The Stata Journal (2013)
13, Number 2, pp. 287–301

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

Abstract. We describe the `spivreg` command, which estimates the parameters of linear cross-sectional spatial-autoregressive models with spatial-autoregressive disturbances, where the model may also contain additional endogenous variables as well as exogenous variables. `spivreg` uses results and the literature cited in Kelejian and Prucha (1998, *Journal of Real Estate Finance and Economics* 17: 99–121; 1999, *International Economic Review* 40: 509–533; 2004, *Journal of Econometrics* 118: 27–50; 2010, *Journal of Econometrics* 157: 53–67); Arraiz et al. (2010, *Journal of Regional Science* 50: 592–614); and Drukker, Egger, and Prucha (2013, *Econometric Reviews* 32: 686–733).

Keywords: st0293, spivreg, spatial-autoregressive models, Cliff–Ord models, generalized spatial two-stage least squares, instrumental-variable estimation, generalized method of moments estimation, spatial econometrics, spatial statistics

1 Introduction

Building on the work of Whittle (1954), Cliff and Ord (1973, 1981) developed statistical models that accommodate forms of cross-unit interactions. The latter is a feature of interest in many social science, biostatistical, and geographic science models. A simple version of these models, typically referred to as spatial-autoregressive (SAR) models, augments the linear regression model by including an additional right-hand-side (RHS) variable known as a spatial lag. Each observation of the spatial-lag variable is a weighted average of the values of the dependent variable observed for the other cross-sectional units. Generalized versions of the SAR model also allow for the disturbances to be generated by a SAR process and for the exogenous RHS variables to be spatial lags of exogenous variables. The combined SAR model with SAR disturbances is often referred to as a SARAR model; see Anselin and Florax (1995).[1]

1. These models are also known as Cliff–Ord models because of the impact that Cliff and Ord (1973, 1981) had on the subsequent literature. To avoid confusion, we simply refer to these models as SARAR models while still acknowledging the importance of the work of Cliff and Ord.

In modeling the outcome for each unit as dependent on a weighted average of the outcomes of other units, SARAR models determine outcomes simultaneously. This simultaneity implies that the ordinary least-squares estimator will not be consistent; see Anselin (1988) for an early discussion of this point. Drukker, Prucha, and Raciborski (2013) discuss the `spreg` command, which implements estimators for the model when the RHS variables are a spatial lag of the dependent variable, exogenous variables, and spatial lags of the exogenous variables.

The model we consider allows for additional endogenous RHS variables. Thus the model of interest is a linear cross-sectional SAR model with additional endogenous variables, exogenous variables, and SAR disturbances. We discuss an estimator for the parameters of this model and the command that implements this estimator, `spivreg`. Kelejian and Prucha (1998, 1999, 2004, 2010) and the references cited therein derive the main results used by the estimator implemented in `spivreg`, with Drukker, Egger, and Prucha (2013) and Arraiz et al. (2010) producing some important extensions that are used in the code.

While SARAR models have a wide range of possible applications, following Cliff and Ord (1973, 1981), much of the original literature was developed to handle spatial interactions; see, for example, Anselin (1988, 2010), Cressie (1993), and Haining (2003). However, space is not restricted to geographic space, and many recent applications employ these techniques in other situations of cross-unit dependence, such as social-interaction models and network models; see, for example, Kelejian and Prucha (2010) and Drukker, Egger, and Prucha (2013) for references. Much of the nomenclature still includes the adjective "spatial", and we continue this tradition to avoid confusion while noting the wider applicability of these models.

Section 2 defines the generalized SARAR model. Section 3 describes the `spivreg` command. Section 4 illustrates the estimation of a SARAR model on example data for U.S. counties. Section 5 describes postestimation commands. Section 6 presents methods and formulas. The conclusion follows.

2 The model

The model of interest is given by

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

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

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 $\boldsymbol{\beta}$ is the corresponding $p \times 1$ parameter vector;
- $\mathbf{W}$ and $\mathbf{M}$ are $n \times n$ spatial-weighting matrices (with 0 diagonal elements);
- $\mathbf{Wy}$ and $\mathbf{Mu}$ are $n \times 1$ vectors typically referred to as spatial lags, and λ and ρ are the corresponding scalar parameters typically referred to as SAR parameters;
- $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of innovations.[2]

The model in equations (1) and (2) is a SARAR model with exogenous regressors and additional endogenous regressors. Spatial interactions are modeled through spatial lags, and the model allows for spatial interactions in the dependent variable, the exogenous variables, and the disturbances.

Because the model in equations (1) and (2) is a first-order SAR process with first-order SAR disturbances, it is also referred to as a SARAR(1,1) model, which is a special case of the more general SARAR(p, q) model. We refer to a SARAR(1,1) model as a SARAR model. Setting $\rho = 0$ yields the SAR model $\mathbf{y} = \mathbf{Y}\boldsymbol{\pi} + \mathbf{X}\boldsymbol{\beta} + \lambda\mathbf{Wy}+\boldsymbol{\epsilon}$. Setting $\lambda = 0$ yields the model $\mathbf{y} = \mathbf{Y}\boldsymbol{\pi} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ with $\mathbf{u} = \rho\mathbf{Mu} + \boldsymbol{\epsilon}$, which is sometimes referred to as the SAR error model. Setting $\rho = 0$ and $\lambda = 0$ causes the model to reduce to a linear regression model with endogenous variables.

