GeoDaSpace: advanced spatial econometrics made easy

Eureka Seminar
Dept. Spatial Economics (VU)

Dani Arribas-Bel

GeoDa Center for Geospatial Analysis and Computation (ASU)

August 2011

Outline

Outline

Outline

Outline

GeoDa Center for Geospatial Analysis and Computation

People

- Five core faculty (Luc Anselin, Sergio Rey, Alan Murray, Elizabeth Wentz, Julia Koschinsky)
- Post-docs, graduate students and visitors
- Multi-disciplinary: geographers, economists, urban planners, computer scientists

About

- Methods development
- Implementation through **software** tools
- Policy relevant research
- Dissemination through training and support

The *GeoDa* family

All GeoDa Center software is freeware...

...some is also open source.

http://geodacenter.asu.edu/software

Freeware

- No cost to use, easy to distribute
- Black box

Open Source

- Source code is released as well
- User can study, modify and run it freely

The *GeoDa* family

Legacy GeoDa - OpenGeoDa

- Interactive ESDA and (a bit of) regression

STARS

- Interactive Exploratory Space-Time Analysis

GeoDaSpace

- Advanced spatial regression

GeoDaNet

- Point patterns on networks

PySAL

- Core Python library

PySAL

One library to rule them all

- Flexible and modular **software library that powers multiple front-ends** (command line, GUIs, web, etc.)
- Rapid development cycle
- State of the art methods in spatial analysis

Functionality

- I/O & weights
- ESDA
- Spatial dynamics
- Inequality
- Regions
- **Spatial Regression**

Release Schedule

- Currently 1.2 (6 months)
- **pysal.org**

GeoDaSpace - Team

- Pedro Amaral
- Luc Anselin
- Dani Arribas-Bel
- David Folch
- Nancy Lozano-Gracia
- Nicholas Malizia
- Serge Rey
- Charles Schmidt
- Ran Wei
- Jing Yao

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)

GeoDaSpace - Models and Methods

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

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

Spatial Lag S2SLS

Model

The dependent variable is spatially lagged

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

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

Models and Methods - Combo + end. reg.

