Stata Data Analysis and Statistical Software

2012 Spanish Stata Users Group meeting

Sppack

The Stata command for Spatial Econometrics

Barcelona
September 2012

Vicente Royuela (AQR-IREA, UB)
vroyuela@ub.edu
http://riscd2.eco.ub.es/~vroyuela/

Recent development in STATA: SPPACK (2012)

.net install sppack.pkg
.net get sppack.pkg

```
Requires: Stata version 11.2

Distribution-Date: 20120125

Author: David M. Drukker , StataCorp LP
Support: email ddrukker@stata.com

Author: Hua Peng, StataCorp LP
Support: email hpeng@stata.com.

Author: Ingmar Prucha, Department of Economics, University of Maryland, College Park, MD.
Support: email prucha@econ.umd.edu

Author: Rafal Raciborski, StataCorp LP
Support: email rraciborski@stata.com
```

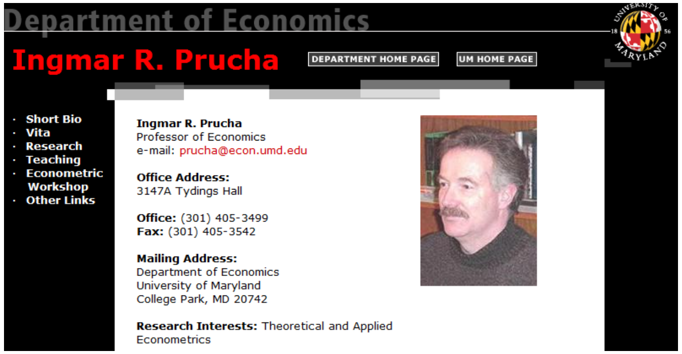
---

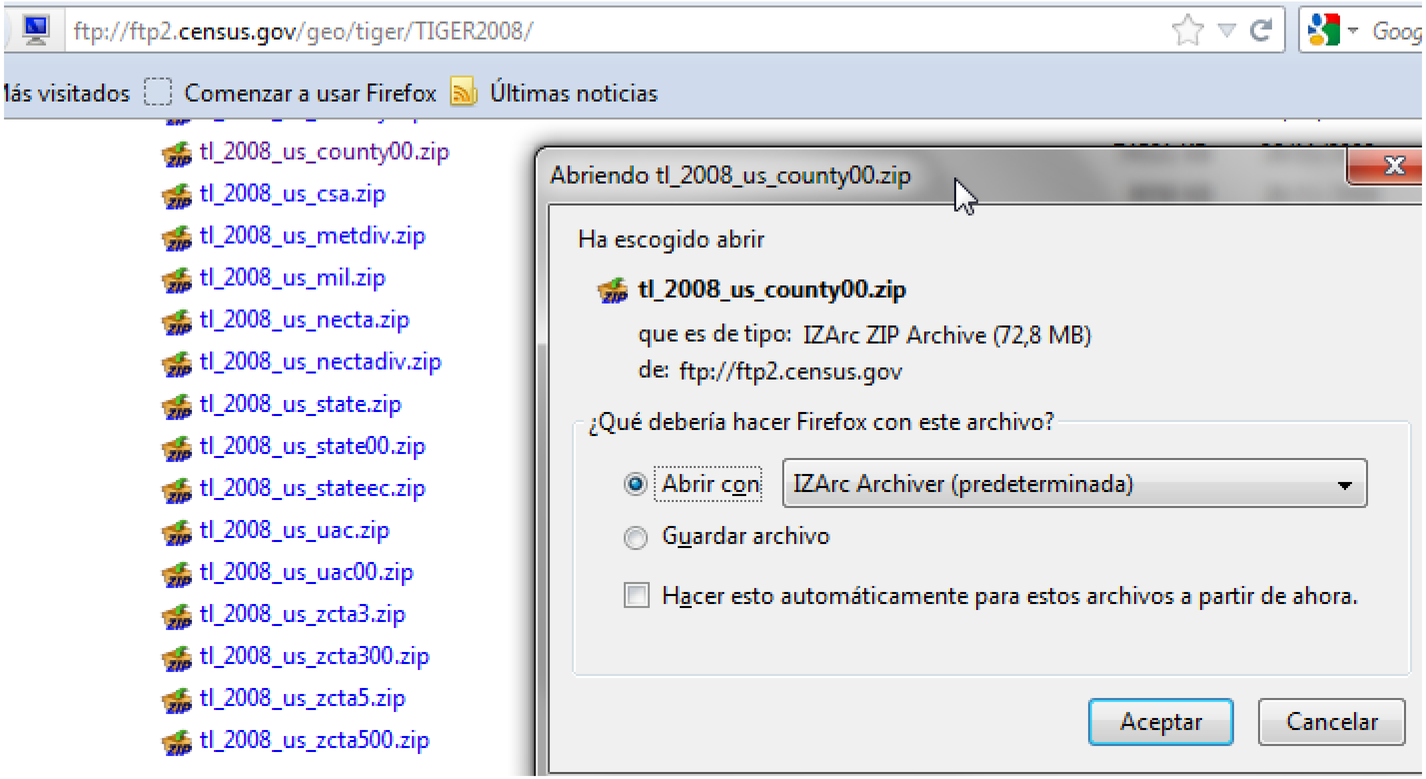
package sppack from http://fmwww.bc.edu/RePEc/bocode/s