The spatial-weighting matrices $\mathbf{W}$ and $\mathbf{M}$ are taken to be known and nonstochastic. These matrices are part of the model definition, and in many applications, $\mathbf{W} = \mathbf{M}$; see Drukker et al. (2013) for more about creating spatial-weighting matrices in Stata. Let $\overline{\mathbf{y}}= \mathbf{Wy}$, let $\overline{y}_i$ and y_i denote the ith element of $\overline{\mathbf{y}}$ and $\mathbf{y}$, respectively, and let w_{ij} denote the (i, j)th element of $\mathbf{W}$. Then

$$\overline{y}_i = \sum_{j=1}^{n} w_{ij} y_j$$

which clearly shows the dependence of y_i on neighboring outcomes via the spatial lag $\overline{y}_i$. The weights w_{ij} will typically be modeled as inversely related to some measure of distance between the units. The SAR parameter λ measures the extent of these interactions.

The innovations $\boldsymbol{\epsilon}$ are assumed to be independent and identically distributed or independent but heteroskedastically distributed. The option `heteroskedastic`, discussed below, should be specified under the latter assumption.

The `spivreg` command implements the generalized method of moments (GMM) and instrumental-variable (IV) estimation strategy discussed in Arraiz et al. (2010) and

2. The variables and parameters in this model are allowed to depend on the sample size; see Kelejian and Prucha (2010) for further discussions. We suppress this dependence for notational simplicity. In allowing, in particular, the elements of $\mathbf{X}$ to depend on the sample size, we find that the specification is consistent with some of the variables in $\mathbf{X}$ being spatial lags of exogenous variables.

Drukker, Egger, and Prucha (2013) for the above class of SARAR models. This estimation strategy builds on Kelejian and Prucha (1998, 1999, 2004, 2010) and the references cited therein. More in-depth discussions regarding issues of model specifications and estimation approaches can be found in these articles and the literature cited therein.

`spivreg` requires that the spatial-weighting matrices **M** and **W** be provided in the form of an `spmat` object as described in Drukker et al. (2013). Both general and banded spatial-weighting matrices are supported.

3 The spivreg command

3.1 Syntax

`spivreg` *depvar* [*varlist1*] `(`*varlist2* `=` [*varlist_iv*]`)` [*if*] [*in*]`, id(`*varname*`)`
[`dlmat(`*objname*`)` `elmat(`*objname*`)` `noconstant` `heteroskedastic` `impower(`*q*`)`
`level(`#`)` *maximize_options*]

3.2 Options

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

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

`heteroskedastic` specifies that `spivreg` use an estimator that allows **e** to be heteroskedastically distributed over the observations. By default, `spivreg` uses an estimator that assumes homoskedasticity.

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

`level(`#`)` specifies the confidence level, as a percentage, for confidence intervals. The default is `level(95)` or as set by `set level`.

maximize_options: `iterate(`#`)`, [`no`]`log`, `trace`, `gradient`, `showstep`, `showtolerance`, `tolerance(`#`)`, `ltolerance(`#`)`, and `from(`*init_specs*`)`; see [R] **maximize** for details. These options are seldom used.

3.3 Saved results

`spivreg` saves the following information in `e()`:

Scalars

`e(N)`	number of observations	`e(converged)`	`1` if GMM stage converged, `0` otherwise
`e(k)`	number of parameters		
`e(rho_2sls)`	initial estimate of ρ		
`e(iterations)`	number of GMM iterations	`e(converged_2sls)`	`1` if 2SLS stage converged, `0` otherwise
`e(iterations_2sls)`	number of 2SLS iterations		

Macros

`e(cmd)`	`spivreg`	`e(exogr)`	exogenous regressors
`e(cmdline)`	command as typed	`e(insts)`	instruments
`e(depvar)`	name of dependent variable	`e(instd)`	instrumented variables
`e(title)`	title in estimation output	`e(constant)`	`noconstant` or `hasconstant`
`e(properties)`	`b V`		
`e(estat_cmd)`	program used to implement `estat`	`e(H_omitted)`	names of omitted instruments in `H` matrix
`e(predict)`	program used to implement `predict`	`e(idvar)`	name of ID variable
`e(model)`	`sarar`, `sar`, `sare`, or `lr`	`e(dlmat)`	name of `spmat` object used in `dlmat()`
`e(het)`	`heteroskedastic` or `homoskedastic`	`e(elmat)`	name of `spmat` object used in `elmat()`
`e(indeps)`	names of independent variables		

Matrices

`e(b)`	coefficient vector	`e(delta_2sls)`	initial estimate of β and λ
`e(V)`	variance–covariance matrix of the estimators		

Functions

`e(sample)`	marks estimation sample

4 Examples

To provide a simple illustration, we use the artificial dataset `spivreg.dta` for the continental U.S. counties.[3] The contiguity matrix for the U.S. counties is taken from Drukker et al. (2013). In Stata, we issue the following commands:

```
. use dui
. spmat use ccounty using ccounty.spmat
```

The spatial-weighting matrix is now contained in the `spmat` object `ccounty`. This minmax-normalized spatial-weighting matrix was created in section 2.4 of Drukker et al. (2013) and was saved to disk in section 11.4.

In the output above, we are just reading in the spatial-weighting-matrix object that was created and saved in Drukker et al. (2013).

