空間迴歸模式

Spatial Regression Modeling

授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw

Oral Presentation and Final Exam

- 6.21: Oral Presentation (term project) 10%
  - 主題 **Web-based *SPATIAL*** analytics app for *specific* purposes (using **R Shiny** or any programming languages)
  - 加期末考或作業分數
  - 入選12組口頭報告 (前2-5組期末考+85；前6-12組期末考+75)
  - 1st place: 學期成績 A+ | 2nd and 3rd place: 學期成績 A
  - 所有組別人氣票選：5組 (或 FB: 100讚+10分享+5則不同留言) (期末考+20)
  - 讚+分享+留言 115
  - 精闢票選評論內容：至多10位同學 (期末考+20)
  - 以上加分獎勵可以從缺
- 6.28: Final Exam (Take-home Exam) 20%
  - 形式時間6.27 12:00pm – 6.28 6:00pm (30-hour)

6/21(Thu.) Oral Presentation 相關規定

- 繳交截止日期 6/18 (Mon.) 10:00pm
- 繳交項目 A4 **ONE**-page (pdf file)，包含；
  - 系統網址
  - 功能：說明與畫面截圖 (what: data, functions and results)
  - 動機：目的 (why)；方法 (how)
- 評選標準：空間應用創意、適當分析方法、系統架構完整
- 公告入選12組 6/19 (Tue.) 6:00pm
- 入選12組繳交PPT 6/21 (Thu.) 2:00pm
- 每組口頭簡報 10 min (含系統展示)
- 期末報告成績公告 6/24 (Sun.) 6:00pm