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

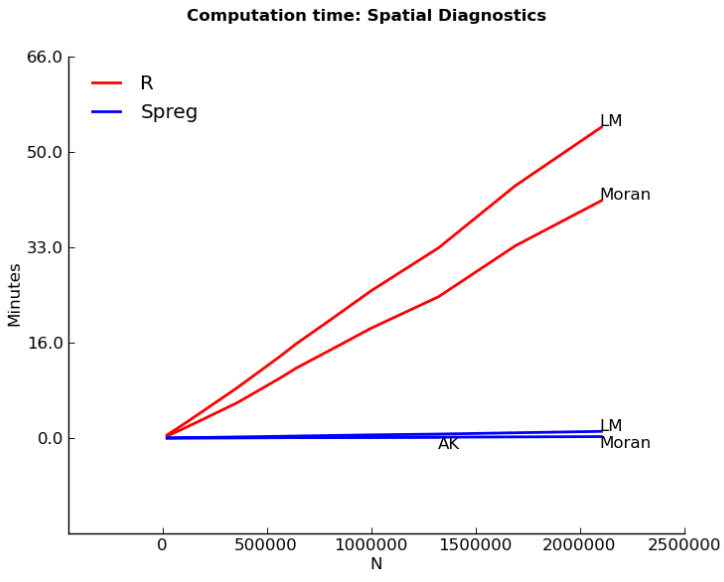
Model

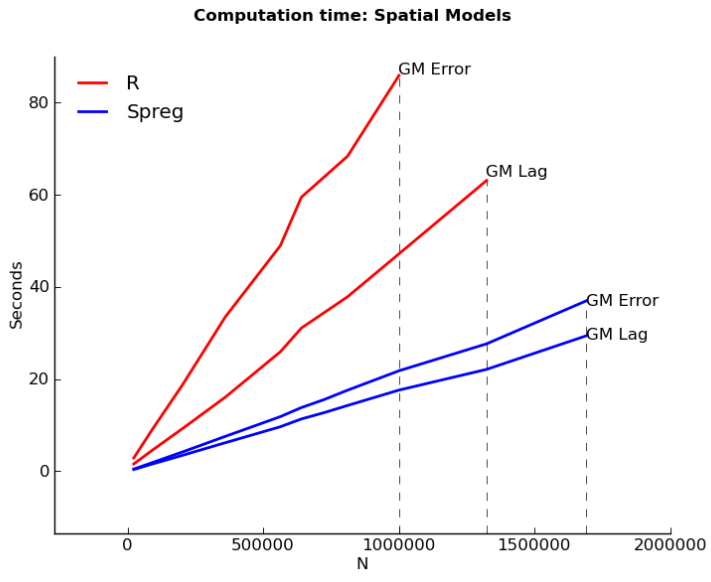
$$y = \rho Wy + \beta X + \gamma Y + u$$
$$u = \lambda Wu + \epsilon$$

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

GeoDaSpace - Performance





i) All simulations run on a 4 core (3GHz) Mac Pro with 16Gb RAM. Software:

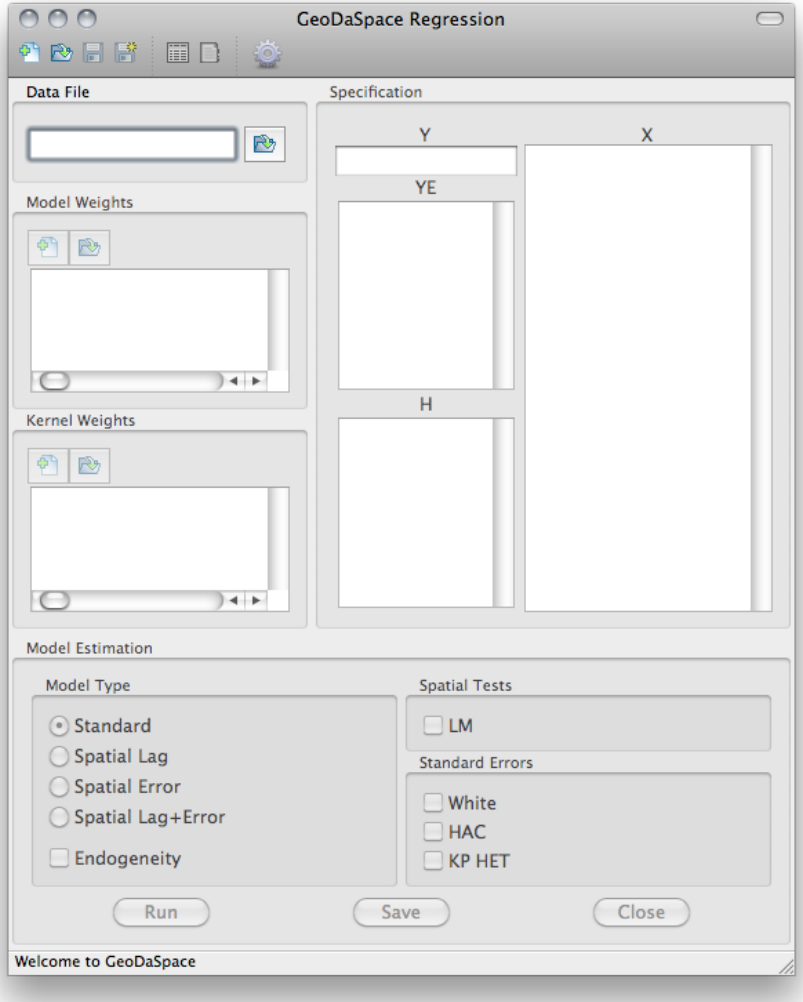
- R (2.12.0), spdep (0.5-29) - Python (2.6.5), PySAL (1.1), numpy (1.4), scipy (0.8.0b1)

ii) Geography setup: square lattice regions ($\sqrt{N}$ by $\sqrt{N}$) and queen contiguity matrices

