

A Short Course on Spatial Econometrics and GIS

Mark L. Burkey¹

¹ North Carolina A&T State University, Greensboro, NC, USA

Received: 20 August 2018/Accepted: 26 August 2018

Abstract. This resource gives a brief overview of a website and playlist of YouTube videos using open source software (R, GeoDa, and QGIS) designed to help get scholars up and running with analyzing their own data using Spatial Econometrics. Sample data, handouts, code, and map files are provided for ease of replication. The course covers the basics of integrating data into a spatial data set, contiguity and spatial correlation, doing basic spatial regressions in GeoDa, and doing more sophisticated specification tests and regressions in R.

JEL classification: R1, C21

Key words: spatial econometrics, instructional videos

1 Introduction

The internet has drastically improved the ability to provide quality distance-learning opportunities over the past decade. The explosion of MOOCs and high-quality educational videos hosted on YouTube have reduced the need for professionals to travel to a certain place in order to learn or refresh their skills in mathematics or data analysis. While there are still many opportunities for taking these courses at various conferences or universities, there are cheaper high quality substitutes becoming available. Specifically when it comes to Spatial Econometrics, there are many opportunities to travel for face-to-face instruction (Table 1), and the author of this article benefitted greatly from a short course offered by Luc Anselin at UI-UC in 2003. However, these courses can easily cost well over \$5,000 when the tuition, travel, and lodging are factored in. In this paper we introduce one complete, yet continuing series of free resources that are available.

2 Overview of the course

The core of the course is comprised of 12 YouTube videos, between 15 and 41 minutes long. These videos use screen capture technology (Burkey 2015), rather than videos of a professor in front of a typical classroom. All of the videos and materials are free of charge, posted at <http://spatial.burkeyacademy.com>. The software used is also all freely available¹, making the course very accessible to faculty and students alike. Each

¹One file is an Excel File, but is also works in some open source applications such as LibreOffice/Open Office.

Table 1: Examples of Short Courses related to Spatial Econometrics and Spatial Data Analysis

Location	Recent Offerings	Cost
Regional Research Institute, West Virginia University, USA	Summer, 1 Week, 2016-2019	\$3,000 + lodging, travel (2018)
Spatial Econometrics Advanced Institute, Università Cattolica del Sacro Cuore of Rome	Summer, 4 weeks, 2018	€ 2,300-3,300 + lodging, travel (2018)
ICPSR, University of Michigan, USA	Summer, 1 Week	\$1,700-3,200 + lodging, travel (2018)
NARSC Workshop, San Anto- nio, TX, USA	1 Day, before the start of the 2018 NARSC Conference	\$95 in addition to the cost of attending NARSC Conference

video has an accompanying handout, dataset, Excel file, or text file containing code, as is appropriate.

3 Brief outline of the course

Though additions to the series will continue to be made based on viewer feedback, the core 12 videos (plus an additional 9-minute brief “Welcome” video) form a complete, though basic introduction to Spatial Econometrics. It is assumed that the viewers are familiar with cross-sectional econometrics and basic matrix notation. Some basic familiarity with map files and R would be helpful, but not required.

3.1 Introduction to thinking spatially with some econometric models

The first video outlines some of the major Spatial Econometric Models and how to think about spatial interaction/spillover in basic terms. The video and handout also discuss some of the major researchers who developed the models, and give some references for books and journal articles where viewers can find more information. In this section we introduce the following models:

OLS Model

$$y = X\beta + \epsilon \quad (1)$$

The OLS model does not contain a spatial relationship, but is often used as a starting place. [Anselin \(1988\)](#) favored the Lagrange Multiplier approach to specification searches. Starting with OLS, Anselin derived 5 Lagrange Multiplier statistics that help determine if the (possibly misspecified) OLS model points toward the Spatial Lag, Spatial Error, or SARMA models, though he suggested that the SARMA model was probably never the correct specification.

Measuring “neighbors” with a Spatial Weights Matrix, W , allows us to mathematically specify how spatial relationships among regions might be structured. Regions might be related with their neighbors in three different ways:

1. The value of y in a region might impact (or be related to) the value of y in a neighboring region
2. The values of X 's in a region might affect (or be related to) the value of y in a neighboring region
3. The residuals ϵ might affect (or be related to) the residuals in a neighboring region (spatial heteroskedasticity or spatial autocorrelation)

Manski Model

$$y = \rho W y + X\beta + WX\theta + u \quad u = \lambda W u + \epsilon \quad (2)$$

The Manski Model builds in all three types of spatial relationship. If $\theta = 0$ then Manski becomes the Kelejian-Prucha Model:

Kelejian-Prucha, SARAR, Cliff-Ord, SAC model

$$y = \rho W y + X\beta + u \quad u = \lambda W u + \epsilon \quad (3)$$

Or if $\lambda = 0$, we get the Spatial Durbin Model (SDM), which involves lagged y and spatially-related residuals. [LeSage, Pace \(2009\)](#) favor starting with this specification (or the Spatial Durbin Error Model, discussed in the last of the core videos), and then using Likelihood Ratio techniques to test to see if the model should be restricted to a simpler, nested model.