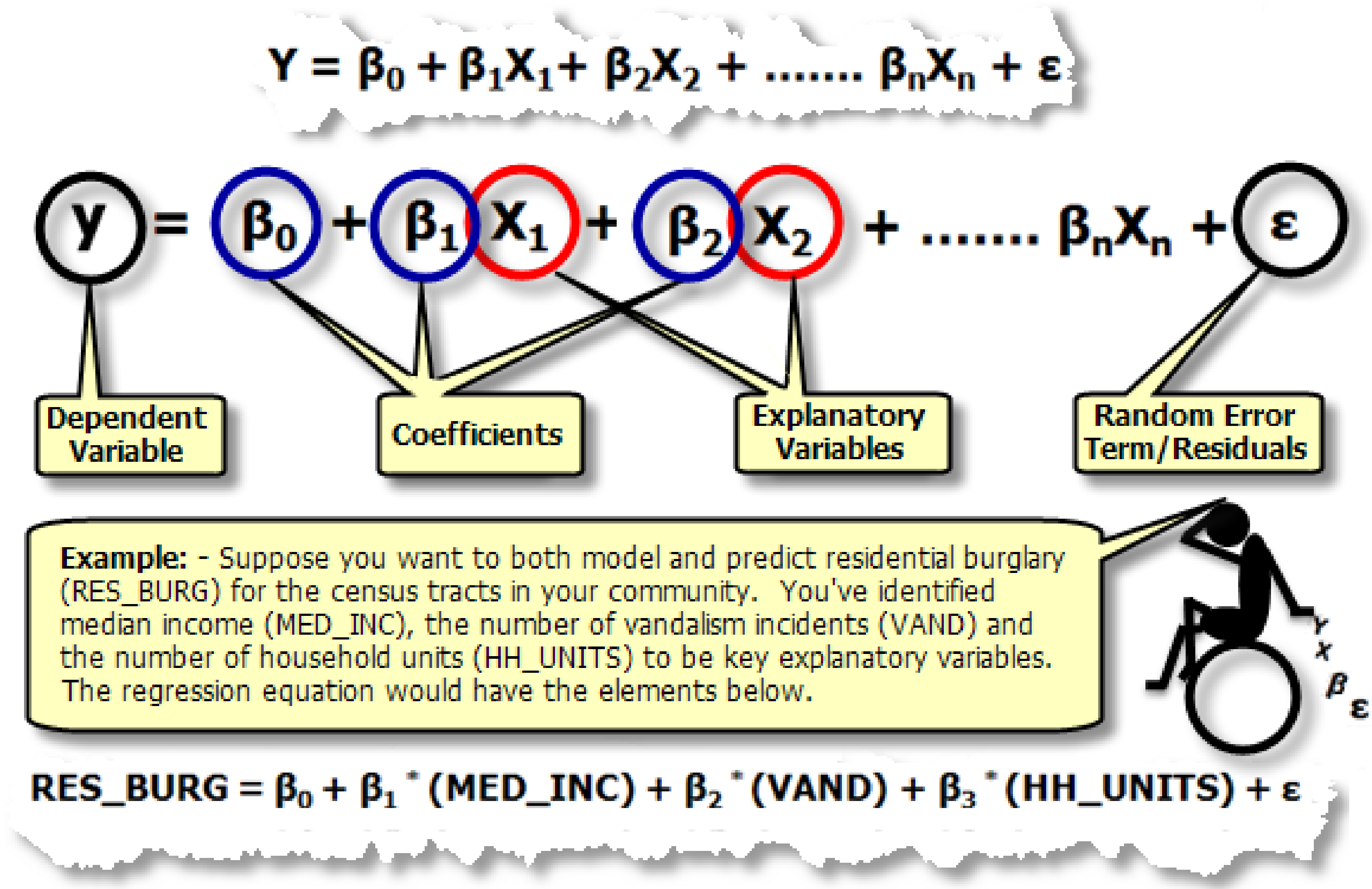
Contents

- OLS regression and residual diagnosis
  - OLS estimation and measurement of model fitness
- Testing spatial autocorrelation for regression residuals
- Spatial regression models
  - Simultaneous Autoregressive Model (SAR)
  - Spatial Lag Model (SLM) and Spatial Error Model (SEM)
  - Spatial Durbin Model (SDM)
- Fitting spatial regression models
  - Maximum Likelihood Estimation (MLE)
- Spatial regression in practice: R code

Basics of Regression

- A statistical method used to examine the relationship between a variable of interest (dependent variable, Y) and **one or more** explanatory variables (predictors, X)
  - Strength of the relationship
  - Direction of the relationship (positive, negative, zero)
  - Goodness of model fit
- Allows you to calculate the amount by which your dependent variable changes when a predictor variable changes by one unit (holding all other predictors constant)
- Often referred to as **Ordinary Least Squares (OLS) regression**
  - Minimize $\Sigma(Y - Y\_hat)^2$
  - Regression with one predictor is called *simple regression*
  - Regression with two or more predictors is called *multiple regression*
- Just like correlation, if an explanatory variable is a significant predictor of the dependent variable, it doesn't imply that the explanatory variable is a ***cause*** of the dependent variable

