ERSA Summer School 2006

Spatial Analysis with GeoDa

4. Spatial Regression

Luc Anselin
University of Illinois, Urbana-Champaign

http://sal.uiuc.edu/

Outline

- Spatial Regression Basics
- Diagnostics for Spatial Autocorrelation
- Estimation

Spatial Regression Basics

Spatial Effects

- Spatial Dependence
  - non-zero correlation between pairs of locations
  - range/structure of dependence
  - two-dimensional and multi-directional
- Spatial Heterogeneity
  - structural change across space
  - spatial structure of the heterogeneity

Four Elements of Spatial Econometrics

- Specifying the Structure of Spatial Dependence
- Testing for the Presence of Spatial Effects
- Estimating Models with Spatial Effects
- Spatial Prediction

Spatial Variables

- Spatial Lag
  - weighted average of neighbors
  - $\sum_j w_{ij} y_j$, with $w_{ii} = 0$
- Spatially Lagged Variables
  - explanatory variables: WX
  - dependent variable: Wy
  - error $W\epsilon$

Spatial Regression Model

- Incorporate Spatially Lagged Variables
- Spatially Lagged Dependent Variable
  - spatial lag model
- Spatially Lagged Error Term
  - spatial error model
- Spatially Lagged Explanatory Variables
  - spatial cross-regressive model

Spatial Lag Model

- Mixed Regressive Spatial Autoregressive Model
  - Wy = spatial autoregressive
  - X = regressive
- $y = \rho Wy + X\beta + \epsilon$
  - ρ = spatial autoregressive coefficient

Spatial Filter

- Remove Effect of Spatial Correlation
  - $y - \rho Wy = X\beta + \epsilon$
  - $(I - \rho W)y = X\beta + \epsilon$
  - $(I - \rho W)$ is spatial filter
- Similar to Detrending
  - but spatial autoregressive coefficient $\neq 1$
  - ρ must be estimated