3. The geographical county location data came from the U.S. Census Bureau and can be found at ftp://ftp2.census.gov/geo/tiger/TIGER2008/. The variables are simulated but inspired by Powers and Wilson (2004) and Levitt (1997).

Our dependent variable, `dui`, is defined as the alcohol-related arrest rate per 100,000 daily vehicle miles traveled (DVMT). Figure 1 shows the distribution of `dui` across counties, with darker colors representing higher values of the dependent variable. Spatial patterns in `dui` are clearly visible.

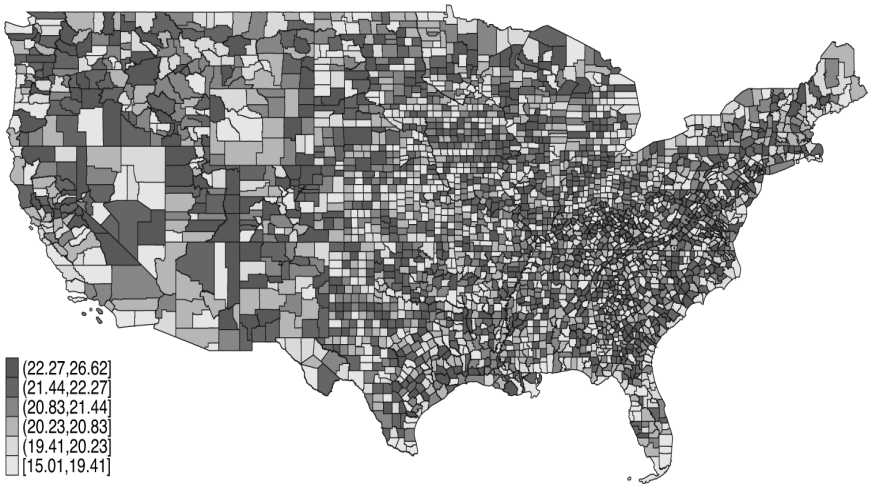

Figure 1. Hypothetical alcohol-related arrests for continental U.S. counties

Our explanatory variables include `police` (number of sworn officers per 100,000 DVMT); `nondui` (nonalcohol-related arrests per 100,000 DVMT); `vehicles` (number of registered vehicles per 1,000 residents); and `dry` (a dummy for counties that prohibit alcohol sale within their borders). Because the size of the police force may be a function of `dui` arrest rates, we treat `police` as endogenous; that is, in this example, $\mathbf{Y} =$ (`police`). All other included explanatory variables, apart from the spatial lag, are taken to be exogenous; that is, $\mathbf{X} =$ (`nondui`, `vehicles`, `dry`, `intercept`). Furthermore, we assume the variable `elect` is a valid instrument, where `elect` is 1 if a county government faces an election and is 0 otherwise. Thus the instrument matrix $\mathbf{H}$ is based on $\mathbf{X}_f =$ (`nondui`, `vehicles`, `dry`, `elect`, `intercept`) as described above.

In Stata, we can estimate the SARAR model with endogenous variables by typing

```
. spivreg dui nondui vehicles dry (police = elect), id(id)
> dlmat(ccounty) elmat(ccounty) nolog
Spatial autoregressive model                    Number of obs   =     3109
(GS2SLS estimates)
```

dui	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
dui						
police	-1.467068	.0434956	-33.73	0.000	-1.552318	-1.381818
nondui	-.0004088	.0008344	-0.49	0.624	-.0020442	.0012267
vehicles	.0989662	.0017653	56.06	0.000	.0955063	.1024261
dry	.4553992	.0278049	16.38	0.000	.4009026	.5098958
_cons	9.671655	.3682685	26.26	0.000	8.949862	10.39345
lambda						
_cons	.7340818	.013378	54.87	0.000	.7078614	.7603023
rho						
_cons	.2829313	.071908	3.93	0.000	.1419941	.4238685

```
Instrumented:  police
Instruments:   elect
```

Given the normalization of the spatial-weighting matrix, the parameter space for λ and ρ is taken to be the interval $(-1, 1)$; see Kelejian and Prucha (2010) for further discussions of the parameter space. The estimate of λ is positive, large, and significant, indicating strong SAR dependence in `dui`. In other words, the alcohol-related arrest rate for a given county is strongly affected by the alcohol-related arrest rates in the neighboring counties. One possible explanation for this may be coordination among police departments. Another may be that strong enforcement in one county may lead some people to drink in neighboring counties.

The estimated ρ is positive, moderate, and significant, indicating moderate spatial autocorrelation in the innovations.

The estimated $\boldsymbol{\beta}$ vector does not have the same interpretation as in a simple linear model, because including a spatial lag of the dependent variable implies that the outcomes are determined simultaneously.

5 Postestimation commands

5.1 Syntax

The syntax for `predict` after `spivreg` is

`predict` [*type*] *newvar* [*if*] [*in*] [`,` *statistic*]

where *statistic* is one of the following:

`naive`, the default, computes $\mathbf{Y}\widehat{\boldsymbol{\pi}} + \mathbf{X}\widehat{\boldsymbol{\beta}} + \widehat{\lambda}\mathbf{W}\mathbf{y}$, which should not be viewed as a predictor for y_i but simply as an intermediate calculation.

`xb` calculates $\mathbf{Y}\widehat{\boldsymbol{\pi}} + \mathbf{X}\widehat{\boldsymbol{\beta}}$.