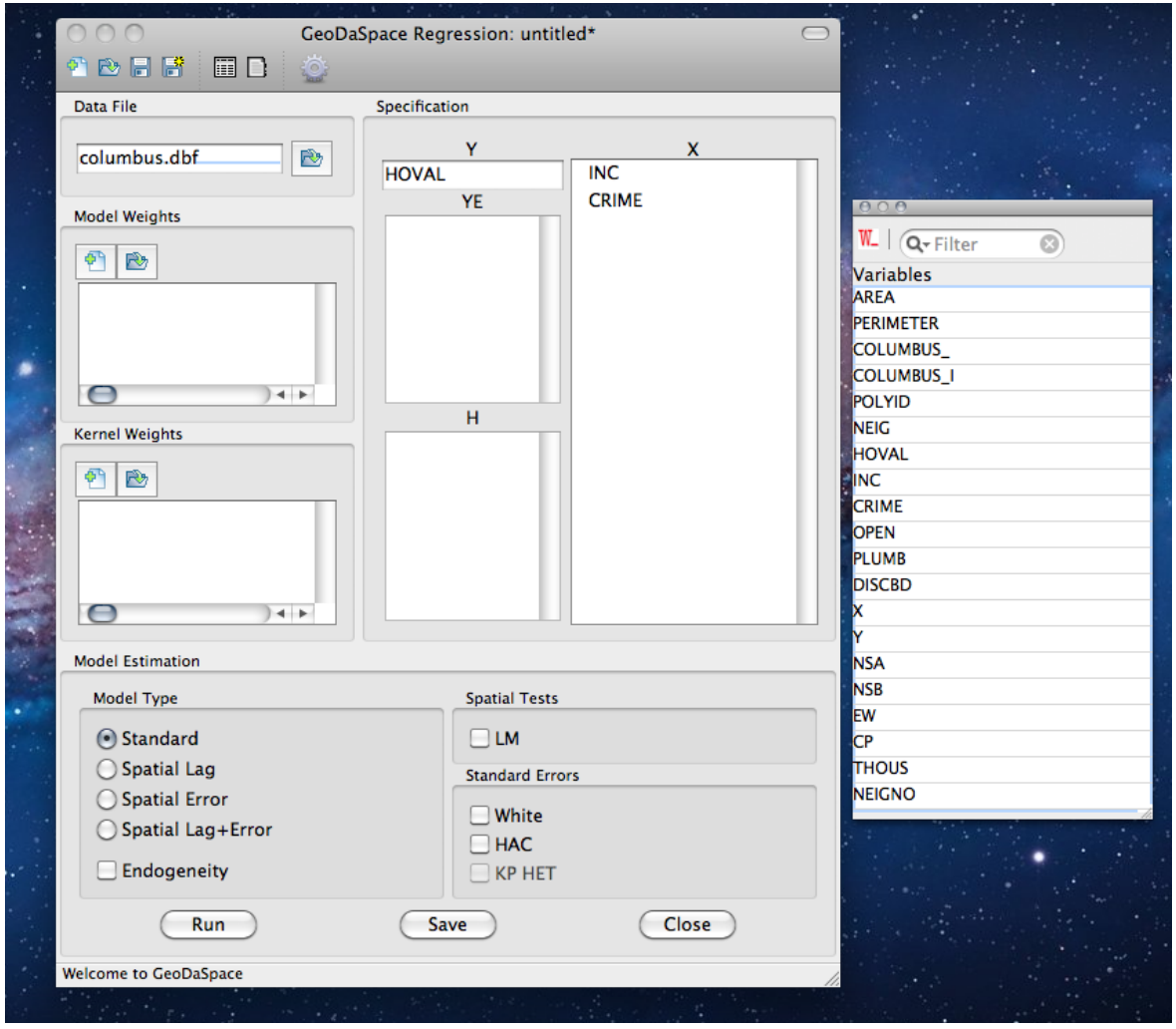
iii) N = 22,500; 90,000; 202,500; 360,000; 562,500; 640,000; 722,500; 810,000; 1,000,000; 1,322,500; 1,690,000; 2,102,500; 2,560,000

iv) Regressions: 5 independent random variables, N(0, 1) against one dependent random variable