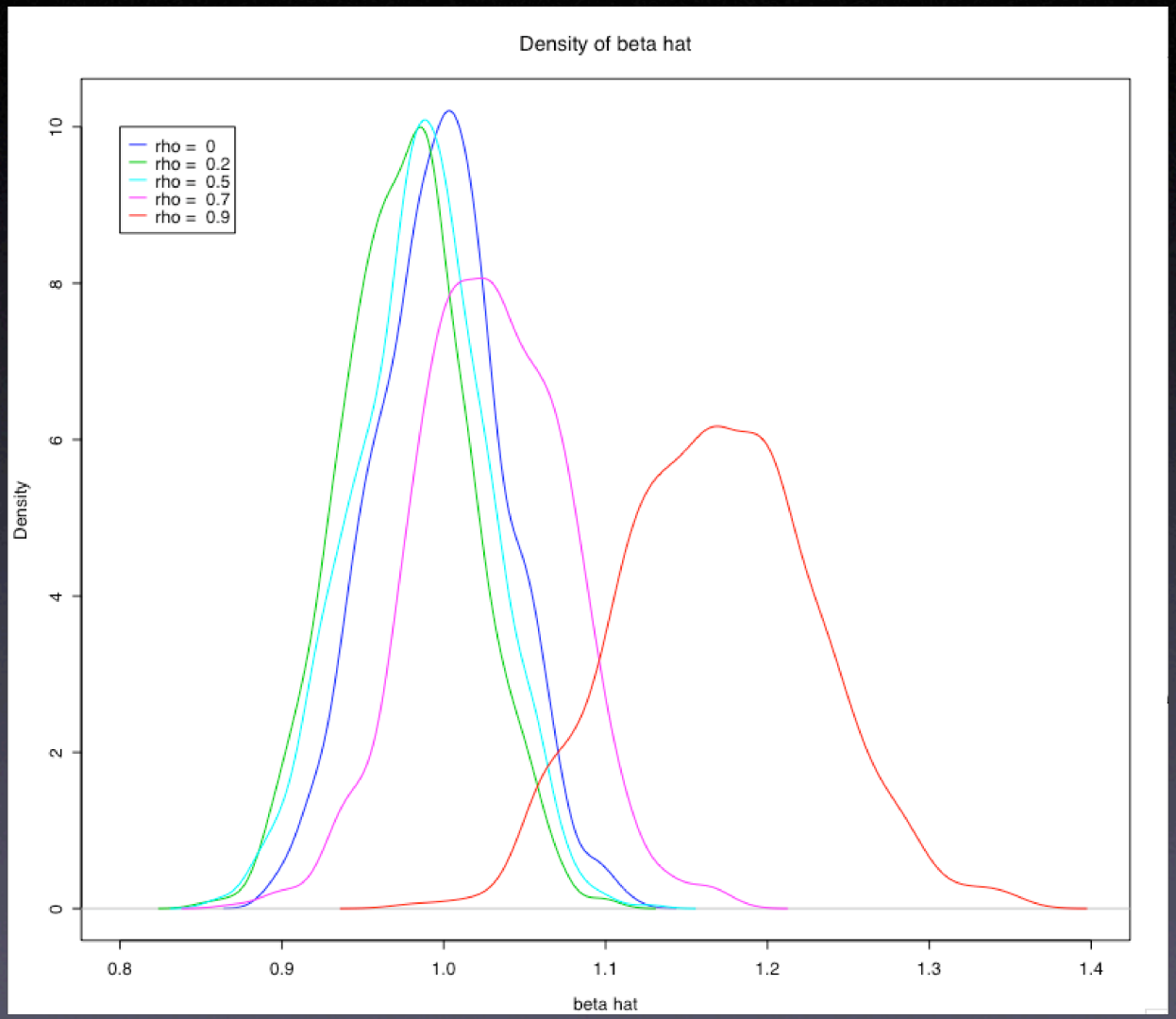
Spatial Multiplier

- Reduced Form
  - $y - \rho W y = X\beta + \epsilon$
  - $(I - \rho W)y = X\beta + \epsilon$
  - $y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\epsilon$
- Spatial Multiplier
  - $E[\, y \mid \Delta X \,] = (I - \rho W)^{-1} (\Delta X)\beta$
  - effect is more than $(\Delta X)\ \beta \Rightarrow$ multiplier

Ignoring Spatial Lag

- Omitted Variable Problem
- OLS Becomes Biased and Inconsistent
  - wrong estimate
  - wrong sign
  - wrong significance
  - wrong fit

Effects of Spatial Lag on OLS



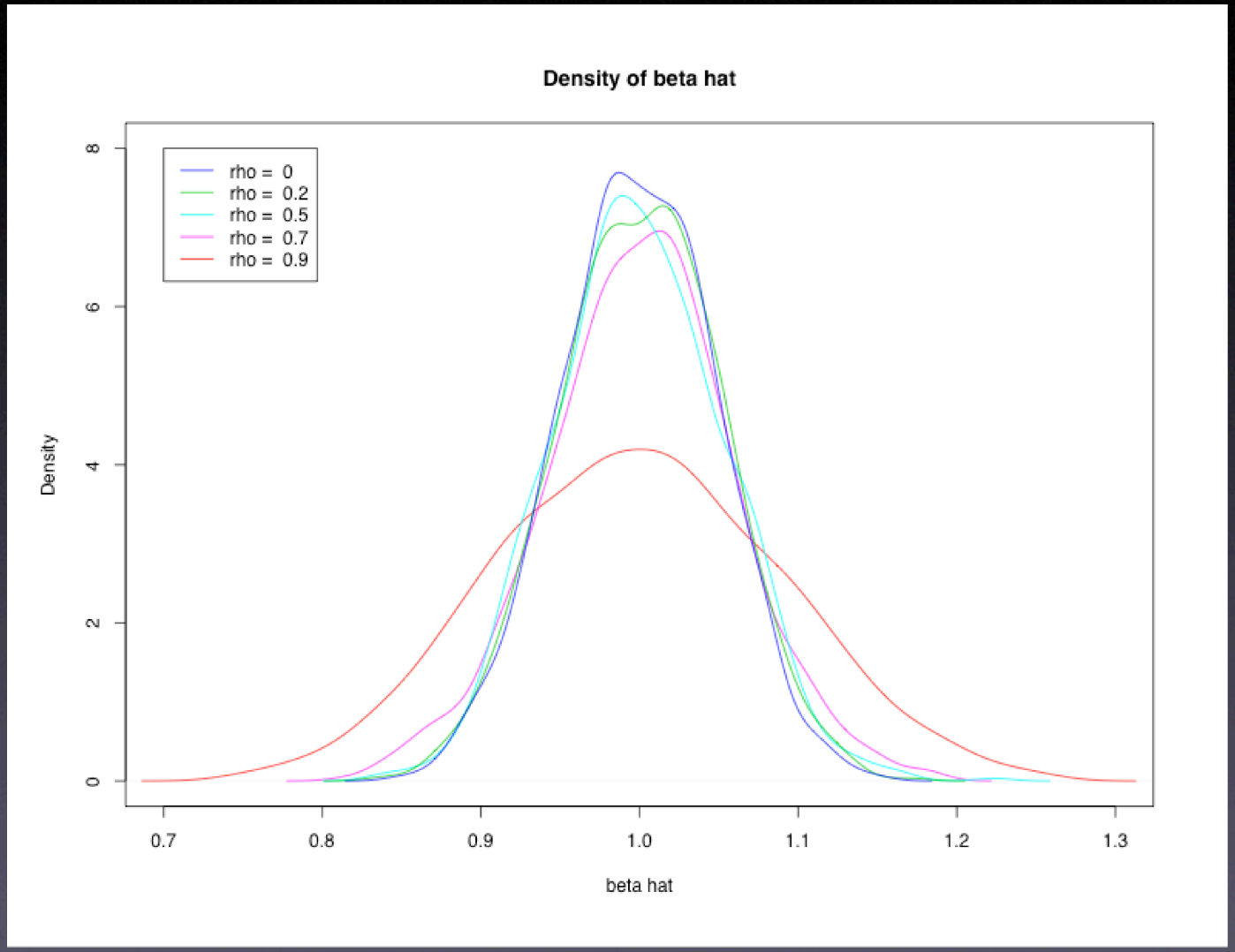
Spatial Error Model

- Non-Diagonal Error Covariance Matrix
- $E[\epsilon\epsilon'] = \Sigma \neq \sigma^2 I$
  - $E[\epsilon_i \epsilon_j] \neq 0$, for $i \neq j$
  - spatial structure in covariance
  - spatial process model
- $y = X\beta + \epsilon, \ \epsilon = \lambda W\epsilon + u$
  - spatial autoregressive error (SAR) model

