Introduction to spatial data and SAR models

Contents

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

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

Looking at data
- [SP] grmap Graph choropleth maps
- [SP] spdistance Calculator for distance between places

Setting the spatial weighting matrix
- [SP] spmatrix Create, manipulate, and import/export weighting matrices
- [SP] spgenerate Generate spatial lag (W × x) variables

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

Postestimation
- [SP] estat moran 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
**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 non-geographic 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. Read them first and read them sequentially.

**Remarks and examples**

Sp provides three estimation commands: `spregress`, `spivregress`, and `spxtregress`. They are extensions of Stata’s `regress`, `ivregress`, and `xtreg` commands.

Before you can use the Sp commands, you must construct the spatial weighting matrix. Usually, you will create the matrix based on shapefiles (maps) that you have obtained from the web or other sources.

That is the subject of the introduction, starting with [SP] intro 1.

The references 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 GS2SLS estimator was derived by Kelejian and Prucha (1998, 1999, 2010) and extended by Arraiz et al. (2010) and Drukker, Egger, and Prucha (2013a).

The formulas for the GS2SLS without higher-order spatial weighting matrices were published in Drukker, Prucha, and Raciborski (2013c). 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 Plümper 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 (2013d, 2013c) and Drukker, Peng, Prucha, and Raciborski (2013b).

References