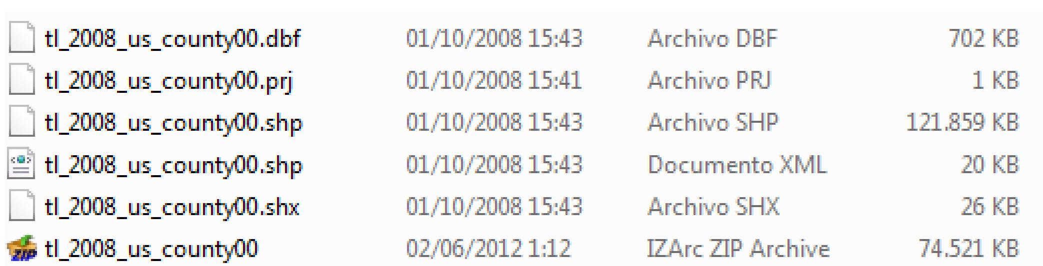
```
TITLE
      'SPPACK': module for cross-section spatial-autoregressive models

DESCRIPTION/AUTHOR(S)

       The spmat, spreg and spivreg commands create spatial-weighting
      matrices, manage spatial-weighting matrices, and estimate the
      parameters of cross-sectional spatial-autoregressive models with
      spatial-autoregressive disturbances that may contain additional
      endogenous covariates.  The estimation method is either maximum
      likelihood or generalized spatial two-stage least squares. The
      working papers that document these commands are available at
      http://econweb.umd.edu/~prucha/Papers/WP_spmat_2011.pdf,
      http://econweb.umd.edu/~prucha/Papers/WP_spreg_2011.pdf, and
      http://econweb.umd.edu/~prucha/Papers/WP_spivreg_2011.pdf .
```