The predictor computed by the option `naive` will generally be biased; see Kelejian and Prucha (2007) for an explanation. Optimal predictors for the SARAR model with additional endogenous RHS variables corresponding to different information sets will be made available in the future. Optimal predictors for the SARAR model without additional endogenous RHS variables are discussed in Kelejian and Prucha (2007).

6 Methods and formulas

In this section, we give a detailed description of the calculations performed by `spivreg`. We first discuss the estimation of the general model as specified in (1) and (2), both under the assumption that the innovations $\boldsymbol{\epsilon}$ are homoskedastic and under the assumption that the innovations $\boldsymbol{\epsilon}$ are heteroskedastic of unknown form. We then discuss the two special cases $\rho = 0$ and $\lambda = 0$, respectively.

6.1 SARAR model

It is helpful to rewrite the model in (1) and (2) as

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u}$$
$$\mathbf{u} = \rho\mathbf{M}\mathbf{u} + \boldsymbol{\epsilon}$$

where $\mathbf{Z} = (\mathbf{Y}, \mathbf{X}, \mathbf{W}\mathbf{y})$ and $\boldsymbol{\delta} = (\boldsymbol{\pi}', \boldsymbol{\beta}', \lambda)'$. In the following, we review the two-step GMM and IV estimation approach as discussed in Drukker, Egger, and Prucha (2013) for the homoskedastic case and in Arraiz et al. (2010) for the heteroskedastic case. Those articles build on and specialize the estimation theory developed in Kelejian and Prucha (1998, 1999, 2004, 2010). A full set of assumptions, formal consistency and asymptotic normality theorems, and further details and discussions are given in that literature.

The IV estimators $\boldsymbol{\delta}$ depend on the choice of a set of instruments, say, $\mathbf{H}$. Suppose that in addition to the included exogenous variables $\mathbf{X}$, we also have excluded exogenous variables $\mathbf{X}_e$, allowing us to define $\mathbf{X}_f = (\mathbf{X}, \mathbf{X}_e)$. If we do not have excluded exogenous

variables, then $\mathbf{X}_f = \mathbf{X}$. Following the above literature, the instruments $\mathbf{H}$ may then be taken as the linearly independent columns of

$$(\mathbf{X}_f, \mathbf{W}\mathbf{X}_f, \ldots, \mathbf{W}^q\mathbf{X}_f, \mathbf{M}\mathbf{X}_f, \mathbf{M}\mathbf{W}\mathbf{X}_f, \ldots, \mathbf{M}\mathbf{W}^q\mathbf{X}_f)$$

The motivation for the above instruments is that they are computationally simple while facilitating an approximation of the ideal instruments under reasonable assumptions. Taking $q = 2$ has worked well in Monte Carlo simulations over a wide range of specifications. At a minimum, the instruments should include the linearly independent columns of $\mathbf{X}_f$ and $\mathbf{M}\mathbf{X}_f$, and the rank of $\mathbf{H}$ should be at least the number of variables in $\mathbf{Z}$.[4] For the following discussion, it proves convenient to define the instrument projection matrix $\mathbf{P}_\mathbf{H} = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$. When there is a constant in the model, it is only included once in $\mathbf{H}$.

The GMM estimators for ρ are motivated by quadratic moment conditions of the form

$$E\left(\boldsymbol{\epsilon}'\mathbf{A}_s\boldsymbol{\epsilon}\right) = \mathbf{0}, \qquad s = 1, \ldots, S$$

where the matrices $\mathbf{A}_s$ satisfy $\mathrm{tr}(\mathbf{A}_s) = 0$. Specific choices for those matrices will be given below. We note that under heteroskedasticity, it is furthermore assumed that the diagonal elements of the matrices $\mathbf{A}_s$ are 0. This assumption simplifies the formula for the asymptotic variance–covariance (VC) matrix; in particular, it avoids the fact that the VC matrix must depend on third and fourth moments of the innovations in addition to second moments.

We next describe the steps involved in computing the GMM and IV estimators and an estimate of their asymptotic VC matrix. The second step operates on a spatial Cochrane–Orcutt transformation of the above model given by

$$\mathbf{y}(\rho) = \mathbf{Z}(\rho)\boldsymbol{\delta} + \boldsymbol{\epsilon}$$

with $\mathbf{y}(\rho) = (\mathbf{I}_n - \rho\mathbf{M})\mathbf{y}$ and $\mathbf{Z}(\rho) = (\mathbf{I}_n - \rho\mathbf{M})\mathbf{Z}$.

Step 1a: Two-stage least-squares estimator

In the first step, we apply two-stage least squares (2SLS) to the untransformed model by using the instruments $\mathbf{H}$. The 2SLS estimator of $\boldsymbol{\delta}$ is then given by

$$\widetilde{\boldsymbol{\delta}} = \left(\widetilde{\mathbf{Z}}'\mathbf{Z}\right)^{-1}\widetilde{\mathbf{Z}}'\mathbf{y}$$

where $\widetilde{\mathbf{Z}} = \mathbf{P}_\mathbf{H}\mathbf{Z}$.

4. Note that if $\mathbf{X}_f$ contains spatially lagged variables, $\mathbf{H}$ will contain collinear columns and will not be full rank. In those cases, we drop collinear columns from $\mathbf{H}$ and return the names of omitted instruments in `e(H_omitted)`.