Ignoring Spatial Error

- Problem of Efficiency
- OLS Remains Unbiased But Inefficient
  - correct estimate
  - wrong significance
  - wrong fit

Effects of Spatial Error on OLS



Diagnostics

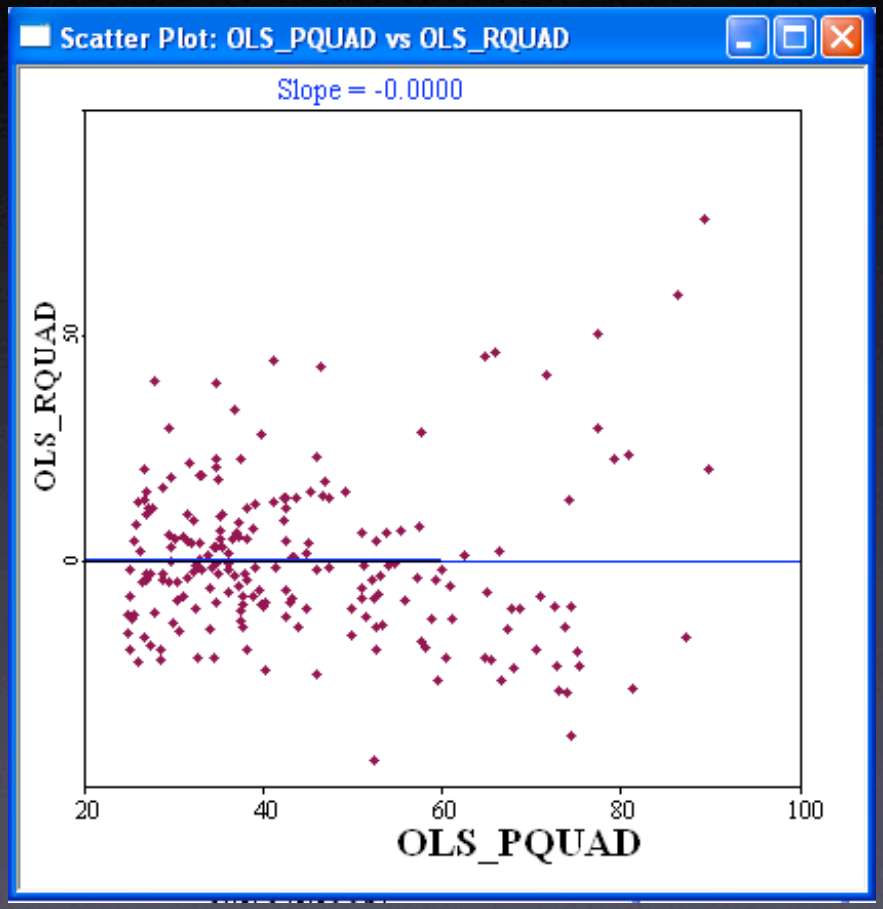
Specification Tests

- Null Hypothesis
  - assumptions of classic linear regression hold
  - error terms constant variance and uncorrelated
- Alternative Hypothesis
  - non-constant variance: heteroskedasticity
  - correlation: spatial lag or spatial error

Informal Diagnostics Non-Spatial

- Residual Plots
- Outlier Detection
  - plot residuals against observation numbers
  - identify large residuals
- Heteroskedasticity Check
  - plot residuals against predicted values
  - informal check on heteroskedasticity

