Introduction to spatial data and SAR models

Description

The Sp commands manage data and fit regressions accounting for spatial relationships. Sp fits SAR models that include spatial lags of dependent and independent variables with spatial autoregressive errors on lattice and areal data, which includes nongeographic data such as social network nodes.

Different fields use different jargon for spatial concepts. SAR stands for (take your pick) spatial autoregressive or simultaneous autoregressive.

Eight short introductions will turn you into an expert on the Sp software. In these introductions, you will learn about spatial weighting matrices and how to create them as you prepare your data for analysis. You will learn about three estimation commands—spregress, spivregress, and spxtrregress—for fitting SAR models. You will also find a worked example that includes data preparation, model fitting, and interpretation. Read the introductions first and read them sequentially.

The introductions and the commands of interest with spatial data are listed below, and each command is described in detail in its respective manual entry.

Learning the system

[Intro 1] A brief introduction to SAR models
[Intro 2] The W matrix
[Intro 3] Preparing data for analysis
[Intro 4] Preparing data: Data with shapefiles
[Intro 5] Preparing data: Data containing locations (no shapefiles)
[Intro 6] Preparing data: Data without shapefiles or locations
[Intro 7] Example from start to finish
[Intro 8] The Sp estimation commands

Preparing data

[zipfile] Compress and uncompress files in zip archive format
[spshape2dta] Translate shapefile to Stata format
[spset] Declare data to be Sp spatial data
[spbalance] Make panel data strongly balanced
[spcompress] Compress Stata-format shapefile

Looking at data

[grmap] Graph choropleth maps
[spdistance] Calculator for distance between places

Setting the spatial weighting matrix

[spmatrix] Create, manipulate, and import/export weighting matrices
[spgenerate] Generate spatial lag (W x x) variables
Fitting models

- **[SP] spregress**: Fit cross-sectional SAR models
- **[SP] spivregress**: Fit cross-sectional SAR models with endogenous covariates
- **[SP] spxtregress**: Fit panel-data SAR models

Postestimation

- **[SP] estat moran**: Moran’s test after regress
- **[SP] spregress postestimation**: Postestimation tools for spregress
- **[SP] spivregress postestimation**: Postestimation tools for spivregress
- **[SP] spxtregress postestimation**: Postestimation tools for spxtregress

Glossary

- **[SP] Glossary**: Jargon

Remarks and examples

The sections below provide more information about SAR models.

**References for learning SAR models**

Spatial models have been applied in a variety of disciplines, such as criminology, demography, economics, epidemiology, political science, and public health. Cressie (1993), Darmofal (2015), LeSage and Pace (2009), and Waller and Gotway (2004) provide textbook introductions.

Darmofal (2015, chap. 2) gives an introduction to spatial weighting matrices.

LeSage and Pace (2009, sec. 2.7) define total, direct, and indirect impacts.

Anselin (1988) gives a classic introduction to the subject.

**Technical references on the development and fitting of SAR models**

SAR models date back to the work of Whittle (1954) and Cliff and Ord (1973, 1981).


The formulas for the GS2SLS without higher-order spatial weighting matrices were published in Drukker, Prucha, and Raciborski (2013a). For the higher-order models, *spregress, gs2sls* implements the estimator derived in Badinger and Egger (2011) and Prucha, Drukker, and Egger (2016).

The properties of the ML estimator were proven by Lee (2004), who also provides the formulas for the robust estimator of the VCE.


Lee and Yu (2011) give formulas and theory for SAR panel models.
Acknowledgments

We thank Ingmar Prucha of the University of Maryland for his work with us on spatial methods and econometrics that led to the methods implemented here.

We also thank Irani Arraiz of the Inter-American Development Bank, Badi Baltagi of Syracuse University, Petter Egger of ETH Zurich, and Harry Kelejian of the University of Maryland for their helpful comments and guidance.

We are grateful to Maurizio Pisati of the Università degli Studi di Milano-Bicocca for allowing us to include grmap, a lightly adapted version of his spmap command (Pisati 2007), which was preceded by his tmap command (Pisati 2004).

We thank Stata users for their contributions on spatial data management and spatial analysis that were published in the Stata Journal. We thank Belotti, Hughes, and Piano Mortari for “Spatial panel-data models using Stata”. We thank Brophy, Daniels, and Musundwa for “gpsbound: A command for importing and verifying geographical information from a user-provided shapefile”. We thank Neumayer and Plumper for “Making spatial analysis operational: Commands for generating spatial-effect variables in monadic and dyadic data”. We thank Müller for “Stata in space: Econometric analysis of spatially explicit raster data”.

StataCorp’s Sp commands are based on earlier versions published in Drukker, Prucha, and Raciborski (2013a, 2013b) and Drukker et al. (2013).

References


