# SPATIAL REGRESSION ANALYSIS

## *DAY 1*

## INTRODUCTION

### *Welcome and Organization*

### *Introduction and Motivation*

- Some examples
- Course outline
- Course project

*Readings*

Anselin L. (1999). The future of spatial analysis in the social sciences. *Geographic Information Sciences* 5, 67-76.

Goodchild M., Anselin L., Appelbaum R., Harthorn B. (2000). Toward spatially integrated social science. *International Regional Science Review* 23, 139-159.

### *Spatial Effects*

- Spatial regression analysis
- Spatial effects

*Readings*

Anselin-Bera, Sections I and II.

## Lab Assignment: Model Setup

The assumption is that you are already somewhat familiar with the *SpaceStat* software and its *ArcView* extensions. These were covered at length in the companion introductory workshop. If you are not familiar with the software, you should work through the relevant tutorials and exercises in the Workbook. These will walk you through the various tasks described below.

The first assignment consists of two main tasks: (1) setting up the various pieces necessary for the forthcoming spatial regression analysis; and (2) getting to know the data. At the end of the day, you should have selected a data set (or use your own), have it stored in *SpaceStat* binary format (Gauss data set) and have constructed a number of different spatial weights. For this, you will need a digital base map (*ArcView* shape file) and the centroid coordinates for the observations (if your data are stored as points, you will not be able to construct contiguity weights, unless you first create Thiessen polygons -- see the Workbook for an example). The spatial weights should include simple contiguity (using rook and queen criterion), higher order contiguity, distance based weights (distance band), and k-nearest neighbors, as well as any others that may be appropriate for your analysis.

Your data set should contain a dependent variable and at least two explanatory variables. You should also carry out a simple exploratory analysis: descriptive statistics, identification of outliers, assessment of normality, simple correlation between the variables, spatial autocorrelation of the variables (Moran's I, LISA). If you are comfortable with *ArcView*, you can use the *SpaceStat-ArcView* extension to visualize the patterns in the data (quartile map, outliers, LISA map).

You should be able to briefly summarize the results of the data exploration in terms of association between the variables, spatial association of each variable (sensitivity of spatial association to the choice of weights), and possible overlapping (or non-overlapping) local spatial clusters in the variables.

*DAY 2*

# MODEL SPECIFICATION

## *Spatial Heterogeneity*

- Issues
- Discrete heterogeneity
- Continuous heterogeneity

*Readings:*

Anselin, Luc. 1990. Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis. *Journal of Regional Science* 30:185-207.

Casetti, Emilio. 1997. The Expansion Method, Mathematical Modeling, and Spatial Econometrics. *International Regional Science Review* 20:9-33.

Fotheringham, A. Stewart, Chris Brundson and Charlton Martin. 1998. Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis. *Environment and Planning A* 30:1905-1927.

## *Spatial Dependence*

- Specifying spatial covariance
- Spatial lag models
- Spatial error models
- Direct representation

*Readings*

Anselin and Bera (1998), pp. 246-252.

Anselin (1988), pp. 34-36.

Cressie (1993), pp. 410-423.

SpaceStat Tutorial, Chapters 31-34.

## Lab Assignment: Spatial Heterogeneity

This assignment involves setting up and trying various model specifications to analyze and interpret the specification of spatial heterogeneity for your model and your data. You will need data on a dependent variable as well as a small number of explanatory variables (presumably representing a more or less meaningful model). The analysis should include:

- selection of meaningful categories for spatial regimes
- spatial anova on your dependent variable
- testing for heteroskedasticity in a base model using the regime categories
- estimation of spatial regimes for your base regression model
- test for spatial homogeneity and assess the extent to which heteroskedasticity has been taken care of
- estimate spatial expansion model and assess the extent to which heteroskedasticity has been taken care of

The starting point is a simple linear regression model and the diagnostics for heteroskedasticity provided by the OLS estimation routine. You also will need to use the various specialized models provided by *SpaceStat.*

At the end of the day, you should be able to defend a choice of a particular specification for spatial heterogeneity. You should make sure you have a good understanding of the differences between these specifications and motivate your choice in function of the results of specification tests.

*DAY 3*

# SPECIFICATION TESTS

## *Specification Tests*

- Spatial autocorrelation tests
- Tests against spatial error
- Tests against spatial lag
- Tests against higher order alternatives
- Specification robust tests

*Readings:*


Anselin and Bera (1998), pp. 264-281.

Anselin (1988), pp. 65-73, 100-105.

Anselin, L. (2001). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference* 97, 113-139.

Anselin, L. and H. Kelejian (1997). Testing for spatial error autocorrelation in the presence of endogenous regressors. *International Regional Science Review* 20, 153-182.

Anselin, L and R.  Florax (1995). Small sample properties of tests for spatial dependence in regression models: some further results. In *New Directions in Spatial Econometrics* pp. 21-74.

Kelejian, H. and D. Robinson (1998). A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results. *Regional Science and Urban Economics* 28, 389-417.