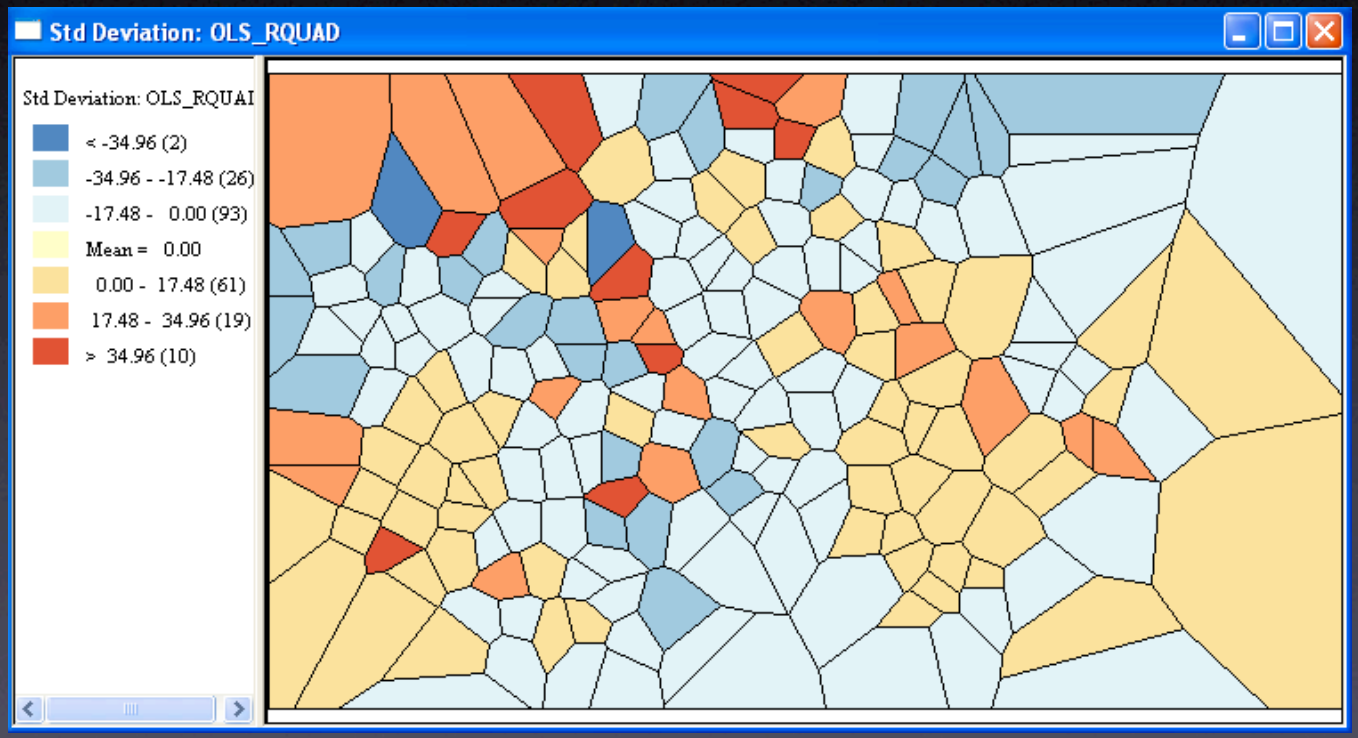
Residual Plot



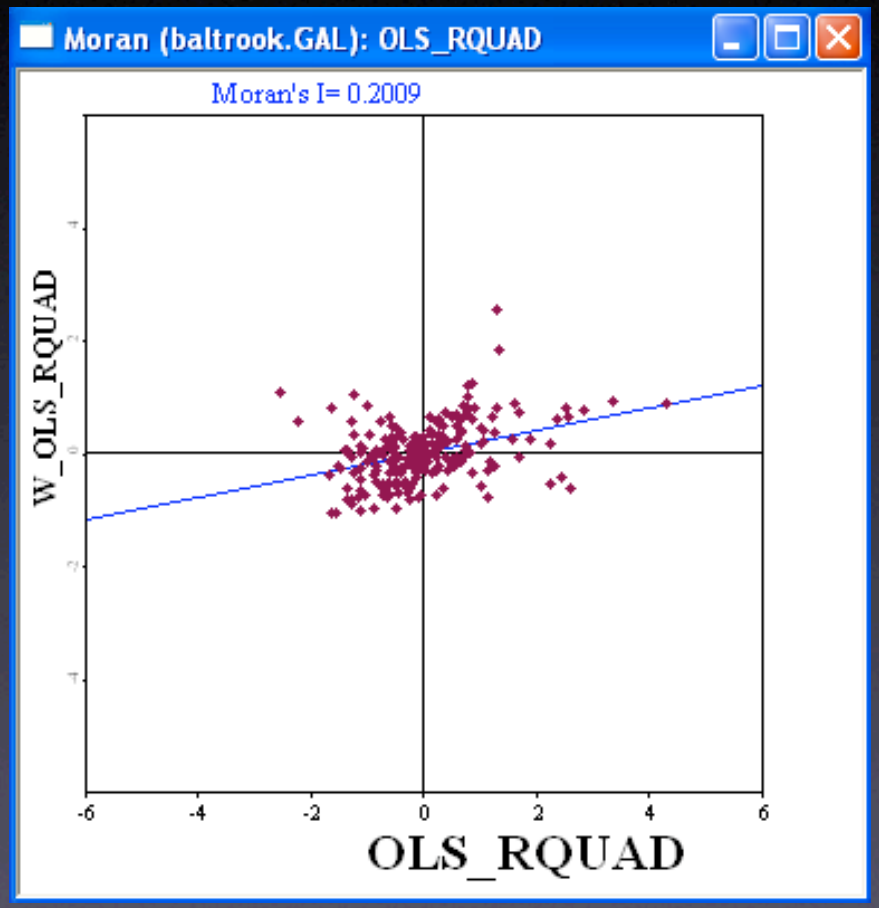
Informal Diagnostics

- Residual Maps
  - standard deviational map
  - informal check for spatial patterns
- Spatial Autocorrelation
  - Moran scatter plot
  - informal only, no inference

Residual Map



Moran Scatter Plot



Tests Against Spatial Autocorrelation

- Constrained Model
  - model without spatial correlation
  - classic regression specification
  - $H_0 : \lambda = 0$ (error) or $\rho = 0$ (lag)
- Unconstrained Model
  - under the alternative
  - spatial error: $H_1: \lambda \neq 0$
  - spatial lag: $H_1: \rho \neq 0$

