The use of geographic information systems (GIS) has been increasingly common in applied microeconometric studies in the past few years. The aim of this course is two-fold: understanding the benefits of using GIS and learning how to use ArcGIS 10, the most widely used GIS software.

Each of the seven 3-hour lectures in this course consists of two parts.

Part I introduces you to a couple of recent applied microeconometric papers that use GIS in detail so that you will learn the benefits of using GIS. In this part, we will also discuss applied microeconometric techniques useful for conducting research.

In Part II of each lecture, you will learn the practical knowledge on how to use ArcGIS 10 for replicating spatial data used by the papers that we discuss in Part I. In particular, we learn (1) which geoprocessing tools should be used for what kind of spatial data editing, and (2) how to write a Python script to implement these tools so that you don’t need to repeat the same data editing process manually and other researchers can replicate what you do.

Grading

Your grade will be only based on the term paper (the submission deadline: 9:00am on Monday 21st January, 2012; email a PDF copy to me at masa@iies.su.se). The term paper will be graded on how much you are aware of econometric issues relevant to your empirical research design. The paper does not need to use the GIS (what’s the point of specifying the data type to encourage you to do research?). My comments on your submitted paper will be returned to you some time during Quarter 3.

See the end of this syllabus for the guideline to write a term paper for this course.

Assignments

There will occasionally be assignments, but these are not compulsory but for familiarizing yourself to ArcGIS.

Using ArcGIS
You are free to use computers with ArcGIS installed in rooms U31 and U32 (they are accessible until around 7 pm on the weekdays). The door code to enter the room, the login name and the password to use the computer will be given during the first lecture.

In the directory T:\Economics, you will find datasets to be used in each lecture. At the beginning of each lecture, you should copy the whole folder for each lecture to your computer’s hard disk drive D:\temp. This folder will be deleted after you log out. If you want to save these files to continue working on them, create your own directory in S:\Economics and copy the files to this directory.

After this course, if you want to use ArcGIS for your dissertation, talk to your supervisor. For students from the Stockholm School of Economics, your supervisor is in charge of funding the software license. For students at Stockholm University, talk further to the Director of Graduate Studies after obtaining the support of your supervisor.

Background readings

For an overview of the use of GIS in economics, the following two papers are recommended.


Melissa Dell at MIT has written a lecture note on GIS in economics with lots of tips. We rely on this lecture note in some of the lectures.


Yale University Library’s website makes public tutorials for its GIS Workshops.

http://guides.library.yale.edu/content.php?pid=29977&sid=1244888

Below I briefly describe what each lesson focuses.

Lesson 1: Introduction

Part I: We learn how spatial datasets expand the set of feasible research questions for applied microeconometric research. The first three papers below use spatial data created from satellite sensor images while the latter three (as well as Michalopoulos and Papaioannou 2012) use spatial data created from scanned maps.


Part II: We learn how to read each type of the spatial datasets. For this purpose, we also learn the basics of GIS: data types and coordinate systems. In addition, we learn how to use the Model Builder to use geoprocessing tools.

**Geoprocessing tools to be learned:**

For file type conversion:

  - Make XY Event Layer, Copy Features, Ascii to Raster, Copy Raster

For data editing:

  - Raster Calculator

For projection:

  - Define Projection, Project

**Lesson 2: Spatial join**

Part I: GIS allows researchers to merge different datasets by location at quite disaggregated a level. This process is known as spatial join, and it allows you to control for fixed effects of very small areas so as to minimize unobservable heterogeneity within fixed effect units (e.g. zip-code areas for developed countries as in Currie and Neidell 2005; villages and town-districts for developing countries as in Kudamatsu et al. 2012). In some cases, spatial join allows you to construct an instrumental variable (as in Feyrer and Sacerdote 2009).

Kudamatsu, Masayuki, Torsten Persson, and David Strömberg. 2012.


Part II: We learn how to write a Python script for replicating the data used in Kudamatsu et al. 2012.

Geoprocessing tools to be learned:

For new data creation:

Create Fishnet, Feature To Point, Add XY Coordinates

For merging datasets:

Spatial Join

Lesson 3: Buffer

Part I: If we collect the data on the geographic coordinate of the unit of observations (households, schools, clinics, etc.) by GPS receivers or some other means, GIS makes it easy to identify each unit’s neighbourhood defined by the distance from it. This allows economists to create useful controls that avoid omitted variable bias (as in Conley and Udry 2010) or to estimate the spatial spill-over effect of policy-interventions (as in Miguel and Kremer 2004 and de Mel et al 2010).


Part II: We learn how to merge each unit of observation with those in its neighbourhood. We also learn about the UTM projection.