Step 1b: Initial GMM estimator of ρ

The initial GMM estimator of ρ is given by

$$\widetilde{\rho} = \arg\min \left[\left\{ \widetilde{\mathbf{\Gamma}} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \widetilde{\boldsymbol{\gamma}} \right\}' \left\{ \widetilde{\mathbf{\Gamma}} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \widetilde{\boldsymbol{\gamma}} \right\} \right]$$

where $\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\widetilde{\boldsymbol{\delta}}$ are the 2SLS residuals, $\widetilde{\overline{\mathbf{u}}} = \mathbf{M}\widetilde{\mathbf{u}}$,

$$\widetilde{\mathbf{\Gamma}} = n^{-1} \begin{bmatrix} \widetilde{\mathbf{u}}'(\mathbf{A}_1 + \mathbf{A}_1')\widetilde{\overline{\mathbf{u}}} & -\widetilde{\overline{\mathbf{u}}}'\mathbf{A}_1\widetilde{\overline{\mathbf{u}}} \\ \vdots & \vdots \\ \widetilde{\mathbf{u}}'(\mathbf{A}_S + \mathbf{A}_S')\widetilde{\overline{\mathbf{u}}} & -\widetilde{\overline{\mathbf{u}}}'\mathbf{A}_s\widetilde{\overline{\mathbf{u}}} \end{bmatrix} \text{ and } \widetilde{\boldsymbol{\gamma}} = n^{-1} \begin{bmatrix} \widetilde{\mathbf{u}}'\mathbf{A}_1\widetilde{\mathbf{u}} \\ \vdots \\ \widetilde{\mathbf{u}}'\mathbf{A}_S\widetilde{\mathbf{u}} \end{bmatrix}$$

Writing the GMM estimator in this form shows that we can calculate it by solving a simple nonlinear least-squares problem. By default, $S = 2$ and `homoskedastic` is specified. In this case,

$$\mathbf{A}_1 = \left[1 + \left\{ n^{-1}\mathrm{tr}(\mathbf{M}'\mathbf{M}) \right\}^2 \right]^{-1} \left\{ \mathbf{M}'\mathbf{M} - n^{-1}\mathrm{tr}(\mathbf{M}'\mathbf{M})\mathbf{I}_n \right\}$$

and

$$\mathbf{A}_2 = \mathbf{M}$$

If `heteroskedastic` is specified, then by default,

$$\mathbf{A}_1 = \mathbf{M}'\mathbf{M} - \mathrm{diag}(\mathbf{M}'\mathbf{M})$$

and

$$\mathbf{A}_2 = \mathbf{M}$$

Step 2a: Generalized spatial two-stage least-squares estimator of $\boldsymbol{\delta}$

In the second step, we first estimate $\boldsymbol{\delta}$ by 2SLS from the transformed model by using the instruments $\mathbf{H}$ and from where the spatial Cochrane–Orcutt transformation uses $\widetilde{\rho}$. The resulting generalized spatial two-stage least-squares (GS2SLS) estimator of δ is now given by

$$\widehat{\boldsymbol{\delta}}(\widetilde{\rho}) = \left\{ \widehat{\mathbf{Z}}(\widetilde{\rho})'\mathbf{Z}(\widetilde{\rho}) \right\}^{-1} \widehat{\mathbf{Z}}(\widetilde{\rho})'\mathbf{y}(\widetilde{\rho})$$

where $\mathbf{y}(\widetilde{\rho}) = (\mathbf{I}_n - \widetilde{\rho}\mathbf{M})\mathbf{y}$, $\mathbf{Z}(\widetilde{\rho}) = (\mathbf{I}_n - \widetilde{\rho}\mathbf{M})\mathbf{Z}$, and $\widehat{\mathbf{Z}}(\widetilde{\rho}) = \mathbf{P}_H\mathbf{Z}(\widetilde{\rho})$.

Step 2b: Efficient GMM estimator of ρ

The efficient GMM estimator of ρ corresponding to GS2SLS residuals is given by

$$\widehat{\rho} = \arg\min \left[\left\{ \widehat{\mathbf{\Gamma}} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \widehat{\boldsymbol{\gamma}} \right\}' \left\{ \widehat{\mathbf{\Psi}}^{\rho\rho}(\widetilde{\rho}) \right\}^{-1} \left\{ \widehat{\mathbf{\Gamma}} \begin{pmatrix} \rho \\ \rho^2 \end{pmatrix} - \widetilde{\boldsymbol{\gamma}} \right\} \right]$$

where $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\delta}}$ denotes the GS2SLS residuals, $\widehat{\overline{\mathbf{u}}} = \mathbf{M}\widehat{\mathbf{u}}$,

$$\widehat{\boldsymbol{\Gamma}} = n^{-1}\begin{bmatrix} \widehat{\mathbf{u}}'(\mathbf{A}_1 + \mathbf{A}_1')\widehat{\overline{\mathbf{u}}} & -\widehat{\overline{\mathbf{u}}}'\mathbf{A}_1\widehat{\overline{\mathbf{u}}} \\ \vdots & \vdots \\ \widehat{\mathbf{u}}'(\mathbf{A}_S + \mathbf{A}_S')\widehat{\overline{\mathbf{u}}} & -\widehat{\overline{\mathbf{u}}}'\mathbf{A}_s\widehat{\overline{\mathbf{u}}} \end{bmatrix} \text{ and } \widehat{\boldsymbol{\gamma}} = n^{-1}\begin{bmatrix} \widehat{\mathbf{u}}'\mathbf{A}_1\widehat{\mathbf{u}} \\ \vdots \\ \widehat{\mathbf{u}}'\mathbf{A}_S\widehat{\mathbf{u}} \end{bmatrix}$$