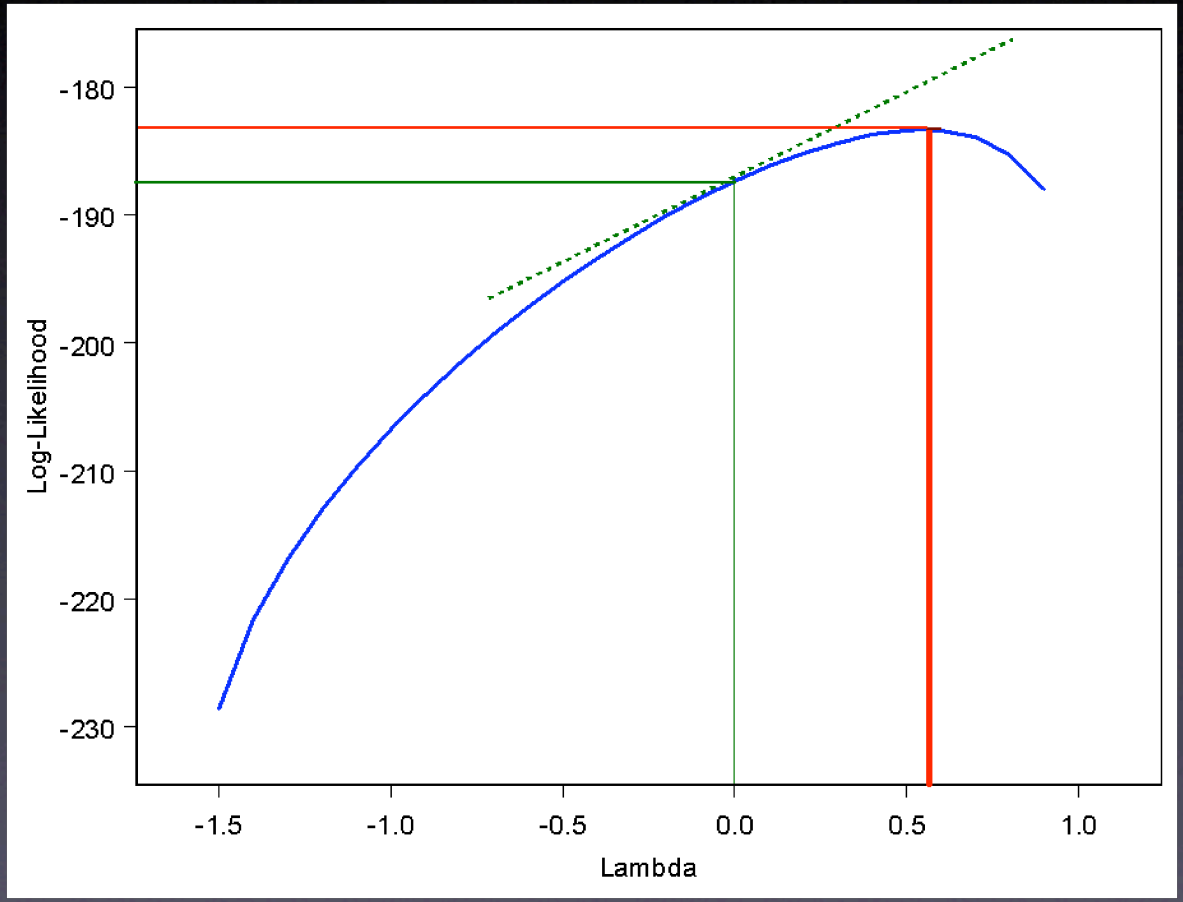
Moran's I

- Based on Regression Residuals
  - $e = y - Xb_{OLS}$
- Moran's I
  - I = [e'We/S0] / [e'e/N]
  - row-standardized weights: $I = e'We / e'e$
- Inference
  - normal approximation for z-value
  - e are residuals → adjusted inference

Tests Based on Maximum Likelihood

- Classic Tests
  - compare unconstrained to constrained results
  - require complete specification of likelihood function
  - assumption of normality

Three Classic Tests



ML Estimation Error Model

Wald Test

- Difference in Value
- Asymptotic t-Test
  - difference ML estimate and 0
  - requires estimation of alternative
  - requires asymptotic variance matrix

Likelihood Ratio Test

- Difference in Fit
- Compare Log-Likelihoods
  - difference max log likelihood spatial model and null model
  - requires estimation of null and alternative
  - requires maximized log likelihood

Lagrange Multiplier Test

- Difference in Slope
- Test on the Score
  - slope of likelihood function (score)
  - only requires estimation of null model (OLS)
  - but full specification of log-likelihood

LM Test Error

- LM Test Against Spatial Error Correlation
  - both SAR and SMA alternatives
  - LM-Err = $[e'We / \sigma^2]^2 / T \sim \chi^2(1)$
  - with e = residuals
  - T = tr(WW + W'W), matrix traces

LM Test Lag

- LM Test Against Spatial Lag
  - LM-Lag = $[e'Wy / \sigma^2]^2 / T_1 \sim \chi^2(1)$
  - $T_1 = (WXb)'M(WXb)/\sigma^2 + T$
  - first term: residual sum of squares of WXb on X