http://econweb.umd.edu/~prucha/Papers/

WP_spmat_2011.pdf
WP_spreg_2011.pdf
WP_spivreg_2011.pdf

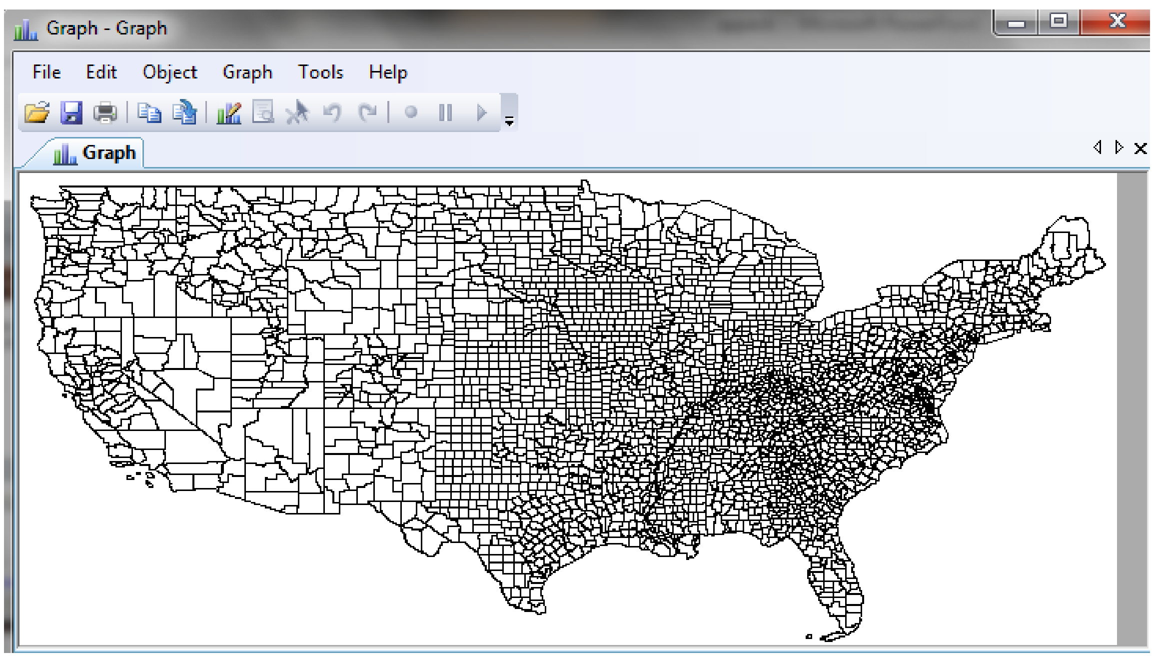



ftp://ftp2.census.gov/geo/tiger/TIGER2008/
lás visitados
Comenzar a usar Firefox
Últimas noticias
tl_2008_us_county00.zip
tl_2008_us_csa.zip
tl_2008_us_metdiv.zip
tl_2008_us_mil.zip
tl_2008_us_necta.zip
tl_2008_us_nectadiv.zip
tl_2008_us_state.zip
tl_2008_us_state00.zip
tl_2008_us_stateec.zip
tl_2008_us_uac.zip
tl_2008_us_uac00.zip
tl_2008_us_zcta3.zip
tl_2008_us_zcta300.zip
tl_2008_us_zcta5.zip
tl_2008_us_zcta500.zip
Abriendo tl_2008_us_county00.zip
Ha escogido abrir
tl_2008_us_county00.zip
que es de tipo: IZArc ZIP Archive (72,8 MB)
de: ftp://ftp2.census.gov
¿Qué debería hacer Firefox con este archivo?
Abrir con
IZArc Archiver (predeterminada)
Guardar archivo
Hacer esto automáticamente para estos archivos a partir de ahora.
Aceptar
Cancelar


tl_2008_us_county00.dbf
01/10/2008 15:43
Archivo DBF
702 KB
tl_2008_us_county00.prj
01/10/2008 15:41
Archivo PRJ
1 KB
tl_2008_us_county00.shp
01/10/2008 15:43
Archivo SHP
121.859 KB
tl_2008_us_county00.shp
01/10/2008 15:43
Documento XML
20 KB
tl_2008_us_county00.shx
01/10/2008 15:43
Archivo SHX
26 KB
tl_2008_us_county00
02/06/2012 1:12
IZArc ZIP Archive
74.521 KB

```
. cd D:\RSAI\SPPACK
D:\RSAI\SPPACK

. shp2dta using tl_2008_us_county00, database(county) coordinates(countyxy) genid(id) gencentroids(c)
```

county	02/06/2012 1:15	Stata Dataset	335 KB
countyxy	02/06/2012 1:15	Stata Dataset	136.952 KB
tl_2008_us_county00.dbf	01/10/2008 15:43	Archivo DBF	702 KB
tl_2008_us_county00.prj	01/10/2008 15:41	Archivo PRJ	1 KB
tl_2008_us_county00.shp	01/10/2008 15:43	Archivo SHP	121.859 KB
tl_2008_us_county00.shp	01/10/2008 15:43	Documento XML	20 KB
tl_2008_us_county00.shx	01/10/2008 15:43	Archivo SHX	26 KB
tl_2008_us_county00	02/06/2012 1:12	IZArc ZIP Archive	74.521 KB



2 Creating a contiguity matrix from geospatial data

`spmat contiguity` computes a contiguity spatial-weighting matrix or a normalized contiguity spatial-weighting matrix from a coordinates dataset and stores it in an `spmat` object.

```
. spmat contiguity ccounty using countyxy, id(id) normalize(minmax)

. spmat summarize ccounty, links

Summary of spatial-weighting object ccounty
```