Spatial Durbin Model

$$y = \rho W y + X\beta + WX\theta + \epsilon \quad (4)$$

If $\rho = 0$, then this becomes the

Spatially Lagged X (SLX) Model

$$y = X\beta + WX\theta + \epsilon \quad (5)$$

If $\theta = 0$, then (4) degenerates into the Spatial Lag Model

Spatial Lag, Spatial Autoregressive (SAR)

$$y = \rho W y + X\beta + \epsilon \quad (6)$$

If $\theta = -\rho\beta$, then (4) simplifies into the spatial error model (because $\lambda = \rho$ in this case)².

$$\begin{aligned} y &= \rho W y + X\beta + WX[\theta = -\rho\beta] + \epsilon \\ y &= \rho W[X\beta + \epsilon] + X\beta + WX[-\rho\beta] + \epsilon \\ y &= \rho W X\beta + \rho W \epsilon + X\beta + WX[-\rho\beta] + \epsilon \end{aligned}$$

Spatial Error (SEM)

$$y = X\beta + u \quad u = \lambda W u + \epsilon, \quad \text{where } \epsilon \sim i.i.d. \quad (7)$$

Of course, setting all of the spatial parameters (λ , ρ , and the vector of θ 's) to zero will restrict the model back to OLS.

3.2 GIS basics

The next two videos give a brief introduction to using QGIS ([QGIS Development Team 2018](#)). The focus is on how to download a SHP file mapping a set of regions, open it in a GIS program, edit it by removing regions that may not be of interest, and add additional data for the regions to be analyzed. Users are provided with a ZIP file to download that contains the map files and extra data to import for practice with the video.

3.3 Understanding the idea of contiguity and spatial correlation

A video focusing on contiguity and spatial correlation (Moran's I) follows, along with a downloadable spreadsheet file. In this file the viewers help to complete small contiguity and weights matrices by hand, using a simple set of regions. The users are shown how simple row standardization can be done. This spreadsheet allows users to change data values in each region, and dynamically see how the value of the spatial correlation changes as the user creates different patterns in the data.

²The lines below – although not mathematically correct – provide intuition for why the SEM is nested with the SDM. For the correct (though possibly less intuitive) derivation, see [LeSage, Pace \(2009, pp. 51-52\)](#).

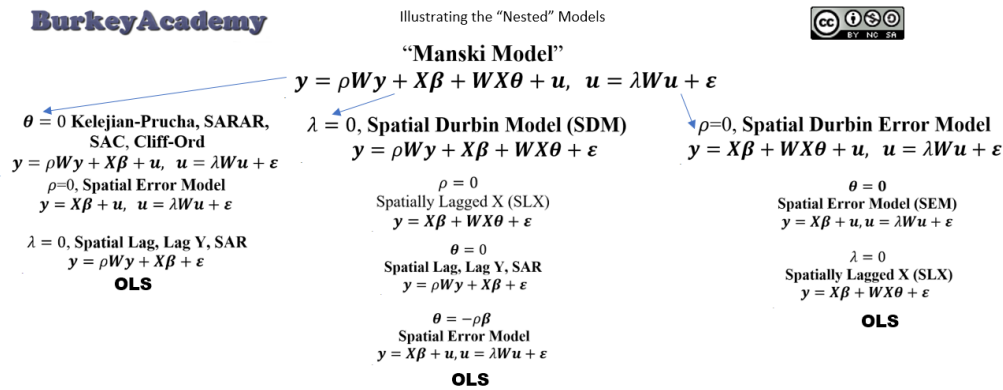


Figure 1: Part of the “Nested Models” Handout

3.4 ESDA, making and visualizing spatial weights for spatial data with GeoDa

The next four videos in the series use GeoDa’s fantastic visualization tools (Anselin et al. 2006) to do basic Exploratory Spatial Data Analysis (ESDA), create and export spatial weights matrices, calculate spatial correlations, and do basic spatial regressions and specification tests. The strengths and weaknesses of GeoDa are discussed, and the user is encouraged to also learn R for spatial data analysis in the remaining videos.