Robust LM Tests

- LM-Error and LM-Lag Tests Have Power Against the Other Alternative
- Robust Forms of the Test Statistics
  - make an asymptotic adjustment
- Use to Find Proper Alternative
  - only use robust forms when BOTH LM-Error and LM-Lag are significant

LM Tests in GeoDa

DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : **baltrook.GAL** (row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.200899	5.8323965	0.0000000
Lagrange Multiplier (lag)	1	20.5454453	0.0000058
Robust LM (lag)	1	1.2626832	0.2611438
Lagrange Multiplier (error)	1	23.6063131	0.0000012
Robust LM (error)	1	4.3235510	0.0375884
Lagrange Multiplier (SARMA)	2	24.8689962	0.0000040

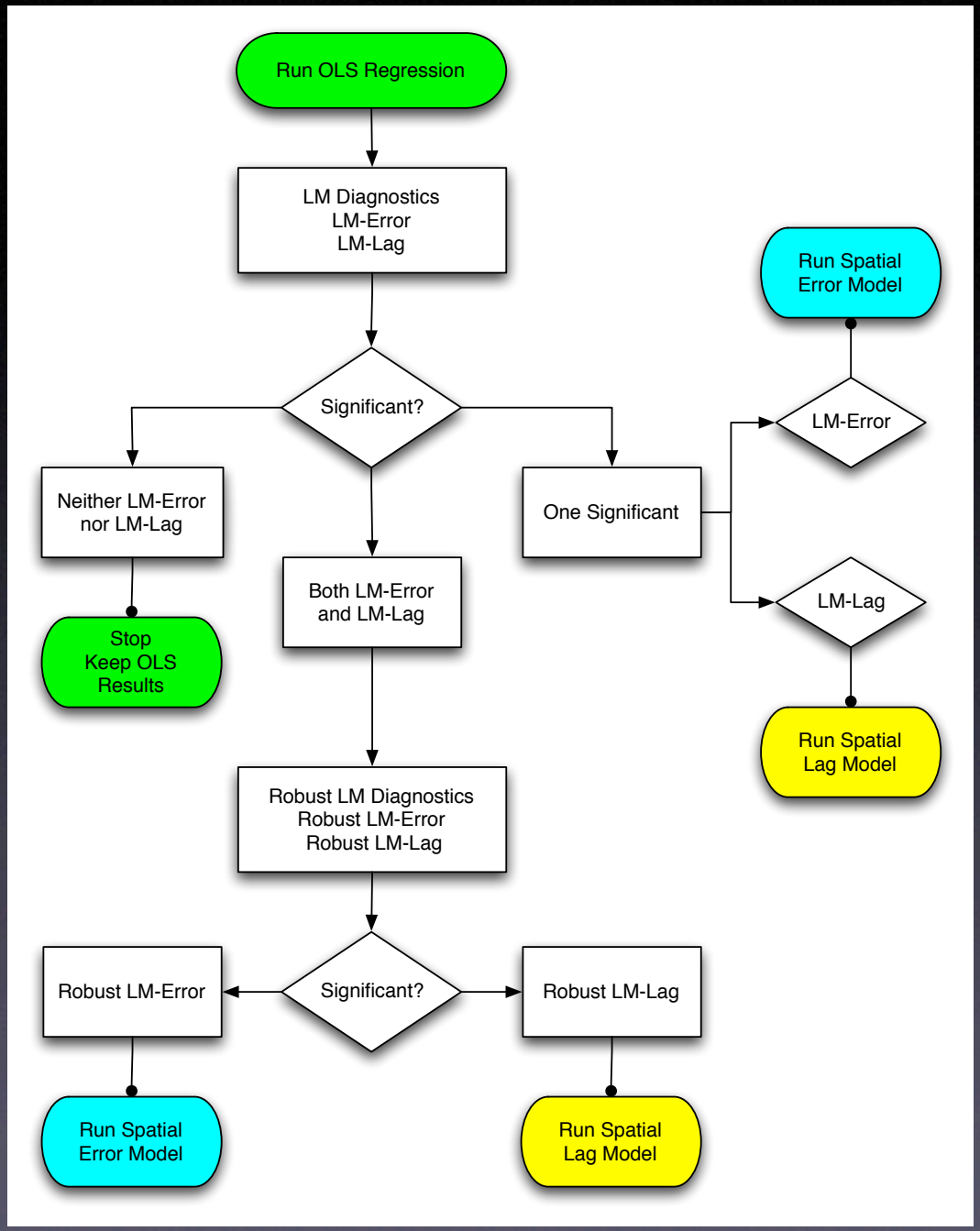
========== END OF REPORT ==========

Specification Search

- Forward Step-Wise Strategy
  - start from constrained model (OLS)
  - base model selection on diagnostics for alternatives
  - advantage: simpler models first
  - problem: pre-testing

Specification Search (2)

- Backward Step-Wise Strategy (Hendry)
  - start from unconstrained model
  - test constraints and proceed to simpler models
  - problem: requires more complex estimation