Matrix	Description
Dimensions	3109 x 3109
Stored as	3109 x 3109
Links	
total	18474
min	1
mean	5.942104
max	14

```
. spmat contiguity ccounty using countyxy, id(id) normalize(row) replace

. spmat summarize ccounty, links

Summary of spatial-weighting object ccounty
```

Matrix	Description
Dimensions	3109 x 3109
Stored as	3109 x 3109
Links	
total	18474
min	1
mean	5.942104
max	14

3 Creating an inverse-distance matrix from data

spmat idistance computes an inverse-distance spatial-weighting matrix from coordinates and stores it in an spmat object.

```
. spmat idistance dcounty longitude latitude, id(id) dfunction(dhaversine)

. spmat summarize dcounty

Summary of spatial-weighting object dcounty
```

Matrix	Description
Dimensions	3109 x 3109
Stored as	3109 x 3109
Values	
min	0
min>0	.0002185
mean	.0012296
max	1.081453

Distances can be :

- euclidean,
- rhaversine,
- dhaversine,
- or p;

Alternative distances

The default behavior of `spmat idistance` is to calculate the Euclidean distance between units s and t, which is given by

$$d_{st} = \sqrt{\sum_{j=1}^{q} \left(\mathrm{x}_j[s] - \mathrm{x}_j[t]\right)^2}$$

The Minkowski distance of order p is given by

$$d_{st} = \sqrt[p]{\sum_{j=1}^{q} \left|\mathrm{x}_j[s] - \mathrm{x}_j[t]\right|^p}$$

The haversine distance measure is useful when the units are located on the surface of the earth and the coordinate variables represent the geographical coordinates of the spatial units. In such cases, we usually wish to calculate a spherical (great-circle) distance between the spatial units. This is accomplished by the haversine formula given by

$$d_{st} = r \times c$$

where

r is the mean radius of the Earth (6,371.009 km or 3,958.761 miles)

$$c = 2\arcsin(\min(1, \sqrt{a}))$$

$$a = \sin^2\phi + \cos(\phi_1)\cos(\phi_2)\sin^2\lambda$$

$$\phi = \frac{1}{2}(\phi_2 - \phi_1) = \frac{1}{2}(\mathrm{x}_2[t] - \mathrm{x}_2[s])$$

$$\lambda = \frac{1}{2}(\lambda_2 - \lambda_1) = \frac{1}{2}(\mathrm{x}_1[t] - \mathrm{x}_1[s])$$

$\mathrm{x}_1[s]$ and $\mathrm{x}_1[t]$ are the longitudes of point s and point t, respectively

$\mathrm{x}_2[s]$ and $\mathrm{x}_2[t]$ are the latitude of point s and point t, respectively

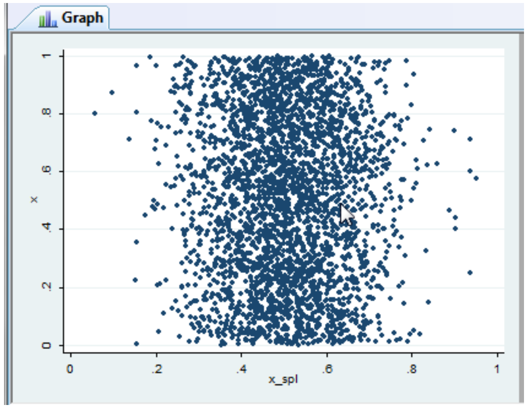
8 Computing spatial lags

`spmat lag` uses a spatial-weighting matrix to compute the weighted averages of a variable known as the spatial lag of a variable.