and where $\widehat{\boldsymbol{\Psi}}^{\rho\rho}(\widetilde{\rho})$ is an estimator for the VC matrix of the (normalized) sample moment vector based on GS2SLS residuals, say, $\boldsymbol{\Psi}^{\rho\rho}$. The estimator $\widehat{\boldsymbol{\Psi}}^{\rho\rho}(\widetilde{\rho})$ and $\boldsymbol{\Psi}^{\rho\rho}$ differ for the cases of homoskedastic and heteroskedastic errors. When `homoskedastic` is specified, the r, s element of $\widehat{\boldsymbol{\Psi}}^{\rho\rho}(\widetilde{\rho})$ is given by $(r, s = 1, 2)$,

$$\begin{aligned} \widehat{\boldsymbol{\Psi}}^{\rho\rho}_{r,s}(\widetilde{\rho}) = {} & \left\{\widetilde{\sigma}^2(\widetilde{\rho})\right\}^2 (2n)^{-1}\mathrm{tr}\left\{(\mathbf{A}_r + \mathbf{A}_r')(\mathbf{A}_s + \mathbf{A}_s')\right\} \\ & + \widetilde{\sigma}^2(\widetilde{\rho})\, n^{-1}\widetilde{\mathbf{a}}_r(\widetilde{\rho})'\widetilde{\mathbf{a}}_s(\widetilde{\rho}) \\ & + n^{-1}\left[\widetilde{\mu}^{(4)}(\widetilde{\rho}) - 3\left\{\widetilde{\sigma}^2(\widetilde{\rho})\right\}^2\right]\mathrm{vec}_D(\mathbf{A}_r)'\,\mathrm{vec}_D(\mathbf{A}_s) \\ & + n^{-1}\widetilde{\mu}^{(3)}(\widetilde{\rho})\left\{\widetilde{\mathbf{a}}_r(\widetilde{\rho})'\,\mathrm{vec}_D(\mathbf{A}_s) + \widetilde{\mathbf{a}}_s(\widetilde{\rho})'\mathrm{vec}_D(\mathbf{A}_r)\right\} \end{aligned} \tag{3}$$

where

$$\begin{aligned} \widehat{\mathbf{a}}_r(\widetilde{\rho}) &= \widehat{\mathbf{T}}(\widetilde{\rho})\widehat{\boldsymbol{\alpha}}_r(\widetilde{\rho}) \\ \widehat{\mathbf{T}}(\widetilde{\rho}) &= \mathbf{H}\widehat{\mathbf{P}}(\widetilde{\rho}) \\ \widehat{\mathbf{P}}(\widetilde{\rho}) &= \widehat{\mathbf{Q}}_{HH}^{-1}\widehat{\mathbf{Q}}_{HZ}(\widetilde{\rho})\left\{\widehat{\mathbf{Q}}_{HZ}(\widetilde{\rho})'\widehat{\mathbf{Q}}_{HH}^{-1}\widehat{\mathbf{Q}}_{HZ}(\widetilde{\rho})'\right\}^{-1} \\ \widehat{Q}_{HH} &= \left(n^{-1}\mathbf{H}'\mathbf{H}\right) \\ \widehat{Q}_{HZ}(\widetilde{\rho}) &= \left\{n^{-1}\mathbf{H}'\mathbf{Z}(\widetilde{\rho})\right\} \\ \mathbf{Z}(\widetilde{\rho}) &= (\mathbf{I} - \widetilde{\rho}\mathbf{M})\mathbf{Z} \\ \widetilde{\boldsymbol{\alpha}}_r(\widetilde{\rho}) &= -n^{-1}\left\{\mathbf{Z}(\widetilde{\rho})'(\mathbf{A}_r + \mathbf{A}_r')\widehat{\boldsymbol{\epsilon}}(\widetilde{\rho})\right\} \\ \widehat{\boldsymbol{\epsilon}}(\widetilde{\rho}) &= (\mathbf{I} - \widetilde{\rho}\mathbf{M})\widehat{\mathbf{u}} \\ \widehat{\sigma}^2(\widetilde{\rho}) &= n^{-1}\widehat{\boldsymbol{\epsilon}}(\widetilde{\rho})'\widehat{\boldsymbol{\epsilon}}(\widetilde{\rho}) \\ \widehat{\mu}^{(3)}(\widetilde{\rho}) &= n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{\epsilon}}_i(\widetilde{\rho})^3 \\ \widehat{\mu}^{(4)}(\widetilde{\rho}) &= n^{-1}\sum_{i=1}^{n}\widehat{\boldsymbol{\epsilon}}_i(\widetilde{\rho})^4 \end{aligned}$$

