 0.0496595 | 4.37  | 0.000  | .1195926 .3142541  |

Wald test of spatial terms: chi2(2) = 408.78 Prob > chi2 = 0.0000
Instrumented: police (W*dui)
Raw instruments: nondui vehicles 1.dry election dui:_cons
When a spatial lag of the dependent variable is included in the model, covariates have both direct and indirect effects. See example 1 of [SP] spregress for a discussion. To obtain the direct, indirect, and total effects of the covariates, we must use `estat impact`:

```
. estat impact
progress : 20% 40% 60% 80% 100%
Average impacts Number of obs = 1,422

<table>
<thead>
<tr>
<th></th>
<th>Delta-Method</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
<td>z</td>
<td>P&gt;</td>
<td>z</td>
<td></td>
</tr>
<tr>
<td>direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>police</td>
<td>-1.313426</td>
<td>.1198948</td>
<td>-10.95</td>
<td>0.000</td>
<td>-1.548416</td>
<td>-1.078437</td>
</tr>
<tr>
<td>nondui</td>
<td>-.0018762</td>
<td>.0026073</td>
<td>-0.72</td>
<td>0.472</td>
<td>-.0069864</td>
<td>.003234</td>
</tr>
<tr>
<td>vehicles</td>
<td>.092742</td>
<td>.0048427</td>
<td>19.15</td>
<td>0.000</td>
<td>.0832504</td>
<td>1.022336</td>
</tr>
<tr>
<td>dry</td>
<td>.4740151</td>
<td>.0788695</td>
<td>6.01</td>
<td>0.000</td>
<td>.3194336</td>
<td>.6285966</td>
</tr>
<tr>
<td>indirect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>police</td>
<td>-.6465736</td>
<td>.1063216</td>
<td>-6.08</td>
<td>0.000</td>
<td>-.8549601</td>
<td>-.4381871</td>
</tr>
<tr>
<td>nondui</td>
<td>-.0009236</td>
<td>.0012928</td>
<td>-0.71</td>
<td>0.475</td>
<td>-.0034576</td>
<td>.0016103</td>
</tr>
<tr>
<td>vehicles</td>
<td>.045655</td>
<td>.0057216</td>
<td>7.98</td>
<td>0.000</td>
<td>.0344409</td>
<td>.0568692</td>
</tr>
<tr>
<td>dry</td>
<td>.2333482</td>
<td>.0464145</td>
<td>5.03</td>
<td>0.000</td>
<td>.1423774</td>
<td>.3243189</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>police</td>
<td>-1.96</td>
<td>.2258604</td>
<td>-8.68</td>
<td>0.000</td>
<td>-2.402678</td>
<td>-1.517322</td>
</tr>
<tr>
<td>nondui</td>
<td>-.0027998</td>
<td>.0038989</td>
<td>-0.72</td>
<td>0.473</td>
<td>-.0104416</td>
<td>.0048419</td>
</tr>
<tr>
<td>vehicles</td>
<td>.138397</td>
<td>.0105248</td>
<td>13.15</td>
<td>0.000</td>
<td>.1177688</td>
<td>.1590253</td>
</tr>
<tr>
<td>dry</td>
<td>.7073633</td>
<td>.123289</td>
<td>5.74</td>
<td>0.000</td>
<td>.4657213</td>
<td>.9490052</td>
</tr>
</tbody>
</table>
```

While it is running, `estat impact` prints percentages at the top of the output to indicate progress. Calculation of the standard errors of the effects can be intensive and take time, so it reports its progress as it does the computations.

The average direct, or own-county, effect of going from a wet county to a dry county on alcohol-related arrest rates is positive. The average indirect, or spillover, effect of going from a wet county to a dry county on alcohol-related arrest rates is also positive. The total effects are the sum of the direct and indirect effects, so these are also positive.
Example 2: SAR models with endogenous regressors and covariate lags

Continuing with example 1, we found that dry, we now add a spatial lag of the covariate dry.

```
. spivregress dui nondui vehicles i.dry (police = elect), dvarlag(W)
> errorlag(W) ivarlag(W: i.dry)
(1422 observations)
(1422 observations (places) used)
(weighting matrix defines 1422 places)
note: exog*W:0b.dry omitted because of collinearity
```