3.5 R for spatial data analysis and regression

In the final four videos of the sequence, the user is introduced to R, and many of the common packages used for spatial data manipulation and analysis (especially spdep (Bivand, Piras 2015, Bivand et al. 2013)). The first two videos focus on getting data into R and creating contiguity matrices. Viewers are shown how to read in spatial data, create weights matrices, import weights created in GeoDa, export weights, and plot contiguity relationships.

The final two videos in the core of the series explore ways to perform a spatial specification search and estimate many of the common spatial econometric models in R. The difference between Lagrange Multiplier (LM) tests and Likelihood Ratio (LR) tests are discussed. In summary, the LM tests begin with estimating a nonspatial OLS model and calculating the score, indicating the rate of improvement in the model fit as we relax the constraints making the model nonspatial. Anselin derived five LM tests: One for the SEM and SAR (with a robust version for these, attempting to filter out some propensity for a false positive between these two models), and a test for a SARMA Model³.

The benefit to the LM approach is that only the OLS model needs to be estimated before calculating the LM statistics. However, this method of spatial specification search is very limiting, because only three types of models can be considered. Thus, many spatial econometricians favor the LR approach, because any model can be tested to see if a simpler, nested model may be more appropriate (See Figure 1). The LR statistic is also very easy to calculate after running the two models to be compared:


$$LR_{\text{stat}} = -2(L_{\text{restricted}} - L_{\text{unrestricted}}) \sim \chi^2(\# \text{ restrictions}) \quad (8)$$

Ho: restricted model is true

LeSage (2014) suggests running the SDM if a global specification is needed, and if a local structure is needed, first run the Spatial Durbin Error Model (SDEM):

$$y = X \beta + W X \theta + u, \quad u = \lambda W u + \epsilon \quad (9)$$

³A SARMA is similar to a Kelejian-Prucha model, with a more quickly-decaying spatial impact.

The Mother of All R Spatial Econometrics Handouts! See spatial.burkeyacademy.com © Mark L. Burkey 

Version 0.65

YouTube Playlist:
https://www.youtube.com/playlist?list=PLInFW8Me14z6Du_chY6o08KcU6hNDkt4k

Work in Progress:
 Friday, June 29, 2018 Version
 For a more comprehensive view of R's spatial and mapping capabilities, check our Roger Bivand's list: <https://cran.r-project.org/web/views/Spatial.html>

Most of these commands rely on the `spdep` library, and you will need to install and load it `library(spdep)` beforehand. Other libraries will be specifically mentioned when called for, e.g., `rgeos`, `rgdal`

Brief R overview:
 Free! Not fast for huge data sets, but few have one ☺
 # is R's comment character, ignores words after #
 R is case sensitive, so `myvar` is not same as `Myvar`. You can name things `my.var` or `my.var.1` or `my.var.1`
`myvar=3` or `myvar<-3` assign value 3 to `myvar`

For creating a SHP spatial data set using QGIS, see <https://www.youtube.com/watch?v=fHRDAGc9-SE>
 Brief commentary on my method: I advise adding your data to the SHP file as in my videos, then reading it in and creating spatial weights (W) matrices in R (or creating weights in GeoDa from the same SHP file, and reading those in). While you can read in your data separately, be careful- it is possible that the ordering of the data and the weights will not match, and then Garbage in, Garbage out.

Reading in Shapefile with (or without) Data:

```
library(rgdal)
spat.data=readOGR(dsn=".",
layer="shpfilename")
  •dsn is directory (. is current)
  •layer is shp file name without .SHP
names(spat.data) show variable names
spplot(spat.data, "var") make map
```

Spatial Weights
 •Reading in external weights files:

```
nb.wts=read.gal("gal.file.gal")
nb.wts=read.gwt2nb("gwt.file.gwt")
```


 •Convert nb to listw (needed for regressions, etc.)

```
listw.wts=nb2listw(nb.wts)
```


 •Creating Spatial Weights in R

```
queen.nb=poly2nb(spat.data)
rook.nb=poly2nb(spat.data,queen=FALSE)
```


 Nearest neighbors requires a matrix of coords, longitude in left column. If you don't have coordinates (centroids):

```
library(rgeos)
cent=gCentroid(spat.data,byid=TRUE)
dnearneigh(coords matrix, d lower bound, d upper bound or all.linked)
```


 Distance-based nearest neighbors

```
knearneigh(coords matrix, k=4):4
nearest neighbors
```

Tools for Working with Weights
`is.symmetric.nb` is structure symmetric?
`diffnb(nb1, nb2)` differences in neighbor lists
`knn2nb` convert knn to nb format
`mat2listw, nb2mat`: convert to/from matrix
 Others: `droplinks`, `summary.nb`, `plot.nb`
`eigenw(listw)`: give eigenvalues of listw
More details: <https://cran.r-project.org/web/packages/spsden/vignettes/nb.pdf>