When `heteroskedastic` is specified, the r, s element of $\boldsymbol{\Psi}^{\rho\rho}$ is estimated by

$$\widehat{\boldsymbol{\Psi}}^{\rho\rho}_{r,s}(\widetilde{\rho}) = (2n)^{-1}\mathrm{tr}\left\{(\mathbf{A}_r + \mathbf{A}_r')\widehat{\boldsymbol{\Sigma}}(\widetilde{\rho})(\mathbf{A}_s + \mathbf{A}_s')\widehat{\boldsymbol{\Sigma}}(\widetilde{\rho})\right\} + n^{-1}\widehat{\mathbf{a}}_r(\widetilde{\rho})'\widehat{\boldsymbol{\Sigma}}(\widetilde{\rho})\widehat{\mathbf{a}}_s(\widetilde{\rho}) \tag{4}$$

where $\widehat{\boldsymbol{\Sigma}}(\widetilde{\rho})$ is a diagonal matrix whose ith diagonal element is $\widehat{\boldsymbol{\epsilon}}_i^2(\widetilde{\rho})$, and $\widehat{\boldsymbol{\epsilon}}(\widetilde{\rho})$ and $\widehat{\mathbf{a}}_r(\widetilde{\rho})$ are as defined above. The last two terms in (3) do not appear in (4) because the $\mathbf{A}_s$ matrices used in the heteroskedastic case have diagonal elements equal to 0.

Having computed the estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\delta}}', \widehat{\rho})$ in steps 1a, 1b, 2a, and 2b, we next compute a consistent estimator for its asymptotic VC matrix, say, $\boldsymbol{\Omega}$. The estimator is given by $n\widehat{\boldsymbol{\Omega}}$ where

$$\begin{aligned}
\widehat{\boldsymbol{\Omega}} &= \begin{pmatrix} \widehat{\boldsymbol{\Omega}}^{\delta\delta} & \widehat{\boldsymbol{\Omega}}^{\delta\rho} \\ \widehat{\boldsymbol{\Omega}}^{\delta\rho\prime} & \widehat{\boldsymbol{\Omega}}^{\rho\rho} \end{pmatrix} \\
\widehat{\boldsymbol{\Omega}}^{\delta\delta} &= \widehat{\mathbf{P}}(\widehat{\rho})'\widehat{\boldsymbol{\Psi}}^{\delta\delta}(\widehat{\rho})\widehat{\mathbf{P}}(\widehat{\rho}) \\
\widehat{\boldsymbol{\Omega}}^{\delta\rho} &= \widehat{\mathbf{P}}(\widehat{\rho})'\widehat{\boldsymbol{\Psi}}^{\delta\rho}(\widehat{\rho}) \left\{\widehat{\boldsymbol{\Psi}}^{\rho\rho}(\widehat{\rho})\right\}^{-1} \widehat{\mathbf{J}} \left[\widehat{\mathbf{J}}' \left\{\widehat{\boldsymbol{\Psi}}^{\rho\rho}(\widehat{\rho})\right\}^{-1} \widehat{\mathbf{J}}\right]^{-1} \\
\widehat{\boldsymbol{\Omega}}^{\rho\rho} &= \left[\widehat{\mathbf{J}}' \left\{\widehat{\boldsymbol{\Psi}}^{\rho\rho}(\widehat{\rho})\right\}^{-1} \widehat{\mathbf{J}}\right]^{-1} \\
\widehat{\mathbf{J}} &= \widehat{\boldsymbol{\Gamma}} \begin{pmatrix} 1 \\ 2\widehat{\rho} \end{pmatrix}
\end{aligned}$$

In the above, $\widehat{\boldsymbol{\Psi}}^{\rho\rho}(\widehat{\rho})$ and $\widehat{\mathbf{P}}(\widehat{\rho})$ are as defined in (3) and (4) with $\widetilde{\rho}$ replaced by $\widehat{\rho}$. The estimators $\widehat{\boldsymbol{\Psi}}^{\delta\delta}(\widehat{\rho})$ and $\widehat{\boldsymbol{\Psi}}^{\delta\rho}(\widehat{\rho})$ are defined as follows:

When `homoskedastic` is specified,

$$\begin{aligned}
\widehat{\boldsymbol{\Psi}}^{\delta\delta}(\widehat{\rho}) &= \widehat{\sigma}^2(\widehat{\rho})\widehat{\mathbf{Q}}_{HH} \\
\widehat{\boldsymbol{\Psi}}^{\delta\rho}(\widehat{\rho}) &= \widehat{\sigma}^2(\widehat{\rho})n^{-1}\mathbf{H}'\left\{\mathbf{a}_1(\widehat{\rho}), \mathbf{a}_2(\widehat{\rho})\right\} + \mu^{(3)}(\widehat{\rho})n^{-1}\mathbf{H}'\left\{\mathrm{vec}_D(\mathbf{A}_1), \mathrm{vec}_D(\mathbf{A}_2)\right\}
\end{aligned}$$

When `heteroskedastic` is specified,

$$\begin{aligned}
\widehat{\boldsymbol{\Psi}}^{\delta\delta}(\widehat{\rho}) &= n^{-1}\mathbf{H}'\widehat{\boldsymbol{\Sigma}}(\widehat{\rho})\mathbf{H} \\
\widehat{\boldsymbol{\Psi}}^{\delta\rho}(\widehat{\rho}) &= n^{-1}\mathbf{H}'\widehat{\boldsymbol{\Sigma}}(\widehat{\rho})\left\{\mathbf{a}_1(\widehat{\rho}), \mathbf{a}_2(\widehat{\rho})\right\}
\end{aligned}$$

We note that the expression for $\widehat{\boldsymbol{\Omega}}^{\rho\rho}$ has the simple form given above because the estimator in step 2b is the efficient GMM estimator.