Elements of an OLS Regression Equation



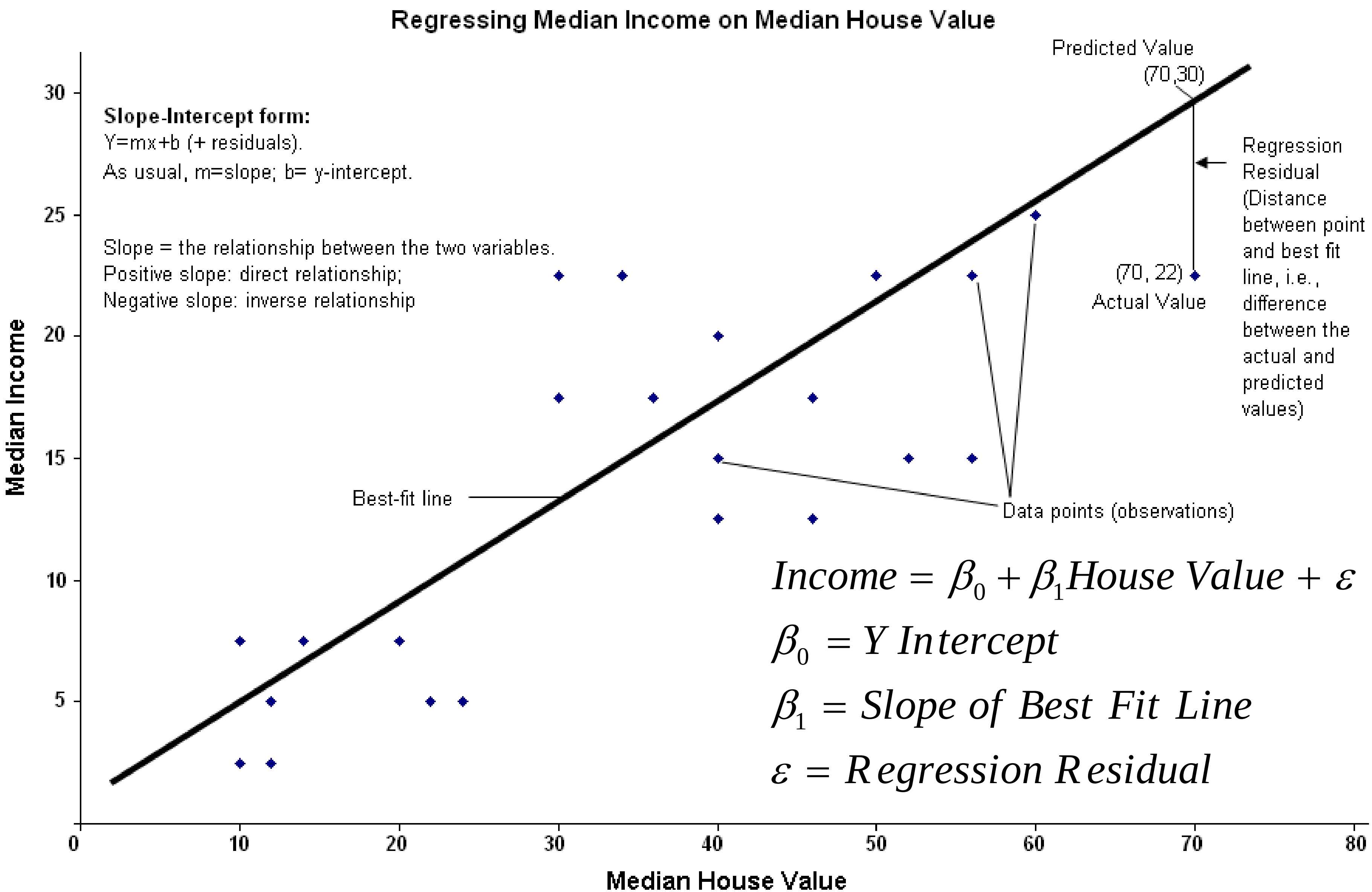
Source: ArcGIS Desktop Help- Regression analysis basics

Terminology of Regression Analysis

- **Dependent variable (y)**
- **Independent/Explanatory variables (X)**
- **Regression coefficients (β)**
  - β_0 is the regression intercept. It represents the expected value for the dependent variable if all of the independent variables are zero.
- **P-Values:** most regression methods perform a statistical test to compute a probability, called a p-value, for the coefficients associated with each independent variable. **The null hypothesis for this statistical test states that a coefficient is not significantly different from zero.**
- **R 2 /R-Squared:** Multiple R-Squared and Adjusted R-Squared are both statistics derived from the regression equation to quantify model performance. The value of R-squared ranges from 0 to 100 %.
- **Residuals:** these are the unexplained portion of the dependent variable, represented in the regression equation as the random error term, ε.

Example

- Assume we have data on median income and median house value in 381 census tracts (unit of measurement)
- Each of the 381 tracts has information on income (call it Y) and on house value (call it X). So, we can create a scatter-plot of Y against X.
  - Through this scatter plot, we can calculate the equation of the line that best fits the pattern (recall: $Y=mx+b$, where m is the slope and b is the y-intercept)
  - This is done by finding a line such that the sum of the *squared* (vertical) *distances* between the points and the line is *minimized*
    - Hence the term ordinary *least squares (OLS)*
- Now, we can examine the relationship between these two variables



$$Income = \beta_0 + \beta_1 House\ Value + \varepsilon$$

$$\beta_0 = Y\ Intercept$$

$$\beta_1 = Slope\ of\ Best\ Fit\ Line$$

$$\varepsilon = Regression\ Residual$$

Extend this to cases with 2+ predictors

- When we have n>1 predictors, rather than getting a line in 2 dimensions, we get a line in n+1 dimensions (the '+1' accounts for the dependent variable)
- Each independent variable will have its own slope coefficient which will indicate the relationship of that particular predictor with the dependent variable, *controlling for all other independent variables in the regression*.
- The equation of the best fit line becomes