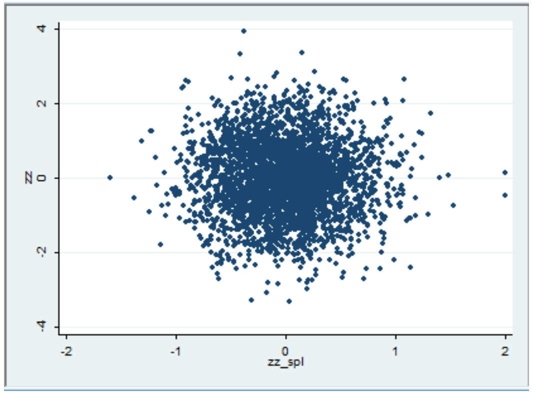
8.1 Syntax

`spmat lag` [*type*] *newvar objname varname*

gen x = **runiform**()
spmat lag x_spl ccounty x
scatter x x_spl



gen zz = **rnormal**()
spmat lag zz_spl ccounty zz
scatter zz zz_spl

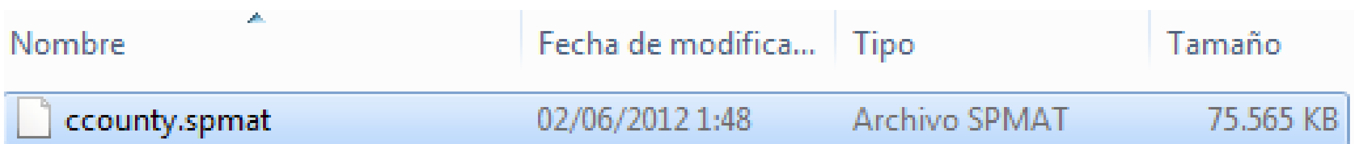


11 Saving an spmat object to disk

spmat save saves an **spmat** object to disk.

11.1 Syntax

spmat save *objname* **using** *filename* **,** [**replace**]



10 Removing an spmat object from memory

spmat drop remove an **spmat** object from memory.

10.1 Syntax

spmat drop *objname*

* Looking at Housing Prices in Barcelona

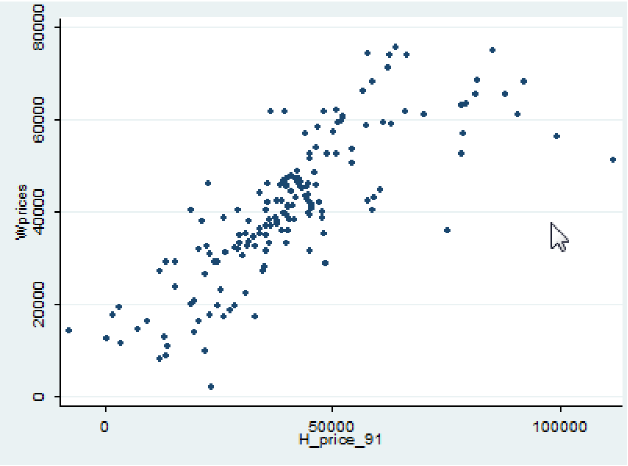
reg H_price_91 pop_91_01 P_25_35_91 P_m65_91 Wealth_91 Unempl_R_91 Cars_91 FUNsis

Source	SS	df	MS
Model	2.9054e+10	7	4.1505e+09
Residual	3.3165e+10	156	212599357
Total	6.2219e+10	163	381712106

Number of obs =	164
F(7, 156) =	19.52
Prob > F =	0.0000
R-squared =	0.4670
Adj R-squared =	0.4430
Root MSE =	14581

H_price_91	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pop_91_01	-2382.528	2999.224	-0.79	0.428	-8306.859	3541.802
P_25_35_91	-171003.2	82135.52	-2.08	0.039	-333244.5	-8761.975
P_m65_91	-242726.9	38283.72	-6.34	0.000	-318348.2	-167105.5
Wealth_91	3460.911	1427.575	2.42	0.016	641.0393	6280.783
Unempl_R_91	417482.8	69639.14	5.99	0.000	279925.4	555040.1
Cars_91	59.29158	18.83766	3.15	0.002	22.08178	96.50137
FUNsis	5578.249	3845.317	1.45	0.149	-2017.358	13173.86
_cons	-8438.675	22796.99	-0.37	0.712	-53469.29	36591.94

spmat lag Wprices cbcn H_price
scatter Wprices H_price



Maximum-likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances

David M. Drukker
StataCorp
College Station, TX
ddrukker@stata.com

Ingmar R. Prucha
Department of Economics
University of Maryland
prucha@econ.umd.edu

Rafal Raciborski
StataCorp
College Station, TX
rraciborski@stata.com

2 A spatial-autoregressive model with spatial-autoregressive disturbances

The spreg command estimates the parameters of the following cross-sectional model ($i = 1, \ldots, n$):

$$y_i = \lambda \sum_{j=1}^{n} w_{ij} y_j + \sum_{p=1}^{k} x_{ip} \beta_p + u_i,$$
$$u_i = \rho \sum_{j=1}^{n} m_{ij} u_j + \varepsilon_i,$$

or more compactly

$$\mathbf{y} = \lambda \mathbf{W} \mathbf{y} + \mathbf{X} \beta + \mathbf{u} \quad (1)$$
$$\mathbf{u} = \rho \mathbf{M} \mathbf{u} + \epsilon \quad (2)$$

The innovations are assumed to be independent and identically distributed (IID) or independent but heteroskedastically distributed, where the heteroskedasticity is of unknown form.

