University of Chicago
Department of Sociology
Spring 2017

SOCI 40217/GEOG 40217, MACS 55000

Spatial Regression Analysis

Luc Anselin

Meet:	Mon, Wed 1:30-2:50pm Saieh Hall for Economics 203
Office:	421 Social Science Research Building
E-Mail:	anselin@uchicago.edu
Office Hours:	Mon, Wed 3:00-4:00pm and by appointment
Prerequisite:	graduate level econometrics or multivariate regression, matrix algebra

SYLLABUS

Course Description

This course covers statistical and econometric methods specifically geared to deal with the problems of spatial dependence and spatial heterogeneity in cross-sectional and panel (space-time) data. The main objective of the course is to gain insight into the scope of spatial regression methods, to be able to apply them in an empirical setting, and to properly interpret the results of spatial regression analysis.

While the focus is on *spatial* aspects, the types of methods covered have general validity in statistical practice. The course covers the specification of spatial regression models in order to incorporate spatial dependence and spatial heterogeneity, as well as different estimation methods and specification tests to detect the presence of spatial autocorrelation and spatial heterogeneity. Special attention is paid to the application to spatial models of generic statistical paradigms, such as Maximum Likelihood and Generalized Method of Moments. An important aspect of the course is an emphasis on computation, with the application of open source software tools such as R, GeoDa/GeoDaSpace and PySAL to solve empirical problems, and the use of R or Python to design simulation exercises.

Objectives

1. Learn principles of spatial regression analysis (aka spatial econometrics) and its application to social science research questions
2. Learn how to incorporate spatial effects such as spatial dependence and spatial heterogeneity in the specification of a regression model

3. Gain an appreciation for the assumptions and limitations associated with different statistical paradigms (ML, GMM) and how they pertain to the estimation of spatial regression models
4. Learn how to interpret and present the results of a spatial regression analysis in a coherent fashion
5. Learn how to use appropriate open source software tools to carry out spatial regression estimation and simulation

Organization

The class will meet twice a week in a mixed lecture/lab format. There is no lab per se, but there will be a heavy emphasis on computation and the use of open source software for the course project. This software will be demonstrated in class, supplemented by materials on the course Canvas site. In addition to the regular class times, there is an optional lab section in the CSIL computer lab, M-W, 3-4pm. This will provide additional opportunities to apply the software and address computational aspects of spatial regression analysis. All the software referred to in class is installed in the CSIL lab.

You are strongly encouraged to use your own laptop and install the software yourself (everything used in the course is cross-platform, free and open source, see instructions below).

The course will use the Canvas site as the main communication mechanism. All required readings, software guides and data will be available from the course site. Note that the course number used is SOCI 40217.

All assignments, papers etc. must be submitted as a **pdf digital file**: NO PAPER and no Word docs, no exceptions.

Requirements

There are two main goals for the course. The main one is for you to complete an empirical/computational exercise that carries out an in-depth spatial regression analysis of a research problem of your choice. There are a few options to accomplish this. One is an empirical examination that applies the techniques covered in class using your own data, or one of the data sets available from the Center for Spatial Data Science web site. A second option is an extended simulation experiment, where the properties of test statistics and estimation methods are assessed using artificially generated data and model specifications. Finally, for those of you more interested in computation, a third option is to develop a collection of estimation routines that would become part of the spreg Python spatial econometrics library. Here, the objective would be to create a professional grade set of Python classes and functions that accomplish estimation of a new set of spatial regression specifications. Examples include spatial panel data models, model with discrete dependent variables, and multilevel models that include spatial dependence.

The second goal is to gain an appreciation of how spatial regression models are (or are not) applied in current social science research. To this end, you will have to write a brief review of a recently published article in your field (after 2010) that uses spatial data and/or addresses a spatial question. Further specifics will be provided in class (tentative deadline is May 1).