(output omitted)

```
Spatial autoregressive model
Number of obs = 1,422
GS2SLS estimates
Wald chi2(6) = 4300.29
Prob > chi2 = 0.0000
Pseudo R2 = 0.7337

|         | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---------|-------|-----------|------|------|----------------------|
| dui     |       |           |      |      |                      |
| police  | -1.301634 | .1155866 | -11.26 | 0.000 | -1.52818 -1.075089   |
| nondui  | -.0018725 | .0025746 | -0.73  | 0.467 | -.0069187 .0031737  |
| vehicles| .091364   | .0045754 | 19.97 | 0.000 | .0823965 .1003316   |
| dry     |       |           |      |      |                      |
| Yes     | .4754855 | .078153   | 6.08  | 0.000 | .3223085 .6286626   |
| _cons   | 8.853401 | 1.07409   | 8.24  | 0.000 | 6.748223 10.95858   |

| W       |       |           |      |      |                      |
| dry     |       |           |      |      |                      |
| Yes     | .2868458 | .2209814 | 1.30  | 0.194 | -.1462697 .7199613  |
| dui     | .38758  | .0196366  | 19.74 | 0.000 | .349093  .4260669  |
| e.dui   | .2196418 | .0497708 | 4.41  | 0.000 | .1220929  .3171908  |

Wald test of spatial terms: chi2(3) = 405.90  Prob > chi2 = 0.0000
Instrumented: police (W*dui)
Raw instruments: nondui vehicles 1.dry election (W*0b.dry) (W*1.dry)
                  dui:_cons
We use `estat impact` to see the effects:

```plaintext
. estat impact
progress : 20% 40% 60% 80% 100%
Average impacts

<table>
<thead>
<tr>
<th></th>
<th>Delta-Method</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dy/dx</td>
<td>Std. Err.</td>
</tr>
<tr>
<td>direct</td>
<td></td>
<td></td>
</tr>
<tr>
<td>police</td>
<td>-1.33</td>
<td>0.12</td>
</tr>
<tr>
<td>nondui</td>
<td>0.19</td>
<td>0.00</td>
</tr>
<tr>
<td>vehicles</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>dry</td>
<td>Yes</td>
<td>0.50</td>
</tr>
<tr>
<td>indirect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>police</td>
<td>-0.66</td>
<td>0.11</td>
</tr>
<tr>
<td>nondui</td>
<td>-0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>vehicles</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>dry</td>
<td>Yes</td>
<td>0.62</td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>police</td>
<td>-1.99</td>
<td>0.23</td>
</tr>
<tr>
<td>nondui</td>
<td>-0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>vehicles</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>dry</td>
<td>Yes</td>
<td>1.12</td>
</tr>
</tbody>
</table>
```

The direct effect of `dry` is little changed when we added a lag of `dry`, going from 0.47 to 0.50. But the indirect effects of `dry` go from 0.23 to 0.62. In these fictional data, the indirect effects of `dry` become larger than the direct effects when there is a lag of `dry` in the model.

Note that `spivregress` does not allow the fitting of spatial lags for `police`, our endogenous regressor, nor for `election`, its instrument.

---

**Example 3: SAR models with endogenous regressors and higher-order lags**

In the previous models, we specified all the spatial lags with a single weighting matrix \( W \), a contiguity weighting matrix with the default spectral normalization. Many researchers use a spatial weighting matrix whose \( (i, j) \)th element is the inverse of the distance between units \( i \) and \( j \). With the GS2SLS estimator used by `spivregress`, we can include spatial lags using two spatial weighting matrices. This can be done to model a “higher-order” approximation to the true spatial process. We will now add lags specified by an inverse-distance matrix, using again a spectral normalization of the matrix.
We create the inverse-distance matrix \( M \) and use \texttt{spmatrix dir} to list our Sp matrices.

\[
\text{. spmatrix create idistance } M \\
\text{. spmatrix dir}
\]

<table>
<thead>
<tr>
<th>Weighting matrix name</th>
<th>( N \times N )</th>
<th>Type</th>
<th>Normalization</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M )</td>
<td>1422 x 1422</td>
<td>idistance</td>
<td>spectral</td>
</tr>
<tr>
<td>( W )</td>
<td>1422 x 1422</td>
<td>contiguity</td>
<td>spectral</td>
</tr>
</tbody>
</table>

We fit the model including both weighting matrices for all the lags:

\[
\text{. spivregress dui nondui vehicles i.dry (police = elect), dvarlag(W)} > \text{errorlag(W) ivarlag(W: i.dry) dvarlag(M) errorlag(M) ivarlag(M: i.dry)} \\
\text{(1422 observations)} \\
\text{(1422 observations (places) used)} \\
\text{(weighting matrices define 1422 places)} \\
\text{note: exog*W:0b.dry omitted because of collinearity} \\
\text{note: exog*M:0b.dry omitted because of collinearity} \\
\text{(output omitted)}
\]

Spatial autoregressive model

| Coef.    | Std. Err. | z     | P>|z|     | [95% Conf. Interval] |
|----------|-----------|-------|---------|----------------------|
| dui      |           |       |         |                      |
| police   | -.9762244 | .0782512 | -12.48 | 0.000 | -.1.129594 | -.8228549 |
| nondui   | -.0010538 | .002093 | -0.50   | 0.615 | -.005156 | .0030483   |
| vehicles | .0786503  | .0031164 | 25.24   | 0.000 | .0725423 | .0847582   |
| dry      |           |       |         |                      |
| Yes      | .4207535  | .0631503 | 6.66    | 0.000 | .2969811 | .5445258   |
| _cons    | 6.067724  | .7490414 | 8.10    | 0.000 | 4.599630 | 7.535818   |

| W        |           |       |         |                      |
| dry      |           |       |         |                      |
| Yes      | .2353895  | .2272276 | 1.04    | 0.300 | -.2099684 | .6807474   |
| dui      | .3335312  | .0134259 | 24.84   | 0.000 | .3072169 | .3598455   |
| e.dui    | .2206942  | .0630468 | 3.50    | 0.000 | .0971248 | .3442636   |

| M        |           |       |         |                      |
| dry      |           |       |         |                      |
| Yes      | -.0923513 | 2.70903 | -0.03   | 0.973 | -5.401952 | 5.217249   |
| dui      | .0005204  | .0112677 | 0.05    | 0.963 | -.0215639 | .0226046   |
| e.dui    | -.1069363 | .5910148 | -0.18   | 0.856 | -1.265304 | 1.051431   |

Wald test of spatial terms: \( \text{chi2(6)} = 649.11 \) \( \text{Prob > chi2} = 0.0000 \)

All the spatial lags specified by the inverse-distance matrix \( M \) are nonsignificant. We conclude that there are no inverse-distance-type effects after we account for contiguity-type effects.
 Stored results

`spivregress` stores the following in `e()`:

Scalars
- `e(N)` number of observations
- `e(k)` number of parameters
- `e(df_m)` model degrees of freedom
- `e(df_c)` degrees of freedom for comparison test
- `e(iterations)` number of generalized method of moments iterations
- `e(df_2sls)` number of two-stage least-squares iterations
- `e(rank)` rank of `e(V)`
- `e(r2_p)` pseudo-$R^2$
- `e(chi2)` \( \chi^2 \)
- `e(chi2_c)` \( \chi^2 \) for comparison test
- `e(p)` \( p \)-value for model test
- `e(p_c)` \( p \)-value for test of spatial terms
- `e(converged)` 1 if generalized method of moments converged, 0 otherwise
- `e(converged_2sls)` 1 if two-stage least-squares converged, 0 otherwise

Macros
- `e(cmd)` `spivregress`
- `e(cmdline)` command as typed
- `e(depvar)` name of dependent variable
- `e(indeps)` names of independent variables
- `e(idvar)` name of ID variable