$$Income = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon, \quad \text{where}$$

$$\beta_0 = Y\ Intercept$$

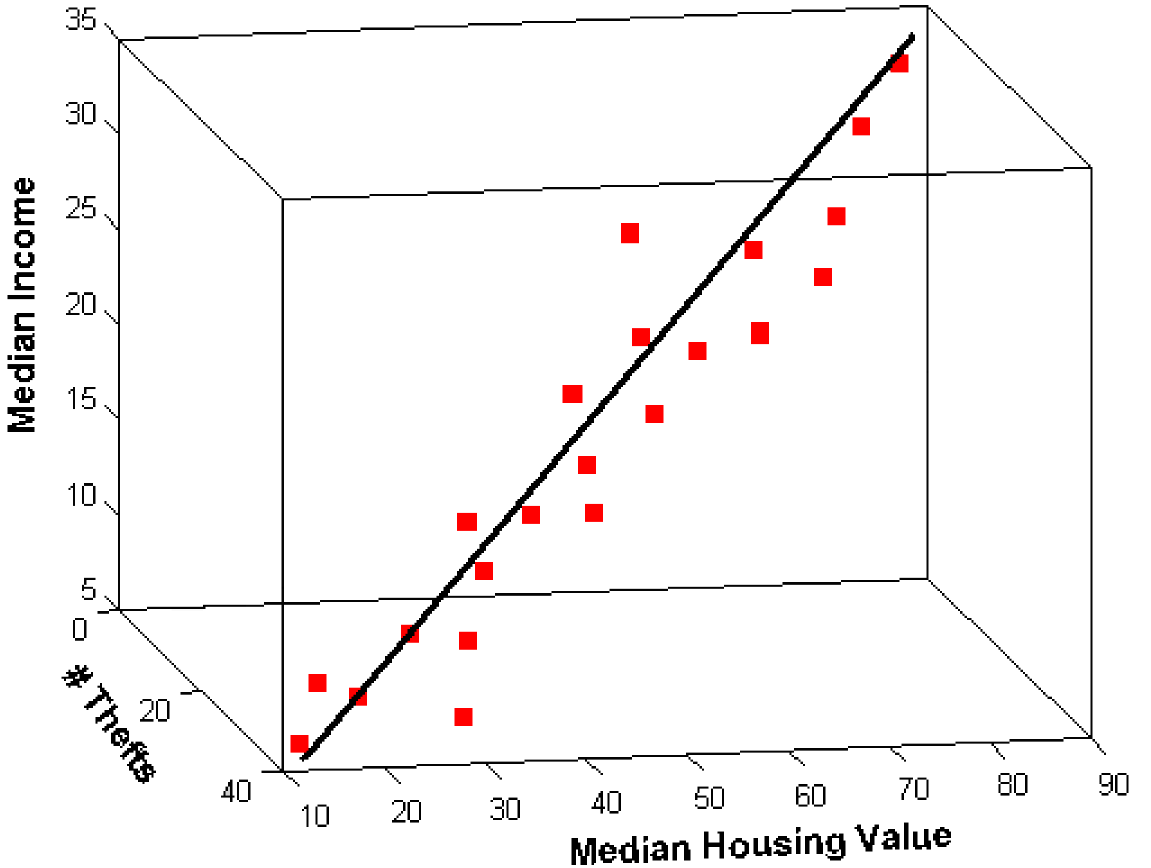
$$\beta_1 .. \beta_n = Coefficients\ of\ Variables\ 1\ ..\ n$$

$$\varepsilon = Residuals\ (should\ be\ random\ noise)$$

- The coefficient β of each predictor may be interpreted as the amount by which the dependent variable changes as the independent variable increases by one unit (holding all other variables constant)

