Intro — Introduction to spatial data and SAR models

Description Remarks and examples Acknowledgments References

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 `spxtregress`—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

[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 ($\mathbf{W} \times \mathbf{x}$) 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

Remarks and examples

The sections below provide more information about SAR models.

References for learning SAR models
Technical references on the development and fitting of 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 (2013).

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 Drukker, Egger, and Prucha (2023).

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

Kelejian and Prucha (2010) give a technical discussion of how normalizing spatial weighting matrices affects parameter definition.

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, Peter 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 (2013a, 2013b) and Drukker et al. (2013).

References

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. New York: Springer. https://doi.org/10.1007/978-94-015-7799-1.

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. https://doi.org/10.1111/j.1467-9787.2009.00618.x.

Badinger, H., and P. H. Egger. 2011. Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances. *Papers in Regional Science* 90: 213–235. https://doi.org/10.1111/j.1435-5957.2010.00323.x.

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

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

Cressie, N. 1993. *Statistics for Spatial Data*. Rev. ed. New York: Wiley. https://doi.org/10.1002/9781119115151.

Darmofal, D. 2015. *Spatial Analysis for the Social Sciences*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139051293.

Drukker, D. M., P. H. 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. https://doi.org/10.1080/07474938.2013.741020.

———. 2023. Simultaneous equations models with higher-order spatial or social network interactions. *Econometric Theory* 39: 1154–1201. https://doi.org/10.1017/S026646662200007X.

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. 2013a. Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. *Stata Journal* 13: 221–241.

———. 2013b. A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. *Stata Journal* 13: 287–301.

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. https://doi.org/10.1023/A:1007707430416.

———. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40: 509–533. https://doi.org/10.1111/1468-2354.00027.

———. 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157: 53–67. https://doi.org/10.1016/j.jeconom.2009.10.025.

Lee, L.-F. 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72: 1899–1925. https://doi.org/10.1111/j.1468-0262.2004.00558.x.

Lee, L.-F., and J. Yu. 2011. Estimation of spatial panels. *Foundations and Trends in Econometrics* 4(1–2): 1–164. https://doi.org/10.1561/0800000015.

LeSage, J., and R. K. Pace. 2009. *Introduction to Spatial Econometrics.* Boca Raton, FL: Chapman and Hall/CRC. https://doi.org/10.1201/9781420064254.

Pisati, M. 2004. Simple thematic mapping. *Stata Journal* 4: 361–378.

———. 2007. spmap: Stata module to visualize spatial data. Statistical Software Components S456812, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s456812.html.

Spinelli, D. 2022. Fitting spatial autoregressive logit and probit models using Stata: The spatbinary command. *Stata Journal* 22: 293–318.

Waller, L. A., and C. A. Gotway. 2004. *Applied Spatial Statistics for Public Health Data.* Hoboken, NJ: Wiley. https://doi.org/10.1002/0471662682.

Whittle, P. 1954. On stationary processes in the plane. *Biometrika* 434–449. https://doi.org/10.2307/2332724.

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.

For suggested citations, see the FAQ on citing Stata documentation.