- `e(estimator)` `gs2sls`
- `e(title)` title in estimation output
- `e(constant)` `hasconstant` or `noconstant`
- `e(exogr)` exogenous regressors
- `e(dlmat)` names of spatial weighting matrices applied to `depvar`
- `e(elmat)` names of spatial weighting matrices applied to errors
- `e(het)` heteroskedastic or homoskedastic
- `e(chi2type)` Wald; type of model \( \chi^2 \) test
- `e(properties)` `b V`
- `e(estat_cmd)` program used to implement `estat`
- `e(predict)` program used to implement `predict`
- `e(marginsok)` predictions allowed by `margins`
- `e(marginsnotok)` predictions disallowed by `margins`
- `e(asbalanced)` factor variables `fvset` as `asbalanced`
- `e(asobserved)` factor variables `fvset` as `asobserved`

Matrices
- `e(b)` coefficient vector
- `e(delta_2sls)` two-stage least-squares estimates of coefficients in spatial lag equation
- `e(rho_2sls)` generalized method of moments estimates of coefficients in spatial error equation
- `e(V)` variance–covariance matrix of the estimators

Functions
- `e(sample)` marks estimation sample

Methods and formulas

We consider a cross-sectional spatial autoregressive model with possible endogenous covariates and spatial autoregressive disturbances (SARAR), allowing for higher-order spatial dependence in the dependent variable, the exogenous variables, and the spatial errors. The model is

\[
\begin{align*}
\mathbf{y} &= \sum_{j=1}^{J} \pi_j \mathbf{y}_j + \sum_{k=1}^{K} \beta_k \mathbf{x}_k + \sum_{p=1}^{P} \gamma_p \mathbf{W}_p \mathbf{x}_p + \sum_{r=1}^{R} \lambda_r \mathbf{W}_r \mathbf{y} + \mathbf{u} \\
\mathbf{u} &= \sum_{s=1}^{S} \rho_s \mathbf{M}_s \mathbf{u} + \mathbf{\epsilon}
\end{align*}
\]
where

\[ \mathbf{y} \text{ is an } n \times 1 \text{ vector of observations on the dependent variable;} \]
\[ \tilde{\mathbf{y}}_j \text{ is an } n \times 1 \text{ vector of observations on the } j\text{th endogenous variable}; \pi_j \text{ is the corresponding scalar parameter;} \]
\[ \mathbf{x}_k \text{ is an } n \times 1 \text{ vector of observations on the } k\text{th exogenous variable}; \beta_k \text{ is the corresponding scalar parameter;} \]
\[ \mathbf{W}_p, \mathbf{W}_r, \text{ and } \mathbf{M}_s \text{ are } n \times n \text{ spatial weighting matrices;} \]
\[ \mathbf{W}_p \mathbf{x}, \mathbf{W}_r \mathbf{y}, \text{ and } \mathbf{M}_s \mathbf{u} \text{ are } n \times 1 \text{ spatial lags for the exogenous variable, dependent variable, and error terms;} \gamma_p, \lambda_r, \text{ and } \rho_s \text{ are scalar parameters; and} \]
\[ \mathbf{\epsilon} \text{ is an } n \times 1 \text{ vector of innovations.} \]

The \( J \) endogenous variables \( \tilde{\mathbf{y}}_j \) are correlated with the errors \( \mathbf{u} \). To estimate the model parameters, we need \( Q \) instrumental variables \( \mathbf{x}_e^1, \mathbf{x}_e^2, \ldots, \mathbf{x}_e^Q \) with \( Q \geq J \) that are correlated with the endogenous variables in \( \tilde{\mathbf{y}}_j \) and uncorrelated with the errors \( \mathbf{u} \).

The model in (1) is frequently referred to as a higher-order spatial autoregressive model with spatial autoregressive disturbances, or namely, a SARAR\((R,S)\) model.

The innovations \( \epsilon \) are assumed to be independent and identically distributed or independent but heteroskedastically distributed, where the heteroskedasticity is of unknown form. The generalized spatial two-stage least-squares (GS2SLS) estimator implemented in \texttt{spivregress} produces consistent estimates in both cases when the heteroskedastic option is specified.


Let’s first rewrite (1) in a compact form.

\[
\mathbf{y} = \mathbf{Z}\delta + \mathbf{u} \\
\mathbf{u} = \mathbf{U}\rho + \mathbf{\epsilon} \tag{2}
\]

where

\( \mathbf{Z} \) is the matrix of observations on all the variables in the equation for \( \mathbf{y} \); \( \mathbf{Z} \) contains the endogenous covariates \( \tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_J \), the exogenous covariates \( \mathbf{x}_1, \ldots, \mathbf{x}_K \), the spatially lagged exogenous covariates \( \mathbf{W}\mathbf{x}_1, \ldots, \mathbf{W}\mathbf{x}_P \), and the spatially lagged dependent variables \( \mathbf{W}\mathbf{y}_1, \ldots, \mathbf{W}\mathbf{y}_R \);

\( \mathbf{U} \) contains all the spatial lags of the errors \( \mathbf{u} \) that appear in (1); \( \mathbf{U} \) contains \( \mathbf{M}_1\mathbf{u}, \ldots, \mathbf{M}_S\mathbf{u} \);

\( \delta = (\pi_1, \ldots, \pi_J, \beta_1, \ldots, \beta_K, \gamma_1, \ldots, \gamma_P, \lambda_1, \ldots, \lambda_R)' \) is a vector of all the coefficients on the variables in the equation for \( \mathbf{y} \); and

\( \rho = (\rho_1, \ldots, \rho_S) \) is the vector of coefficients on the spatially lagged errors.

Given these definitions, the estimator implemented in \texttt{spivregress} is a simple extension to the GS2SLS estimator documented in the \textit{Methods and formulas} of \texttt{spregress}.

Specifically, after adding the instrumental variables \( \mathbf{x}_e^1, \mathbf{x}_e^2, \ldots, \mathbf{x}_e^Q \) to the list of exogenous variables \( \mathbf{X}_f \) used to create the matrix of instruments \( \mathbf{H}_1 \) in \texttt{spregress}, the other formulas in \texttt{spregress} specify how the estimator implemented in \texttt{spivregress} works. See \textit{Methods and formulas} in [SP] \texttt{spregress} for further details.
References


Also see

[SP] spivregress postestimation — Postestimation tools for spivregress
[SP] estat moran — Moran’s test of residual correlation with nearby residuals
[SP] Intro — Introduction to spatial data and SAR models
[SP] spmatrix — Categorical guide to the spmatrix command
[SP] spregress — Spatial autoregressive models
[SP] spxtrreregress — Spatial autoregressive models for panel data
[R] ivregress — Single-equation instrumental-variables regression
[U] 20 Estimation and postestimation commands