An Example with 2 Predictors: Income as a function of House Value and Crime

$$Income = \beta_0 + \beta_1 HouseValue + \beta_2 Thefts + \varepsilon$$



Some Basic Regression Concepts

- The so-called *p-value* associated with the variable
  - For any statistical method, including regression, we are testing some hypothesis. In regression, we are testing the *null hypothesis* that the coefficient (i.e., slope) *β* is equal to zero (i.e., that the explanatory variable is not a significant predictor of the dependent variable).
  - Formally, the *p-value* is the probability of observing the value of *β* as extreme (i.e., as different from 0 as its estimated value is) when in reality it equals to zero (i.e., when the Null Hypothesis holds). If this probability is small enough (generally, $p<0.05$), we reject the null hypothesis of $\beta=0$ for an *alternative hypothesis* of $\beta<>0$.
    - Again, when the null hypothesis (of $\beta=0$) cannot be rejected, the dependent variable is not related to the independent variable.
    - The rejection of a null hypothesis (i.e., when $p <0.05$) indicates that the independent variable is a statistically significant predictor of the dependent variable
  - One p-value per independent variable

Some Basic Regression Concepts (Cont'd)

- The *sign* of the coefficient of the independent variable (i.e., the slope of the regression line)
  - One coefficient per independent variable
  - Indicates whether the relationship between the dependent and independent variables is positive or negative
  - We should look at the sign when the coefficient is statistically significant (i.e., significantly different from zero)

Some Basic Regression Concepts (Cont'd)

- *R-squared* (Coefficient of Determination): the percent of variance in the dependent variable that is explained by the predictors
  - In the single predictor case, R-squared is simply the square of the correlation between the predictor and dependent variable
  - The more independent variables included, the higher the R-squared
  - Adjusted R-squared: percent of variance in the dependent variable explained, adjusted by the number of predictors
  - One R-squared for the regression model

Common Regression Problems

- Omitted explanatory variables (misspecification)
- Non-linear relationships
- Data outliers
- Multicollinearity
- Non-stationarity
- Inconsistent residual variance (heteroskedasticity)
- Autocorrelated residuals
- Normal distribution bias

Source: ArcGIS Desktop Help- Regression analysis basics

Some (but not all) Regression Assumptions

- The dependent variable should be normally distributed (i.e., the histogram of the variable should look like a bell curve)
  - Ideally, this will also be true of independent variables, but this is not essential. Independent variables can also be binary (i.e., have two values, such as 1 (yes) and 0 (no))
- The predictors should not be strongly correlated with each other (i.e., no multicollinearity)
- Very importantly, the observations should be independent of each other. (The same holds for regression residuals). If this assumption is violated, our coefficient estimates could be wrong!

- General rule of thumb: 10 observations per independent variable

Additional Regression Methods

- Logistic regression/Probit regression
  - When your dependent variable is binary
    - E.g., Employment Indicator (Are you employed? Yes/No)
- Multinomial logistic regression
  - When your dependent variable is categorical and has more than two categories
    - E.g., Race: Black, Asian, White, Other
- Ordinal logistic regression
  - When your dependent variable is ordinal and has more than two categories
    - E.g., Education (1=Less than High School, 2=High School, 3=More than High School)
- Poisson regression
  - When your dependent variable is a count
    - E.g., Number of traffic violations (0, 1, 2, 3, 4, 5, etc)

殘差的重要性與殘差檢定

- 根據高斯-馬可夫定理 (Gauss-Markov Theorem)，只要殘差符合**某些特定的假設**，使用**一般最小平方法(OLS)**來估計迴歸係數時，就可以得到具有「**最佳線性不偏估計量**」**(Best Linear Unbiased Estimator, BLUE)** 的性質。

不偏性 (unbiasedness)
有效性 (efficiency)
一致性 (consistency)

殘差假設的檢定

迴歸模型的殘差 (residual) 必需符合以下的性質：

- 殘差期望值為零 (zero mean) :
  - $E(u) = 0$
- 殘差具同質變異 (homoskedasticity) :
  - $var(u) = \sigma^2$，σ^2 為一固定常數
- 殘差無自我相關 (non-autocorrelation) :
  - cov(ut, ut-s) = 0, for s ≠0
- 自變數與殘差無相關 (orthogonality) :
  - cov(x, u) = 0, for any I
- 殘差為常態性 (normality)

符合以上要求之殘差稱為獨立相同分配 **(independently identical distribution)** 殘差，英文縮寫為 **iid**，用符號表示則為：**u ~ iid (0, σ2)**

OLS Estimation

Linear regression model:
Relationsship between a dependent variable Y and a set of explanatory variables $X_1, X_2, \ldots, X_k$.