6.2 SAR model without spatially correlated errors

Consider the case $\rho = 0$, that is, the case where the disturbances are not spatially correlated. In this case, only step 1a is necessary, and `spivreg` estimates $\boldsymbol{\delta}$ by 2SLS using as instruments $\mathbf{H}$ the linearly independent columns of $\{\mathbf{X}_f, \mathbf{W}\mathbf{X}_f, \ldots, \mathbf{W}^q\mathbf{X}_f\}$. The 2SLS estimator is given by

$$\widetilde{\boldsymbol{\delta}} = \left(\widetilde{\mathbf{Z}}'\mathbf{Z}\right)^{-1}\widetilde{\mathbf{Z}}'\mathbf{y}$$

where $\widetilde{\mathbf{Z}} = \mathbf{P}_{\mathbf{H}}\mathbf{Z}$.

When `homoskedastic` is specified, the asymptotic VC matrix of $\widetilde{\boldsymbol{\delta}}$ can be estimated consistently by

$$\widetilde{\sigma}^2 \left(\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\right)^{-1}$$

where $\widetilde{\sigma}^2 = n^{-1}\sum_{i=1}^{n} \widetilde{\mathbf{u}}_i^2$ and $\widetilde{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\widetilde{\boldsymbol{\delta}}$ denotes the 2SLS residuals.

When `heteroskedastic` is specified, the asymptotic VC matrix of $\widetilde{\boldsymbol{\delta}}$ can be estimated consistently by the sandwich form

$$\left(\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\right)^{-1} \widetilde{\mathbf{Z}}'\widetilde{\boldsymbol{\Sigma}}\widetilde{\mathbf{Z}} \left(\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}}\right)^{-1}$$

where $\widetilde{\boldsymbol{\Sigma}}$ is the diagonal matrix whose ith element is $\widetilde{\mathbf{u}}_i^2$.

6.3 Spatially correlated errors without a SAR term

Consider the case $\lambda = 0$, that is, the case where there is no spatially lagged dependent variable in the model. In this case, we use the same formulas as in section 6.1 after redefining $\mathbf{Z} = \mathbf{Y}, \mathbf{X}$, $\boldsymbol{\delta} = (\boldsymbol{\pi}', \boldsymbol{\beta}')'$, and we take $\mathbf{H}$ to be composed of linearly independent columns of $(\mathbf{X}_f, \mathbf{M}\mathbf{X}_f)$.

6.4 No SAR term or spatially correlated errors

When the model does not contain a SAR term or spatially correlated errors, the 2SLS estimator provides consistent estimates, and we obtain our results by using `ivregress` (see [R] **ivregress**). When `homoskedastic` is specified, the conventional estimator of the asymptotic VC is used. When `heteroskedastic` is specified, the `vce(robust)` estimator of the asymptotic VC is used. When no endogenous variables are specified, we obtain our results by using `regress` (see [R] **regress**).

7 Conclusion

We have described the `spivreg` command for estimating the parameters of a SARAR model with additional endogenous RHS variables. In the future, we plan to add options for optimal predictors corresponding to different information sets.

8 Acknowledgment

We gratefully acknowledge financial support from the National Institutes of Health through the SBIR grants R43 AG027622 and R44 AG027622.

9 References

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

———. 2010. Thirty years of spatial econometrics. *Papers in Regional Science* 89: 3–25.

Anselin, L., and R. J. G. M. Florax. 1995. Small sample properties of tests for spatial dependence in regression models: Some further results. In *New Directions in Spatial Econometrics*, ed. L. Anselin and R. J. G. M. Florax, 21–74. Berlin: Springer.

Arraiz, I., D. M. Drukker, H. H. Kelejian, and I. R. Prucha. 2010. A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. *Journal of Regional Science* 50: 592–614.

Cliff, A. D., and J. K. Ord. 1973. *Spatial Autocorrelation*. London: Pion.

———. 1981. *Spatial Processes: Models and Applications*. London: Pion.

Cressie, N. A. C. 1993. *Statistics for Spatial Data*. Revised ed. New York: Wiley.

Drukker, D. M., P. Egger, and I. R. Prucha. 2013. On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews* 32: 686–733.

Drukker, D. M., H. Peng, I. R. Prucha, and R. Raciborski. 2013. Creating and managing spatial-weighting matrices with the spmat command. *Stata Journal* 13: 242–286.

Drukker, D. M., I. R. Prucha, and R. Raciborski. 2013. Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. *Stata Journal* 13: 221–241.

Haining, R. 2003. *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

Kelejian, H. H., and I. R. Prucha. 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17: 99–121.

———. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40: 509–533.

———. 2004. Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics* 118: 27–50.

———. 2007. The relative efficiencies of various predictors in spatial econometric models containing spatial lags. *Regional Science and Urban Economics* 37: 363–374.

———. 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157: 53–67.

Levitt, S. D. 1997. Using electoral cycles in police hiring to estimate the effect of police on crime. *American Economic Review* 87: 270–290.

Powers, E. L., and J. K. Wilson. 2004. Access denied: The relationship between alcohol prohibition and driving under the influence. *Sociological Inquiry* 74: 318–337.

Whittle, P. 1954. On stationary processes in the plane. *Biometrika* 41: 434–449.

About the authors

David Drukker is the director of econometrics at StataCorp.

Ingmar Prucha is a professor of economics at the University of Maryland.

Rafal Raciborski is an econometrician at StataCorp.