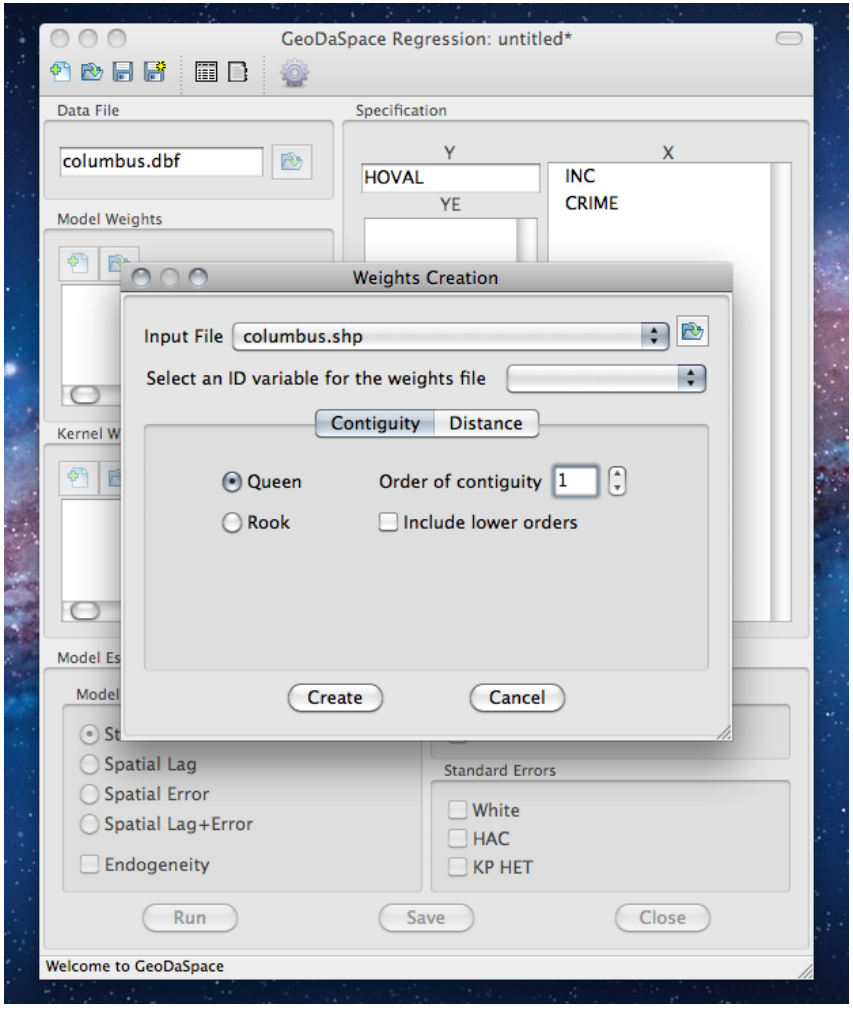
GeoDaSpace - GUI

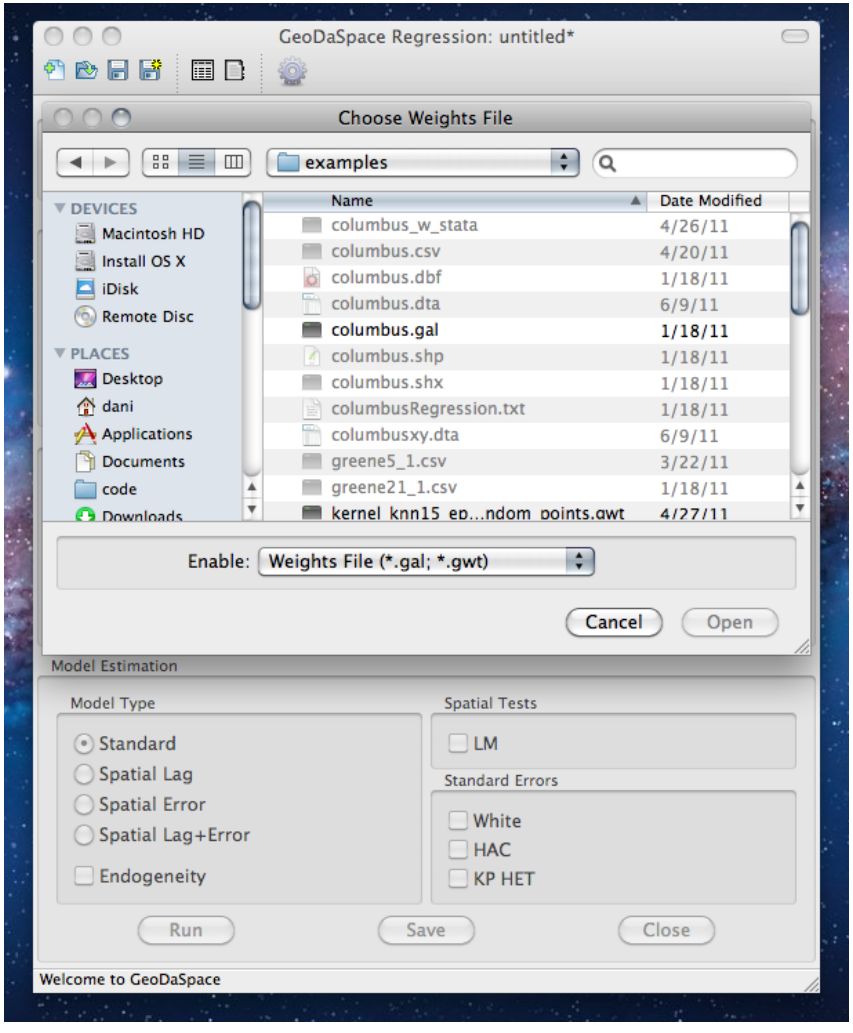


GeoDaSpace - GUI

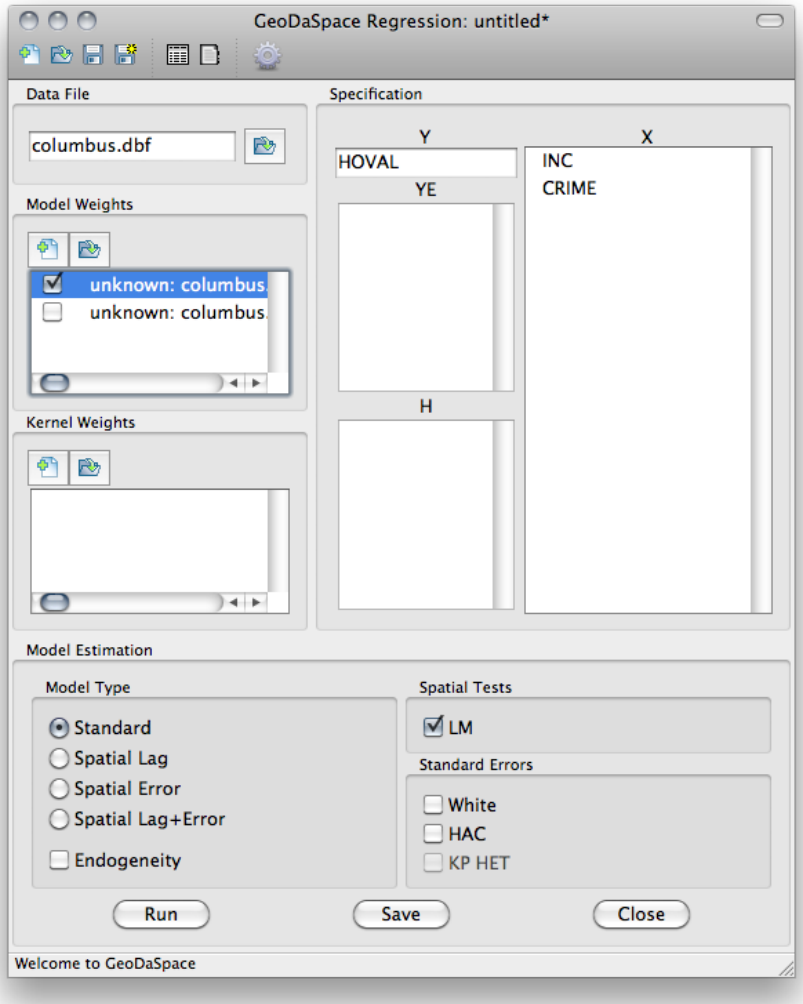


GeoDaSpace - GUI - Weights





GeoDaSpace - GUI - Run model

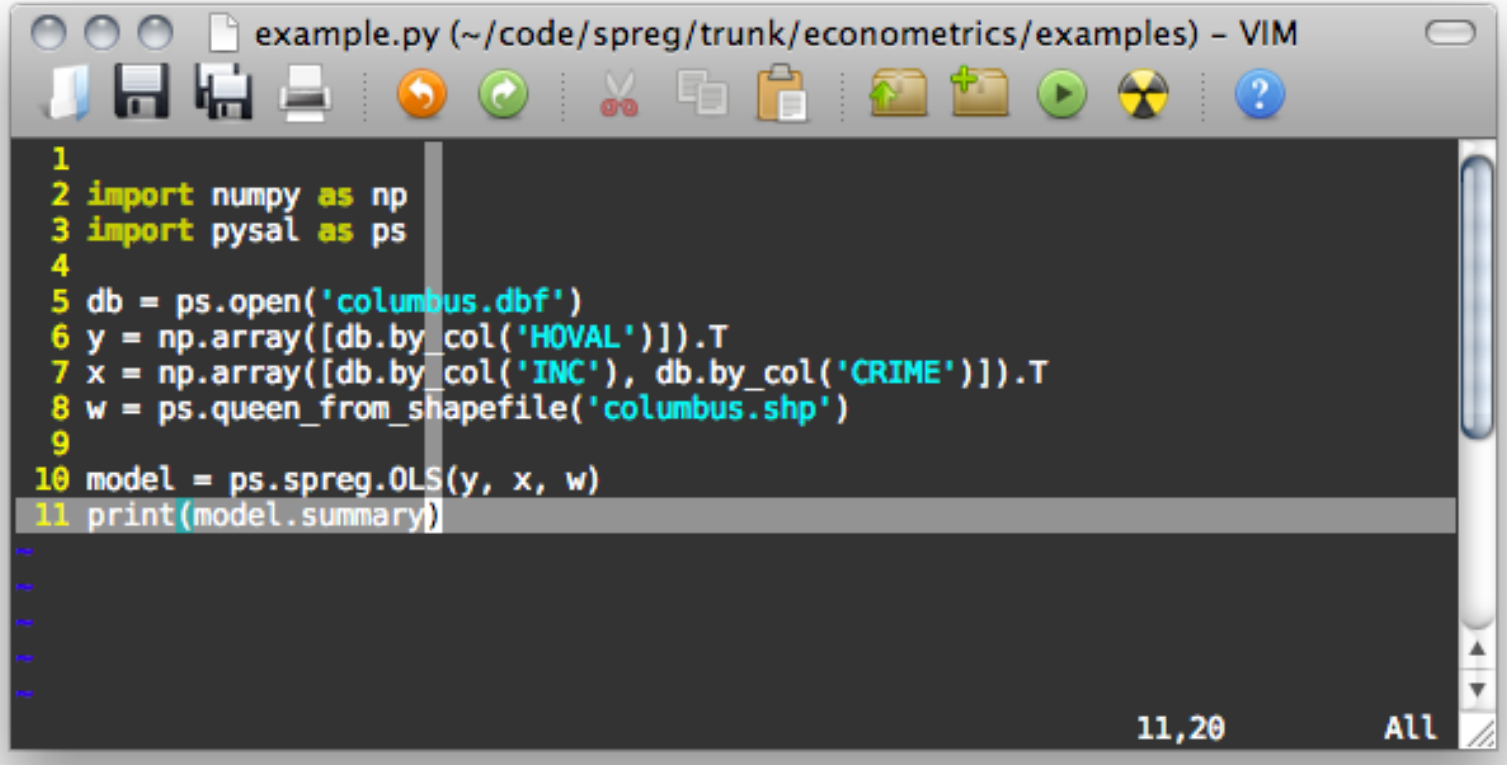


GeoDaSpace - GUI - Report

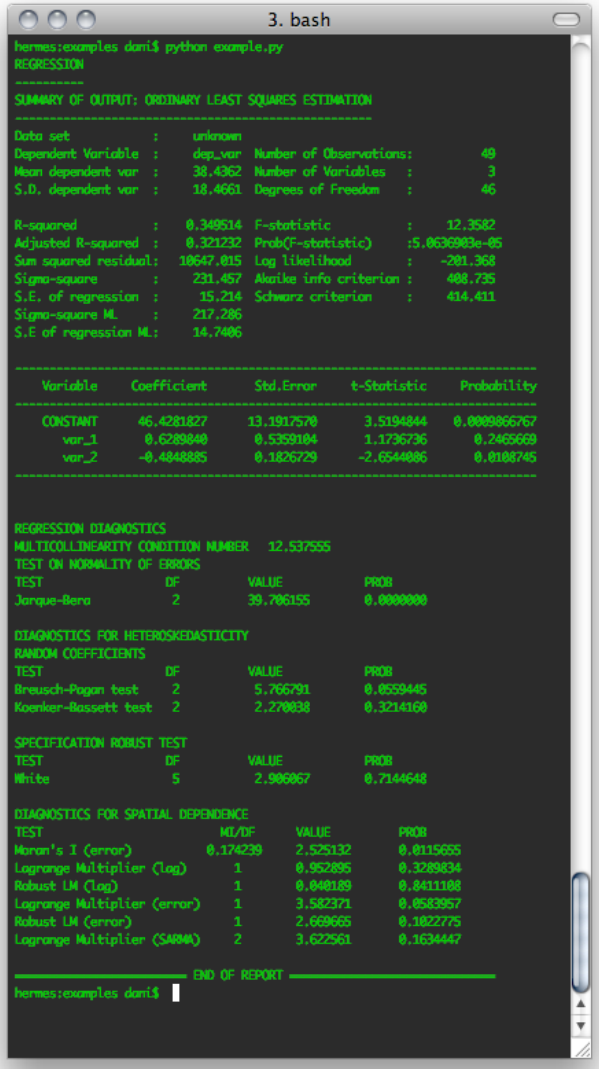
Untitled*

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
---------------------------------------------------
Data set            :/Users/dani/code/spreg/trunk/econometrics/examples/columbus.dbf
Dependent Variable  :       HOVAL  Number of Observations:          49
Mean dependent var  :     38.4362  Number of Variables   :           3
S.D. dependent var  :     18.4661  Degrees of Freedom    :          46

R-squared           :    0.349514  F-statistic           :     12.3582
