Title

stata.com

intro — Introduction to spatial data and SAR models

Contents

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. Read them first and read them sequentially.

Remarks and examples

stata.com

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.

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, 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

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. New York: 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.

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.

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.

Darmofal, D. 2015. *Spatial Analysis for the Social Sciences*. New York: Cambridge University Press.

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

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

——. 2013d. 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.

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

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

Lee, L.-f. 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72: 1899–1925.

Lee, L.-f., and J. Yu. 2011. Estimation of spatial panels. *Foundations and Trends in Econometrics* 4(1–2): 1–164.

LeSage, J., and R. K. Pace. 2009. *Introduction to Spatial Econometrics*. Boca Raton, FL: Chapman & Hall/CRC.

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.

Prucha, I. R., D. M. Drukker, and P. H. Egger. 2016. Simultaneous equations models with higher-order spatial or social network interactions. Working paper, Department of Economics, University of Maryland. http://econweb.umd.edu/~prucha/papers/WP_IRP_PHE_DMD_2016.pdf.

Waller, L. A., and C. A. Gotway. 2004. *Applied Spatial Statistics for Public Health Data*. Hoboken, NJ: Wiley.

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