## Lab Assignment: Testing for Spatial Dependence

This assignment involves testing various model specifications to assess the extent and type of spatial dependence in the base line models. The starting point is a simple linear regression model and the diagnostics provided by the OLS estimation routine. From this, you can assess the types of spatial effects that may be present. Check out variations of your base model that include spatial regimes and groupwise heteroskedasticity. Compare the indications of the various tests and for various spatial weights. If possible/appropriate, assess whether the indication of spatial effects changes when new variables are introduced or variables are dropped from the base model. You have to complete the following tasks:

- construct three different spatial weights, for example, a contiguity based one, a distance based one and a k-nearest neighbor one
- construct second and third order from the first order spatial contiguity weights
- run all the tests for spatial dependence based on the OLS residuals for all five weights
- interpret the tests and suggest the most likely alternative
- compare results between the different weights
- assess the extent to which the tests remain significant (or not) when you introduce spatial regimes
- if the diagnostics for heteroskedasticity are significant, run a heteroskedastic model and test again

At the end of the day, you should be able to defend a choice of a particular specification, spatial lag or spatial error and/or regimes/heteroskedasticity. You should make sure you have a good understanding of the differences between these specifications and motivate your choice in function of the results of specification tests.

## *DAY 4*

# ESTIMATION

## *Maximum Likelihood Estimation*

- General principles of ML estimation
- ML estimation of spatial lag model
- ML estimation of spatial error model
- ML estimation of spatial dependence and heteroskedasticity

*Readings*

Anselin and Bera (1998), pp. 255-258.

Anselin (1988), pp. 57-65.

Ord, J.K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120-126.

*IV-GMM Estimation of Spatial Models*

- Spatial two stage least squares
- GM estimation spatial error model
- GMM estimation spatial error model
- Estimation of higher order models

*Readings*

Kelejian, H. and I. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.

Conley, T. (1999). GMM estimation with cross-sectional dependence. *Journal of Econometrics* 92, 1-45.

Kelejian, H. and I. 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.

Bell, K. and N. Bockstael (2000). Applying the generalized moments estimation approach to spatial problems involving microlevel data. *The Review of Economics and Statistics* 82, 72-82.

Greene, W.H. (1997). *Econometric Analysis*, pp. 517-531.

Davidson R. and J.G. MacKinnon (1993). *Estimation and inference in econometrics*, Chapter 17.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1054.

Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817-858.

## Lab Assignment: Estimating Spatial Models

This assignment involves estimating various model specifications that incorporate spatial dependence in the form of lag or error dependence, possibly in combination with spatial regimes and/or groupwise heteroskedasticity. Use maximum likelihood estimation and assess the extent to which spatial dependence has been accounted for by the model (i.e., test for remaining spatial association). If using spatial regimes, test for parameter constancy across regimes. Compare the model fit between the lag and error specification.

You will also be re-estimating the various models by means of IV and GM techniques and comparing the results to maximum likelihood estimation.

At the end of the day, you should be pretty close to your final model choice and have a good idea of how the spatial effects have been accounted for through the incorporation of spatial dependence and/or spatial heterogeneity.

*Note*: in order to carry out maximum likelihood estimation, you will need to convert your spatial weighs from a sparse format to a “full” (fmt) format.

*DAY 5*

# ADVANCED TOPICS

## *Space-Time Dependence*

- Panel data
- Spatial panel data models
- Fixed effects
- Spatial Seemingly Unrelated Regression (SUR)

*Readings*

Anselin (1988), Chapter 12.

Anselin (2000), Spatial econometrics, Section 3.2.

Elhorst, J. Paul (2001). Dynamic Models in Space and Time. *Geographical Analysis* 33:119-140.

Pace, R.K., R. Barry, J. Clapp and M. Rodriguez (1998). Spatiotemporal Autoregressive Models of Neighborhood Effects. *Journal of Real Estate Finance and Economics* 17, 15-33.

Anselin, L. (1988). A Test for Spatial Autocorrelation in Seemingly Unrelated Regressions. *Economics Letters* 28, 335-341.

Baltagi, B., S.H. Song and W. Koh (2000). Testing Panel Data Regression Models with Spatial Error Autocorrelation. Working Paper, Dept. of Economics, Texas A&M University.

*Spatial Probit Models*

- Qualitative response models
- Spatial probit specification
- Testing for spatial effects in probit
- Estimating spatial probit

*Readings*


Anselin (2000), Spatial econometrics, Section 3.3.

Pinkse, J. and M. Slade (1998). Contracting in space: an application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85, 125-154.

Beron, K. and W. Vijverberg (2002). Probit in a spatial context: a Monte Carlo approach. In Anselin and Florax, *Advances in Spatial Econometrics* (forthcoming).

Kelejian, H. and I. Prucha (1999). On the Asymptotic Distribution of the Moran I Test Statistic with Applications. Working Paper, Department of Economics, University of Maryland.

Pinkse, J. (1998). Asymptotic Properties of the Moran and Related Tests and a Test for Spatial Correlation in Probit Models. Working Paper, Department of Economics, University of British Columbia.

Pinkse, J. (2002). Moran-Flavored Tests with Nuisance Parameters, Examples. In Anselin and Florax, *Advances in Spatial Econometrics* (forthcoming).

Fleming, M. (2002). A Review of the Techniques for Estimating Spatially Dependent Discrete Choice Models. In Anselin and Florax, *Advances in Spatial Econometrics* (forthcoming).

## Lab Assignment: Putting it all Together

At this point, you should be ready to summarize your findings and defend and interpret the final model specification both in technical as well as in methodological terms.

# PROJECT PRESENTATIONS