Example

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : baltrook.GAL (row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.200899	5.8323965	0.0000000
Lagrange Multiplier (lag)	1	20.5454453	0.0000058
Robust LM (lag)	1	1.2626832	0.2611438
Lagrange Multiplier (error)	1	23.6063131	0.0000012
Robust LM (error)	1	4.3235510	0.0375884
Lagrange Multiplier (SARMA)	2	24.8689962	0.0000040

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

- LM-Lag and LM-Error both significant at $p < 0.000$
- Robust LM-Lag not significant $p = 0.26$
- Robust LM-Error significant at $p < 0.05$

⇒ Spatial Error Model is suggested alternative

Estimation

Estimation Problem

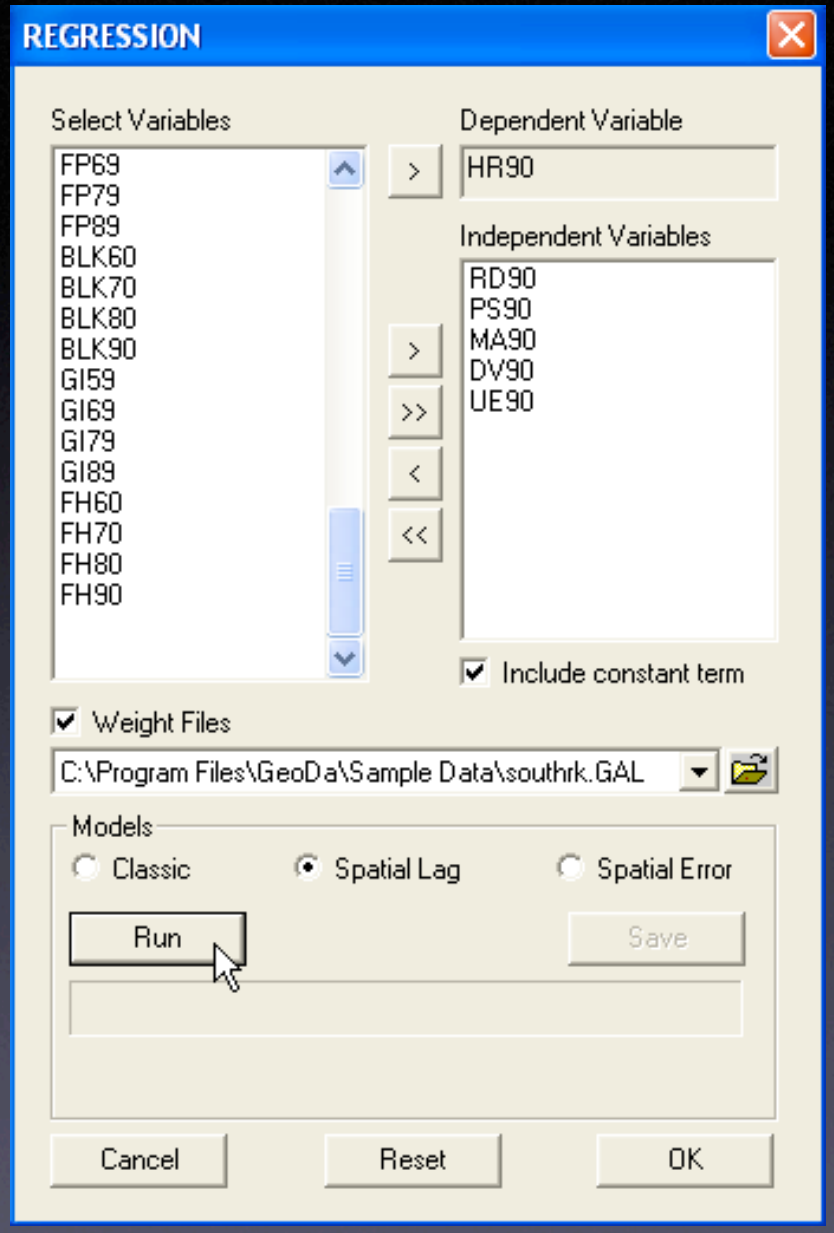
- Spatial Lag Model
  - simultaneity Wy and error term
  - model simultaneity or use instruments
- Spatial Error Model
  - non-spherical error variance matrix
  - Feasible Generalized Least Squares

Maximum Likelihood Estimation

- Likelihood Principle
  - obtain estimator as maximum of the joint density of the "sample"
  - joint density function = likelihood
- Likelihood Equation
  - first order conditions for maximum
  - second order conditions: information matrix ⇒ asymptotic variance

Measures of Fit

- R^2 Not Appropriate
  - based on unweighted sum of squared residuals
  - does not compensate for loss of information
- Likelihood Criteria