Estimating regressions
 Save model to a name: `reg.eql=y~x1+x2`
 OLS: $y = X\beta + \epsilon$

```
reg1=lm(reg.eql, data=spat.data)
```


 SLX Spatially Lagged X $y = X\beta + Wx\theta + \epsilon$

```
reg2=lmSLX(reg.eql, data=spat.data, listw)
```


 SAR Spatial Lag Model $y = \rho Wy + X\beta + \epsilon$

```
reg3=lagsarlm(reg.eql, data=spat.data, listw)
```


 SEM Spatial Error Model
 $y = X\beta + u, u = \lambda Wu + \epsilon$

```
reg4=errorsarlm(reg.eql, data=spat.data, listw)
```


 SDEM Spatial Durbin Error Model (add lag X to SEM) $y = X\beta + Wx\theta + u, u = \lambda Wu + \epsilon$

```
reg5=errorsarlm(reg.eql, data=spat.data, listw, etype="emixed")
```


 SDM Spatial Durbin Model (add lag X to SAR)
 $y = \rho Wy + X\beta + WX\theta + \epsilon$

```
reg6=lagsarlm(reg.eql, data=spat.data, listw, type="mixed")
```

Test Model Restrictions:
 LR: `sarlm(reg6, reg4)` likelihood ratio test to see if SDM should be restricted to the SEM.
 H₀ restrictions are true Can only test restriction to a simpler, nested model. SDM and SDEM are not nested, as one cannot be simplified into the other.

Marginal Effects of Models:
 When using any model with a ρWy , you cannot interpret the coefficients as marginal effects from `summary(reg1)` command. You must calculate the direct, indirect, and total marginal effects using the `impacts` command. (Also useful for total impacts for models with SLX)

```
impacts(reg1, listw=listw.wts)
summary(impacts(reg1, listw=listw.wts, R=500), zstats=TRUE) #for p values
```

Figure 2: Snapshot of the Spatial Econometrics Cheat Sheet (Burkey 2018)

This model is “local” because it does not contain a lag y term; so while neighbors affect each other, this effect does not propagate throughout the entire space. Links to the data, commands, and handouts are included, along with a two-page spatial econometrics in R reference sheet with commands and tips (Figure 2).

4 Conclusion

While this web page and YouTube series are not a perfect substitute for a short course, it is hoped that it helps to introduce Spatial Econometrics to a wider audience, and serve as a resource for those who cannot afford to take a face to face course. The author welcomes any suggestions and corrections from the academic community. The author hopes to add some interviews with leading researchers in the field to this series of videos, as well as add additional topics of interest.

Acknowledgement

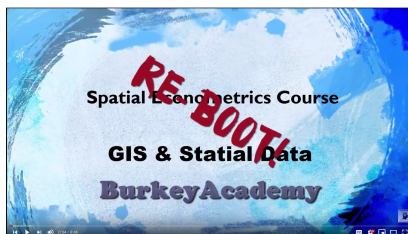
Thanks to Stephanie Kelly for editorial assistance.

References

- Anselin L (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht. [CrossRef](#).
- Anselin L, Syabri I, Kho Y (2006) Geoda: An introduction to spatial data analysis. *Geographical Analysis* 38[1]: 5–22. [CrossRef](#).
- Bivand RS, Hauke J, Kossowski T (2013) Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis* 45[2]: 150–179. [CrossRef](#).

- Bivand RS, Piras G (2015) Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software* 63[18]: 1–36. [CrossRef](#).
- Burkey ML (2015) Making educational and scholarly videos with screen capture software. *REGION* 2[2]: R2–10. [CrossRef](#).
- Burkey ML (2018) The mother of all R spatial econometrics handouts. Available from <http://spatial.burkeyacademy.com>
- LeSage JP (2014) What regional scientists need to know about spatial econometrics. *The Review of Regional Studies* 44[1]: 13–32
- LeSage JP, Pace RK (2009) *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, FL. [CrossRef](#).
- QGIS Development Team (2018) QGIS geographic information system. Open source geospatial foundation project. <http://qgis.osgeo.org>

A list of the videos with descriptions and links to the supplementary material can be found at <http://spatial.burkeyacademy.com>.



The playlist of videos is at https://www.youtube.com/playlist?list=PLlnEW8MeJ4z6Du_cb-Y6o08KsU6hNDkt4k



© 2018 by the authors. Licensee: REGION – The Journal of ERSA, European Regional Science Association, Louvain-la-Neuve, Belgium. This article is distributed under the terms and conditions of the Creative Commons Attribution, Non-Commercial (CC BY NC) license (<http://creativecommons.org/licenses/by-nc/4.0/>).