The **GS2SLS** estimator produces consistent estimates in either case, given that the heteroskedastic option is specified in the case of heteroskedasticity.

The **ML** estimator produces consistent estimates in the IID case, but generally not in the heteroskedastic case;

3 `spreg` command

3.1 Syntax

`spreg ml` *depvar* [*indepvars*] [*if*] [*in*] [, *options* *ml_options*]

`spreg gs2sls` *depvar* [*indepvars*] [*if*] [*in*] [, *options* *gs2sls_options*]

The unconcentrated log-likelihood function is

$$\begin{aligned}\ln L(y|\beta,\sigma^2,\lambda,\rho) = & -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \ln||\mathbf{I}-\lambda\mathbf{W}|| + \ln||\mathbf{I}-\rho\mathrm{M}|| \\ & - \frac{1}{2\sigma^2}[(\mathbf{I}-\lambda\mathbf{W})\mathbf{y}-\mathbf{X}\beta]^T(\mathbf{I}-\rho\mathrm{M})^T(\mathbf{I}-\rho\mathrm{M})[(\mathbf{I}-\lambda\mathbf{W})\mathbf{y}-\mathbf{X}\beta] \quad (10)\end{aligned}$$

concentrated log-likelihood

$$L_c(y|\lambda,\rho) \quad = \quad -\frac{n}{2}\{\ln(2\pi)+1\} - \frac{n}{2}\ln(\widehat{\sigma}^2(\lambda,\rho)) + \ln||\mathbf{I}-\lambda\mathbf{W}|| + \ln||\mathbf{I}-\rho\mathrm{M}||.$$

3 `spreg` command

3.1 Syntax

`spreg ml` *depvar* [*indepvars*] [*if*] [*in*] [, *options* *ml_options*]

`spreg gs2sls` *depvar* [*indepvars*] [*if*] [*in*] [, *options* *gs2sls_options*]

The GS2SLS estimator requires instruments. Kelejian and Prucha (1998, 1999) suggest using as instruments $\mathbf{H}$ the linearly independent columns of

$$\mathbf{X}, \mathbf{WX}, \ldots, \mathbf{W}^q\mathbf{X}, \mathbf{MX}, \mathbf{MWX}, \ldots, \mathbf{MW}^q\mathbf{X}$$

where $q = 2$ has worked well in Monte Carlo simulations spreg **ml** H_price_91 pop_91_01 P_25_35_91 P_m65_91 Wealth_91 Unempl_R_91 Cars_91 FUNsis, id(id) dlmat(cbcn) elmat(cbcn) nolog

```
Spatial autoregressive model                     Number of obs   =       164
(Maximum likelihood estimates)                   Wald chi2(7)    =   49.1563
                                                 Prob > chi2     =    0.0000
```

H_price_91	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
H_price_91						
pop_91_01	2263.196	2329.913	0.97	0.331	-2303.351	6829.742
P_25_35_91	-113381.5	63020.92	-1.80	0.072	-236900.2	10137.28
P_m65_91	-138932.9	31249.69	-4.45	0.000	-200181.2	-77684.66
Wealth_91	3842.096	1441.154	2.67	0.008	1017.485	6666.706
Unempl_R_91	242063.6	100910.1	2.40	0.016	44283.4	439843.8
Cars_91	45.99367	13.98261	3.29	0.001	18.58826	73.39908
FUNsis	3684.323	2774.676	1.33	0.184	-1753.942	9122.587
_cons	-29776.77	22369.59	-1.33	0.183	-73620.36	14066.82
lambda						
_cons	.5082226	.2188418	2.32	0.020	.0793007	.9371446
rho						
_cons	.3129106	.3333841	0.94	0.348	-.3405103	.9663315
sigma2						
_cons	1.14e+08	1.28e+07	8.90	0.000	8.91e+07	1.39e+08

spreg **gs2sls** H_price_91 pop_91_01 P_25_35_91 P_m65_91 Wealth_91 Unempl_R_91 Cars_91 FUNsis, id(id) dlmat(cbcn) elmat(cbcb) nolog

```
Spatial autoregressive model                       Number of obs   =      164
(GS2SLS estimates)
```