Geoprocessing tools to be learned:

For new data creation:

Buffer
Lesson 4: Distance as an instrument

Part I: Under certain circumstances, distance can be used as an instrumental variable. We will learn when distance can be a valid instrument and how ArcGIS makes the computation of distance easy.


Part II: We learn how to extract geographic coordinates to measure distances between those points on the earth such as the centroid of a country and the nearest point on the coastline. We also learn how to calculate the area of a polygon feature.

Geoprocessing tools to be learned:

For new data creation:

Select, Dissolve, Intersect, Feature to Line

For data editing:

Add Field, Calculate Field, Near

Lesson 5: Zonal statistics

Part I: GIS allows researchers to calculate any summary statistics for any levels of administrative areas (e.g. countries, provinces) and for any “squares” defined by parallels and meridians. The latter is useful to overcome the possible endogeneity of the boundary of administrative areas.


Part II: We learn how to calculate statistics for each spatial zone. We also learn how to create a loop in Python programming.

Geoprocessing tools to be learned:

For merging datasets:

Zonal Statistics as Table, Extract Values By Point
Lesson 6: Elevation as an instrument or an exogenous regressor

GIS makes easy the derivation of various elevation measures that can be used as an exogenous variable (conditional on various other geographic measures. This has so far been the most popular application of GIS in applied microeconometrics. We mainly discuss Duflo and Pande (2007) while other papers listed below are briefly mentioned.


Part II: We learn how to calculate slopes from the elevation data and how to handle raster datasets.

Geoprocessing tools to be learned:

For projection:

Project Raster

For new data creation:

Reclassify, Slope

For data editing:

Join Field

Lesson 7: Spatial regression discontinuity design

Part I: If a treatment or a government intervention targets a well-specified geographic area, we can use two-dimensional regression discontinuity design to obtain credible estimates on the impact of the treatment.


Part II: We learn Python scripts used by Dell (2010) as a review of what we have learned in the previous lessons.

Geoprocessing tools to be learned:

For new data creation:

   Extract By Mask, Clip (Analysis)

For data editing:

   Surface Length
Guideline for writing a term paper

1. Choosing a topic

Any topic is fine as long as it is an empirical question. For tips to find a good research topic, see the following pieces of writing by leading economists:

http://econ.lse.ac.uk/staff/spischke/phds/How%20to%20start.pdf
http://www.princeton.edu/~dixitak/home/dixitwrk.pdf (see pages 4-6 in particular)
http://people.ischool.berkeley.edu/~hal/Papers/how.pdf (see Sections 1 and 2 in particular)
http://faculty.haas.berkeley.edu/LEVINE/cheap_advice.html#dissertation

2. Originality

A literature survey is NOT allowed. Your term paper needs to be a piece of original research.

3. Proposal or full paper

A research proposal is fine. Maybe you haven’t obtained the datasets needed. In such cases, write a proposal. I do not want to force you to write a full paper at this stage. To write a full paper, you may need to opt for a research question that is more feasible but less important. I do not want that.

Of course, this doesn’t mean that you should not write a full paper. If you have already obtained some findings, you should write them up in the form of a full paper, which will be a draft chapter in your PhD thesis.

4. Structure of full paper / proposal

For a full paper, follow the structure of the papers that you read for the course (or any other paper published in top journals) as closely as possible. Of course, the structure differs depending on the nature of the paper. Pick the paper that’s closest to your paper in terms of empirical research design (which may be on a different topic than yours) and imitate its structure.

For a research proposal, follow the guideline below:

You should first clearly state the research question (the specific one, not the general one such as “Does credit market failure impede growth?”).

Then explain why this question is important and original. Here you may need to cite the existing literature, but never write a lengthy literature review. The purpose of citing the literature is to show why your research is original and will be an important contribution to the literature.

Finally, describe the empirical research method you use to answer this research question.

Describe the data you will use or collect (the sample period, the number of observations, the list of variables to be used, the sampling method employed if it’s a survey data). Write summary statistics tables or figures with the content empty. You can at least specify what variables to be reported in these tables and figures. Here you may also want to create a table in which you compare the means between treated and control observations with t-statistics.
reported. Then write down the equation to be estimated. Explain what estimation method you use (OLS, IV, SUR, Probit etc., including how standard errors will be computed). Describe the identifying assumption (the assumption that ensures the consistency of estimated coefficients on the variables of interest). Finally, write down regression tables. You don’t have results yet, but you can still write them down to describe what empirical specification each column estimates. What is the dependent variable? How is the sample restricted? Which regressors are included? What F-tests will be reported? And so forth. Add notes to the table, to explain how standard errors are computed, what F-statistics refers to, etc. Write paragraphs to explain these regression tables. This is where you need to think hard about the potential threat to your identifying assumption. To deal with each threat, what kind of robustness checks need to be done?

You don’t need to write implications of your finding or the conclusion, because you haven’t obtained any finding yet.