Adjusted R-squared  :    0.321232  Prob(F-statistic)     :5.0636903e-05
Sum squared residual:   10647.015  Log likelihood        :    -201.368
Sigma-square        :     231.457  Akaike info criterion :     408.735
S.E. of regression  :      15.214  Schwarz criterion     :     414.411
Sigma-square ML     :     217.286
S.E of regression ML:     14.7406

------------------------------------------------------------------------------
    Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------
    CONSTANT      46.4281827      13.1917570       3.5194844    0.0009866767
         INC       0.6289840       0.5359104       1.1736736       0.2465669
       CRIME      -0.4848885       0.1826729      -2.6544086       0.0108745
------------------------------------------------------------------------------


REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   12.537555
TEST ON NORMALITY OF ERRORS
TEST                     DF          VALUE            PROB
Jarque-Bera               2          39.706155        0.0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                     DF          VALUE            PROB
Breusch-Pagan test        2           5.766791        0.0559445
Koenker-Bassett test      2           2.270038        0.3214160
SPECIFICATION ROBUST TEST
TEST                     DF          VALUE            PROB
White                     5           2.906067        0.7144648

DIAGNOSTICS FOR SPATIAL DEPENDENCE
TEST                           MI/DF       VALUE           PROB
Moran's I (error)           0.174239       2.525132        0.0115655
Lagrange Multiplier (lag)          1       0.952895        0.3289834
Robust LM (lag)                    1       0.040189        0.8411108
Lagrange Multiplier (error)        1       3.582371        0.0583957
Robust LM (error)                  1       2.669665        0.1022775
Lagrange Multiplier (SARMA)        2       3.622561        0.1634447

======================== END OF REPORT =============================
```

GeoDaSpace - Command line



GeoDaSpace - Command line



Timeline

Current

- OLS (PySAL)
- Diagnostics (PySAL)

Imminent future

- GUI release (soon!!!)
- All models and methods as part of PySAL (Jan. 2012)

Medium run

- Maximum likelihood
- Spatial regimes

Get in touch!

http://geodacenter.asu.edu

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

@geodacenter

https://www.facebook.com/geodacenter