H_price_91	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
H_price_91						
pop_91_01	2764.139	2068.197	1.34	0.181	-1289.452	6817.731
P_25_35_91	-128539.3	57968.45	-2.22	0.027	-242155.3	-14923.19
P_m65_91	-108808.4	29571.41	-3.68	0.000	-166767.3	-50849.48
Wealth_91	2886.179	910.5099	3.17	0.002	1101.612	4670.746
Unempl_R_91	97547.39	51213.91	1.90	0.057	-2830.036	197924.8
Cars_91	41.55607	13.10802	3.17	0.002	15.86482	67.24733
FUNsis	3003.949	2711.069	1.11	0.268	-2309.65	8317.547
_cons	-16982.28	15236.41	-1.11	0.265	-46845.1	12880.54
lambda						
_cons	.8560284	.0737623	11.61	0.000	.7114568	1.0006
rho						
_cons	-.2822312	.2074991	-1.36	0.174	-.6889219	.1244595

gs2sls_options

dlmat(*objname*) specifies an spmat object that contains the spatial-weighting matrix $\mathbf{W}$ to be used in the spatial-autoregressive term.

elmat(*objname*) specifies an spmat object that contains the spatial-weighting matrix $\mathbf{M}$ to be used in the spatial-error term.

heteroskedastic specifies that spreg use an estimator that allows the errors to be heteroskedastically distributed over the observations. By default, spreg uses an estimator that assumes homoskedasticity.

impower(*q*) specifies how many powers of $\mathbf{W}$ to include in calculating the instrument matrix $\mathbf{H}$ with $q = 2$ being the default. The allowed values of q are integers in the set $2, 3, \ldots, \lfloor\sqrt{n}\rfloor$, where n is the number of observations.

maximize_options control the maximization process; seldom used. The options include iterate(#), [no]log, trace, gradient, showstep, showtolerance, tolerance(#),

and ltolerance(#), see [R] **maximize** for details. from(*init_specs*) is also allowed, but because ρ is the only parameter in this optimization problem, only initial values for ρ can be specified.

```
Spatial autoregressive model                    Number of obs    =      164
(GS2SLS estimates)
```

H_price_91	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
H_price_91						
pop_91_01	3028.06	2209.735	1.37	0.171	-1302.94	7359.06
P_25_35_91	-119237.6	58779.67	-2.03	0.043	-234443.6	-4031.533
P_m65_91	-113486.4	30393.2	-3.73	0.000	-173056	-53916.83
Wealth_91	3239.553	1017.629	3.18	0.001	1245.038	5234.069
Unempl_R_91	117994.7	58534.12	2.02	0.044	3269.953	232719.5
Cars_91	42.86427	13.53228	3.17	0.002	16.34149	69.38704
FUNsis	2997.119	2753.412	1.09	0.276	-2399.469	8393.706
_cons	-22924.09	16315.62	-1.41	0.160	-54902.12	9053.926
lambda						
_cons	.8210292	.0850833	9.65	0.000	.654269	.9877894

No spatial error (rho=0)

```
Spatial autoregressive model                    Number of obs    =      164
(GS2SLS estimates)
```

H_price_91	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
H_price_91						
pop_91_01	1412.64	2521.666	0.56	0.575	-3529.735	6355.014
P_25_35_91	-124146	59350.15	-2.09	0.036	-240470.2	-7821.867
P_m65_91	-168088.7	34459.29	-4.88	0.000	-235627.6	-100549.7
Wealth_91	4015.707	1411.349	2.85	0.004	1249.514	6781.899
Unempl_R_91	390626.4	85413.83	4.57	0.000	223218.4	558034.4
Cars_91	48.93055	14.47935	3.38	0.001	20.55154	77.30956
FUNsis	5295.438	2784.482	1.90	0.057	-162.0464	10752.92
_cons	-24300.99	21280.09	-1.14	0.253	-66009.2	17407.22
rho						
_cons	.6974938	.063693	10.95	0.000	.5726578	.8223297