(4.1) $\mathbf{y} = \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}$

nx1 vector of the dependent variable: $\mathbf{y} = [y_1 \quad y_2 \quad \cdots \quad y_n]'$

nxk matrix with observations of the k explanatory variables:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{12} & \ldots & x_{1k} \\ 1 & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \ldots & x_{nk} \end{bmatrix}$$

x_{ij}: observation of the j*th* variable at the i*th* statistical unit
1st column of **X**: vector of ones (for intercept)

The explanatory variables are treated as fixed and not random.

kx1 vector of regression coefficients: $\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \ldots \quad \beta_k]'$

nx1 vector of disturbances (error terms): $\boldsymbol{\varepsilon} = [\varepsilon_1 \quad \varepsilon_2 \quad \ldots \quad \varepsilon_n]'$

OLS Estimation (cont'd)

- **Ordinary least squares (OLS) estimation**

An important task of regression analysis is to estimate the unknown vector of regression coefficients, $\boldsymbol{\beta}$, in order to assess the influence of the regressors X_1, X_2, …, X_k on the dependent variable Y. Under the standard assumptions, ordinary least squares (OLS) estimation yields best linear unbiased estimators (blue property).

Least squares criterion:

(4.2a) $$Q(\boldsymbol{\beta}) = \sum_{i=1}^{n} \varepsilon_i^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Q has to be minimized with respect to $\boldsymbol{\beta}$ for which we use the equivalent expression:

(4.2b) $$Q(\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

First order condition for a minimum of Q:

$$\frac{dQ(\boldsymbol{\beta})}{d\boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + 2(\mathbf{X}'\mathbf{X})\boldsymbol{\beta} = \mathbf{o}$$

OLS estimator of $\boldsymbol{\beta}$;

(4.3) $$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Fitted values: (4.4) $\hat{\mathbf{y}} = \mathbf{X} \cdot \hat{\boldsymbol{\beta}}$

Residuals: (4.5a) $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ or (4.5b) $\mathbf{e} = \mathbf{y} - \mathbf{X} \cdot \hat{\boldsymbol{\beta}}$

Residual variance (unbiased estimate of σ^2):

(4.6) $$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^{n} e_i^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} \qquad (\bar{e} = 0)$$

OLS Estimation : Model Fitness

Decomposition of the total sum of squares of the dependent variable Y:

(4.8) SST = SSE + SSR

Total sum of squares:

(4.9) $$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Explained sum of squares:

(4.10) $$SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

Residual sum of squares:

(4.11) $$SSR = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e}$$

Coefficient of determination:

(4.12a) $$R^2 = \frac{SSE}{SST}$$ or (4.12b) $$R^2 = 1 - \frac{SSR}{SST}$$

Range of R^2: $0 \leq R^2 \leq 1$

OLS Parameter Estimation : Model Fitness

Adjusted coefficient of determination:
By the adjustment regression models with different numbers of regressors are made comparable.

(4.13) $$\overline{R}^2 = 1 - \frac{n-1}{n-k}(1-R^2)$$

Information criteria

Information measure the goodness of fit where model complexity in terms of the number of explanatory variables is penalized. Goodness of fit is covered by the log likelihood function ln L,

(4.14) $$\ln L = C - \frac{n}{2}\cdot\ln(\hat{\sigma}^2),$$

which is mainly composed of the sum of squared residuals. By penalizing fits with a larger number of regressors, regression models with different k are made comparable. According to the infor-mation criteria, the model with the lowest value is the best.

- Akaike information criterion (AIC)

(4.15a) $AIC = -2\cdot\ln(L) + 2k$

- Schwartz criterion (SC)

(4.15b) $SC = -2\cdot\ln(L) + k\cdot\ln(n)$

http://www.uni-kassel.de/~rkosfeld/

Why Geospatial: Neighborhood Structure

Regression Assumption:
the observations should be independent of each other.