Over the course of the semester, there will be three intermediate project deadlines, with deliverables that are part of the grading of the final paper (deadlines are tentative, subject to adjustment):

- A fully spelled-out research question (including a brief description of the data and variables to be used as well as a tentative model specification, the general outline of the design of the simulation experiment, or an initial organization of a coding project) – 2pp. due April 12
- Either: (1) An initial analysis of a non-spatial specification (by means of OLS or 2SLS as appropriate) with diagnostics for spatial effects and an interpretation of the results; (2) The design of the simulation experiment, fully spelled out; or (3) a fully organized set of classes and functions for the coding project
  max 5pp. (not including graphs and figures), due May 1
- Either: (1) Estimation results for the spatial model(s) with an initial interpretation of the results; (2) Initial results of the simulation experiments with summary and interpretations; (3) Full set of code with working examples
  max 5pp., due May 24

The final paper is due June 2 (May 26 for those graduating in Spring), max 20pp (not including tables and graphs)

More details will be provided in class and on the course web site as the quarter progresses.

Software

The class uses only open source software (free and cross-platform). Everything can be readily downloaded from the web.

- **GeoDa** and **GeoDaSpace**, available from http://geodacenter.github.io/download.html
- **R** and its associated spatial regression analysis packages, especially `spdep`, `sphet` and `splm`, everything available from http://cran.r-project.org
- Recommended: **jupyter** notebooks, available from http://jupyter.readthedocs.io/en/latest/install.html (requires a Python installation, preferably through the Anaconda distribution – see instructions on the jupyter web site; make sure to install the IRkernel package in R, see

the jupyter web site for detailed instructions, starting here https://irkernel.github.io/requirements/).
- Alternatively, a graphical user interface to R: **RStudio**, available from https://www.rstudio.com/products/rstudio/download/

See also the separate installation note for more details. This software is installed in the CSIL lab, which is available for exclusive use by this class M-W, 3-4pm.

Readings

The course text is:

- Luc Anselin and Sergio Rey (2014). *Modern Spatial Econometrics in Practice, A Guide to GeoDa, GeoDaSpace and PySAL*. GeoDa Press, Chicago, IL. (available from Amazon).

A pdf of this text will be made available on the course web site. In addition, supplementary readings and other materials (e.g., software guides) will be on the course web site as well.

Grading

- Class participation: 10%
- Review paper: 30%
- Project Paper: 60%
  - research question 10%
  - specification tests/simulation design 10%
  - estimation results/simulation results 10%
  - final paper/presentation 30%

Tentative Course Outline (subject to change)

- Week 1: 3/27, 3/29
  - Introduction and overview of the course
  - Boot camp: review of matrix algebra, spatial weights, spatial autocorrelation, OLS and 2SLS basics
  - Introduction and overview of the software (GeoDa/GeoDaSpace, R, Python, numpy)
- Week 2: 4/3, 4/5
  - Spatial process models and spatial covariance
  - Simulating spatial processes
- Week 3: 4/10, 4/12
  - Specification of spatial regression models: lag, error, spatial Durbin, SLX, global and local effects, spatial multipliers, direct and indirect effects, interpretation of the coefficients, spatial heterogeneity

- Week 4: 4/17, 4/19
    - Estimation principles – Maximum Likelihood: ML for spatial lag and spatial error model, computational issues
- Week 5: 4/24, 4/26
    - Specification tests: classic tests in cross-sectional regression, Moran's I, LM tests, robust LM tests, LR tests, specification search
- Week 6: 5/1, 5/3
    - Estimation principles – GMM: IV/GMM for spatial lag and spatial error models, HAC estimator, effect of heteroskedasticity
- Week 7: 5/8 (no class 5/10)
    - Spatial panel models: pooled, fixed and random effects, different error specifications, spatial SUR
- Week 8: 5/15, 5/17
    - Spatial panel models (continued): ML and GMM estimation, factor models
    - Specification tests in spatial panel models
- Week 9: 5/22, 5/24
    - Spatial effects in discrete dependent variable models: spatial probit, spatial tobit, spatial effects in count models (hierarchical models)
    - Specification tests, ML and GMM estimation, simulation estimators
- Week 10: 5/31 (no class 5/29 – Memorial Day)
    - Project presentations or review (last day before reading days)