No spatial lag (lambda=0)

A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables

David M. Drukker
StataCorp
College Station, TX
ddrukker@stata.com

Ingmar R. Prucha
Department of Economics
University of Maryland
College Park, MD
prucha@econ.umd.edu

Rafal Raciborski
StataCorp
College Station, TX
rraciborski@stata.com

2 The model

$$\mathbf{y} = \mathbf{Y}\boldsymbol{\pi} + \mathbf{X}\beta + \lambda\mathbf{W}\mathbf{y} + \mathbf{u} \tag{1}$$

$$\mathbf{u} = \rho\mathbf{M}\mathbf{u} + \epsilon \tag{2}$$

where

- $\mathbf{y}$ is an $n \times 1$ vector of observations on the dependent variable;
- $\mathbf{Y}$ is an $n \times p$ matrix of observations on p RHS endogenous variables, and $\boldsymbol{\pi}$ is the corresponding $p \times 1$ parameter vector;
- $\mathbf{X}$ is an $n \times k$ matrix of observations on k RHS exogenous variables (where some of the variables may be spatial lags of exogenous variables), and β is the corresponding $p \times 1$ parameter vector;
- $\mathbf{W}$ and $\mathbf{M}$ are $n \times n$ spatial-weighting matrices (with zero diagonal elements);
- $\mathbf{W}\mathbf{y}$ and $\mathbf{M}\mathbf{u}$ are $n \times 1$ vectors typically referred to as spatial lags, and λ and ρ are the corresponding scalar parameters typically referred to as spatial-autoregressive parameters;
- ϵ is an $n \times 1$ vector of innovations.

3 spivreg command

3.1 Syntax

`spivreg` *depvar* [*varlist1*] (*varlist2* = *varlist_iv*) [*if*] [*in*] [, *options*]

3.2 Options

`id(`*varname*`)` specifies a numeric variable that contains a unique identifier for each observation. This option is required.

`dlmat(`*objname*`)` specifies an `spmat` object that contains the spatial-weighting matrix **W** to be used in the spatial-autoregressive term.

`elmat(`*objname*`)` specifies an `spmat` object that contains the spatial-weighting matrix **M** to be used in the spatial-error term.

`noconstant` suppresses the constant term.

`heteroskedastic` tells `spivreg` to use the methods developed for heteroskedastic innovations.

`impower(`q`)` specifies how many powers of **W** are to be used in constructing the instrument matrix **H** with $q = 2$ being the default. The allowed values of q are integers in the set $\{2, 3, \ldots, \lfloor\sqrt{n}\rfloor\}$.

`level(#)` sets confidence level; default is `level(95)`.

maximize_options control the maximization process; seldom used. The options include `difficult`, `iterate(#)`, `[no]log`, `trace`, `gradient`, `showstep`, `showtolerance`, `tolerance(#)`, `ltolerance(#)`, and `from(`*init_specs*`)`; see [R] **maximize** for details.

spivreg H_price_91 pop_91_01 P_25_35_91 P_m65_91 (**Wealth_91=Stud_Y_91**) Unempl_R_91 Cars_91 FUNsis, id(id) dlmat(cbcn) elmat(cbcn) nolog

```
Spatial autoregressive model                        Number of obs   =      164
(GS2SLS estimates)
```

H_price_91	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
H_price_91						
Wealth_91	8407.414	2218.251	3.79	0.000	4059.722	12755.11
pop_91_01	2839.639	2391.51	1.19	0.235	-1847.634	7526.912
P_25_35_91	-121047.9	62726.37	-1.93	0.054	-243989.4	1893.498
P_m65_91	-150632	34534.85	-4.36	0.000	-218319.1	-82944.95
Unempl_R_91	107768.5	65681.02	1.64	0.101	-20963.97	236500.9
Cars_91	36.0374	14.80008	2.43	0.015	7.029779	65.04502
FUNsis	3721.101	2945.796	1.26	0.207	-2052.553	9494.756
_cons	-59324.22	22290.37	-2.66	0.008	-103012.5	-15635.9
lambda						
_cons	.7903311	.0928853	8.51	0.000	.6082793	.9723829
rho						
_cons	.0993877	.182029	0.55	0.585	-.2573826	.456158

```
Instrumented:  Wealth_91
Instruments:   Stud_Y_91
```