There are ***NO***
Spatial Associations/ Orientations

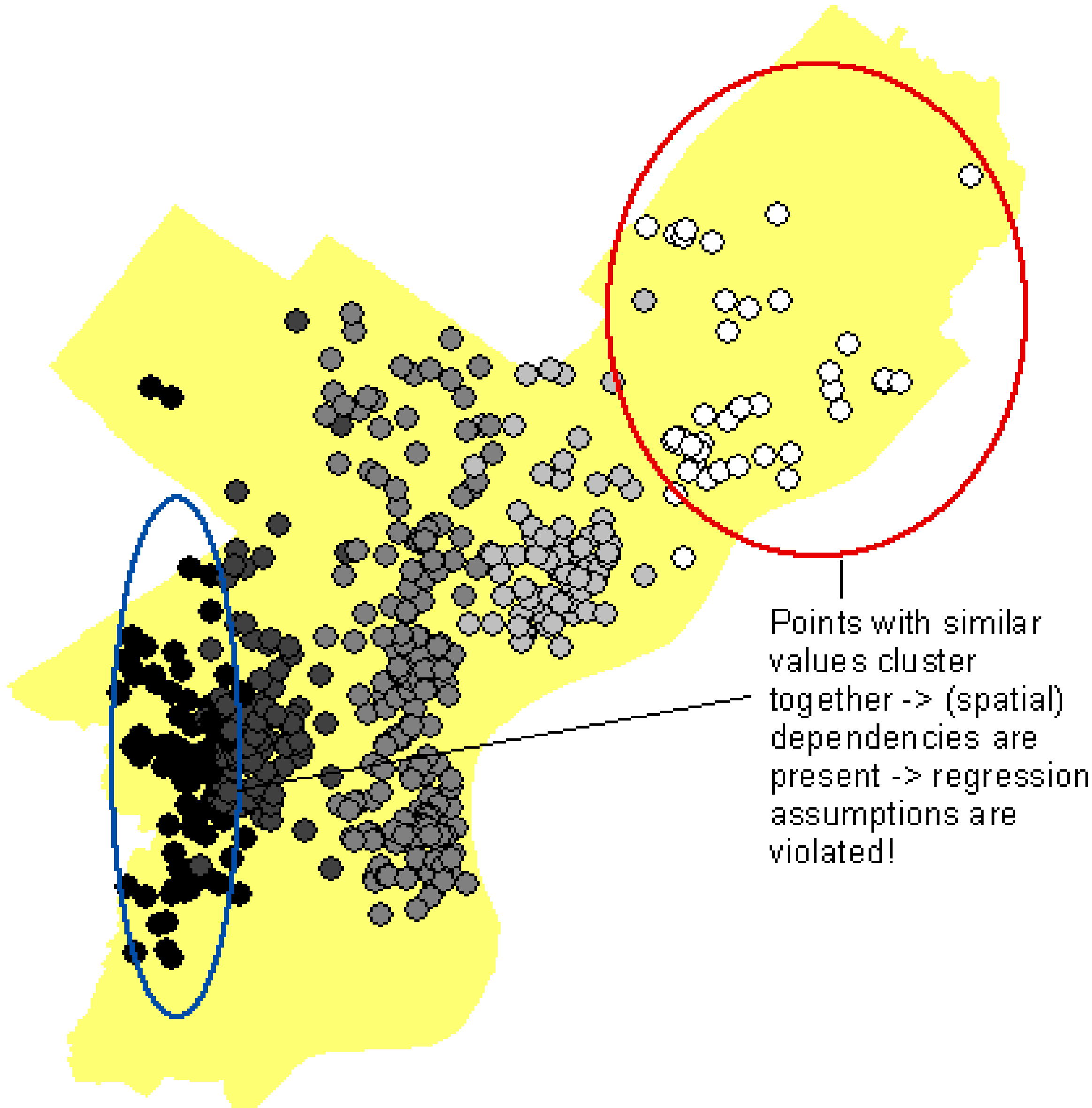
（各鄉鎮市區的人口結構）

縣市	鄉鎮	男性	女性	不識字/自修	國小	國初中	高中職	大專	碩博士	單身	有偶/同居	離婚/分居	喪偶
臺北縣	板橋市	213734	200822	29742	160248	64378	64423	26627	653	235651	166498	4089	8318
臺北縣	三重市	180898	169485	29198	150394	55594	46694	13823	223	203018	135752	3445	8168
臺北縣	永和市	108561	105069	12795	59639	30677	46937	34164	1369	117658	89023	2385	4564
臺北縣	中和市	146888	138477	18358	95658	43810	54030	28182	782	158348	118559	2715	5743
臺北縣	新店市	91671	84992	11795	53631	24710	36151	27058	1081	96421	73846	1953	4443
臺北縣	新莊市	94267	88356	12875	75322	27494	24401	7775	212	103484	74575	1405	3159
臺北縣	樹林市	39101	36599	5875	30780	12143	11172	3122	60	43067	30104	475	2054
臺北縣	鶯歌鎮	24327	22873	4835	20152	7675	5521	1550	22	27356	18063	324	1457
臺北縣	三峽鎮	27621	23761	6807	22862	8917	4731	1206	18	31400	17869	396	1717
臺北縣	淡水鎮	35586	32612	7793	26187	11151	9197	4096	171	38796	26570	608	2224
臺北縣	汐止市	36306	33725	6978	28553	11020	10070	3500	80	39818	27154	696	2363
臺北縣	瑞芳鎮	34683	31661	9853	26317	12036	8835	1923	32	39319	23203	650	3172
臺北縣	土城市	33530	27317	5341	25374	9629	8476	2561	61	34642	24089	606	1510
臺北縣	蘆洲市	25817	23425	4304	22168	7116	5512	1994	28	28522	19053	358	1309
臺北縣	五股鄉	18207	16362	3212	15261	5619	4204	1197	18	20074	13361	320	814
臺北縣	泰山鄉	21504	19890	2833	16297	6324	6059	1748	53	23260	16932	362	840
臺北縣	林口鄉	13014	11643	3458	10002	4104	2761	798	14	14496	9215	247	699
臺北縣	深坑鄉	5101	4559	1133	4038	1658	1219	353	9	5624	3584	80	372
臺北縣	石碇鄉	5020	3994	1413	3854	1762	816	158	0	5393	3113	88	420
臺北縣	坪林鄉	2855	2482	894	2293	983	463	72	1	3185	1893	17	242
臺北縣	三芝鄉	8661	7595	3096	6833	2611	1216	255	8	9154	6306	123	673
臺北縣	石門鄉	5697	4859	2283	4366	1613	731	133	1	6109	3826	114	507