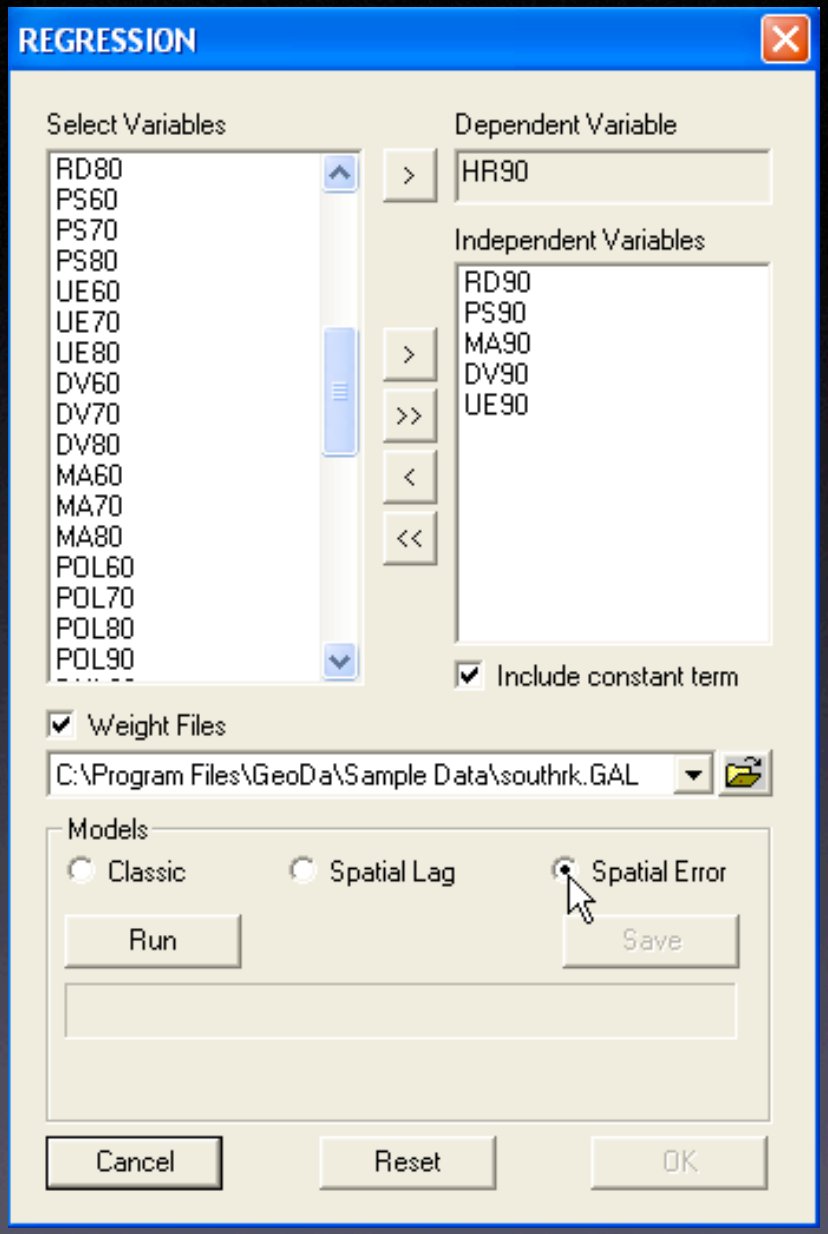


REGRESSION

SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set	: south		
Spatial Weight	: southrk.GAL		
Dependent Variable	: HR90	Number of Observations:	1412
Mean dependent var	: 9.54929	Number of Variables :	7
S.D. dependent var	: 7.03636	Degrees of Freedom :	1405
Lag coeff. (Rho)	: 0.22622		
R-squared	: 0.337480	Log likelihood :	-4474.92
Sq. Correlation	: -	Akaike info criterion :	8963.84
Sigma-square	: 32.8016	Schwarz criterion :	9000.61
S.E of regression	: 5.72727		

Variable	Coefficient	Std.Error	z-value	Probability
W_HR90	0.2262204	0.0335461	6.743568	0.0000000
CONSTANT	5.100989	1.793351	2.84439	0.0044498
RD90	4.030911	0.22555	17.87147	0.0000000
PS90	1.786308	0.2018053	8.851641	0.0000000
MA90	-0.01129424	0.04793461	-0.2356177	0.8137294
DV90	0.4769045	0.1125612	4.236845	0.0000227
UE90	-0.4393495	0.06870696	-6.394542	0.0000000



REGRESSION

SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION

Data set	: **south**		
Spatial Weight	: **southrk.GAL**		
Dependent Variable	: **HR90**	Number of Observations:	1412
Mean dependent var	: 9.549293	Number of Variables	: 6
S.D. dependent var	: 7.036358	Degree of Freedom	: 1406
Lag coeff. (Lambda)	: 0.291609		
R-squared	: 0.345458	R-squared (BUSE)	: -
Sq. Correlation	: -	Log likelihood	:-4471.317119
Sigma-square	: 32.406602	Akaike info criterion	: 8954.63
S.E of regression	: 5.69268	Schwarz criterion	: 8986.150813

Variable	Coefficient	Std.Error	z-value	Probability
CONSTANT	6.693515	1.958045	3.418469	0.0006298
RD90	4.407397	0.237668	18.54434	0.0000000
PS90	1.766328	0.2256524	7.82765	0.0000000
MA90	-0.01663971	0.05298999	-0.3140161	0.7535089
DV90	0.4991464	0.1249123	3.995975	0.0000645
UE90	-0.3878414	0.07847802	-4.942039	0.0000008
LAMBDA	0.2916094	0.03727543	7.823098	0.0000000