`GeoDaSpace`: advanced spatial econometrics made easy

V Spatial Econometrics workshop in honour of Jean Paelink
Coimbra (Portugal)

Dani Arribas-Bel

Dpt. of Spatial Economics
VU University - Amsterdam (Netherlands)

October 2012

Outline

Outline

Outline

Outline

Outline

Material from this presentation

GeoDaSpace

`http://geodacenter.asu.edu/software`

Data

`http://geodacenter.asu.edu/sdata` → **Phoenix ACS dataset**

Slides and PySAL tutorial

`http://github.com/darribas/coimbra2012`

Background - PySAL

Who

Sergio J. Rey & Luc Anselin

Where

- GeoDa Center for Geospatial Analysis and Computation at ASU
- Worldwide at http://pysal.org

What

- Python library of (state of the art) spatial analysis
- Weights, ES(T)DA, inequality, networks, **spatial regression**

Why

- Not *reinvent the wheel*
- One code base, multiple delivery formats (command line, desktop GUI, web...)

Overview of GeoDaSpace

- Front-end GUI for regression modules in PySAL:
  - **I/O**: csv, dbf, shp
  - **weights**: creation, manipulation
  - **spreg**: state-of-the-art spatial econometrics
- Speed and scalability (sparse matrices)
- Intuitive and easy to use:
  - "Point and click"
  - Save model specification and load them later
  - Save results
- Cross-platform: Windows, Mac, (Linux)

Implemented models

Models	Methods	Refs.
Non spatial	OLS	
	2SLS	
Spatial Lag	S2SLS	
Spatial Error	GMM	KP98-99
	GMM-Het	Arraiz et al. 2010
	GMM-Hom	Drukker et al. 2010
	sHAC	KP2007
Endogenous regressors + Error	2SLS + GMM	KP98/99
	2SLS + GMM-Het	Arraiz et al. 2010
	2SLS + GMM-Hom	Drukker et al. 2010
	2SLS + sHAC	KP2007
Combo (Lag+Error)	S2SLS + GMM	KP98/99
	S2SLS + GMM-Het	KP98/99 + Arraiz et al. 2010
	S2SLS + GMM-Hom	KP98/99 + Drukker et al. 2010
	S2SLS + sHAC	KP98/99 + KP2007
Combo (Lag+Error) + Endogenous Regressors	S2SLS + GMM	KP98/99
	S2SLS + GMM-Het	KP98/99 + Arraiz et al. 2010
	S2SLS + GMM-Hom	KP98/99 + Drukker et al. 2010
	S2SLS + sHAC	KP98/99 + KP2007

Example data

- MSA of Phoenix
- 2010 US Census tracts (n=685)

Dependent variable

- Sampling error in estimates for p/c income as percentage of the actual estimate: accuracy of income data
- Source: American Community Survey 2005/09

Explanatory variables

- Socio-demographics from 2010 Decenial Census
- Housing units, rate of vacant housing units, population density,

rate of renters, race (rate of white, black and hispanic), rate of female head of the household (no husband)

Hands on!!!

Models and Methods - Non spatial

| Non spatial | OLS | |
| | 2SLS | |

Model

Traditional basic model

- $y = \beta X + \epsilon$

Non spatial endogenous regressors

- $y = \beta X + \gamma Y + \epsilon$

Methods

- OLS
- Two stages least squares

Models and Methods - Spatial lag

Model

The dependent variable is spatially lagged

- $y = \rho W y + \beta X + \epsilon$

Method

- Spatial Two Stages Least Squares

Models and Methods - Spatial error

Spatial Error		GMM	KP98-99
		GMM-Het	Arraiz et al. 2010
		GMM-Hom	Drukker et al. 2010
		sHAC	KP2007

Model

$$y = \beta X + u$$
$$u = \lambda W u + \epsilon$$

Methods

- OLS + Basic GM (λ as point estimate)
- OLS + GM allowing for heteroskedasticity in the residuals
- OLS + GM assuming homoskedasticity in the residuals
- OLS + Spatial Heteroscedasticity and Autocorrelation Consistent (spatial HAC) of the residuals - Does not assume error structure

Models and Methods - Endogenous reg. + sp. error

Endogenous regressors + Error	2SLS + GMM	KP98/99
	2SLS + GMM-Het	Arraiz et al. 2010
	2SLS + GMM-Hom	Drukker et al. 2010
	2SLS + sHAC	KP2007

Model

$$y = \beta X + \gamma Y + u$$
$$u = \lambda W u + \epsilon$$

Methods

- 2SLS + Basic GM (λ as point estimate)
- 2SLS + GM allowing for heteroskedasticity in the residuals
- 2SLS + GM assuming homoskedasticity in the residuals
- 2SLS + Spatial Heteroscedasticity and Autocorrelation Consistent (spatial HAC) of the residuals - Does not assume error structure

Models and Methods - Combo

Combo (Lag+Error)	S2SLS + GMM	KP98/99
	S2SLS + GMM-Het	KP98/99 + Arraiz et al. 2010
	S2SLS + GMM-Hom	KP98/99 + Drukker et al. 2010
	S2SLS + sHAC	KP98/99 + KP2007

Model

$$y = \rho Wy + \beta X + u$$
$$u = \lambda Wu + \epsilon$$

Methods

- S2SLS + Basic GM (λ as point estimate)
- S2SLS + GM allowing for heteroskedasticity in the residuals
- S2SLS + GM assuming homoskedasticity in the residuals
- S2SLS + Spatial Heteroscedasticity and Autocorrelation Consistent (spatial HAC) of the residuals - Does not assume error structure

Combo (Lag+Error) + Endogenous Regressors	S2SLS + GMM	KP98/99
	S2SLS + GMM-Het	KP98/99 + Arraiz et al. 2010
	S2SLS + GMM-Hom	KP98/99 + Drukker et al. 2010
	S2SLS + sHAC	KP98/99 + KP2007

Model

$$y = \rho Wy + \beta X + \gamma Y + u$$
$$u = \lambda Wu + \epsilon$$

Methods

- 2SLS + Basic GM (λ as point estimate)
- 2SLS + GM allowing for heteroskedasticity in the residuals
- 2SLS + GM assuming homoskedasticity in the residuals
- 2SLS + Spatial Heteroscedasticity and Autocorrelation Consistent (spatial HAC) of the residuals - Does not assume error structure

A sneak peek into the future of GeoDaSpace

Inminent

- Spatial regimes

Medium/long run

- Maximum Likelihood (PySAL)
- Spatial Panels

Spatial Econometrics with PySAL

Contact and resources

- GeoDaSpace/`spreg` tutorial by L. Anselin coming out soon!

`http://geodacenter.asu.edu`

`http://groups.google.com/group/openspace-list`

@geodacenter

`https://www.facebook.com/geodacenter`