臺北縣	八里鄉	7849	6964	2254	6664	2386	1049	258	8	8759	5335	105	614
臺北縣	平溪鄉	5603	4919	1730	4019	2166	1335	183	1	6431	3506	83	502

Spatial Autocorrelation

- Recall:
  - There is spatial autocorrelation in a variable if observations that are closer to each other in space have related values (Tobler's Law)
  - One of the regression assumptions is independence of observations. If this doesn't hold, we obtain inaccurate estimates of the β coefficients, and the error term ε contains spatial dependencies (i.e., meaningful information), whereas we want the error to not be distinguishable from random noise.

Imagine a problem with a spatial component…



This example is obviously a dramatization, but nonetheless, in many spatial problems points which are close together have similar values

R Sample data: columbus

```
> data(columbus)
> head(columbus)
          AREA PERIMETER COLUMBUS. COLUMBUS.I POLYID NEIG  HOVAL    INC    CRIME     OPEN
1005 0.309441  2.440629         2          5      1    5 80.467 19.531 15.72598 2.850747
1001 0.259329  2.236939         3          1      2    1 44.567 21.232 18.80175 5.296720
1006 0.192468  2.187547         4          6      3    6 26.350 15.956 30.62678 4.534649
1002 0.083841  1.427635         5          2      4    2 33.200  4.477 32.38776 0.394427
1007 0.488888  2.997133         6          7      5    7 23.225 11.252 50.73151 0.405664
1008 0.283079  2.335634         7          8      6    8 28.750 16.029 26.06666 0.563075
         PLUMB DISCBD     X     Y   AREA NSA NSB EW CP THOUS NEIGNO    PERIM
1005 0.217155   5.03 38.80 44.07 10.391   1   1  1  0  1000   1005 2.440629
1001 0.320581   4.27 35.62 42.38  8.621   1   1  0  0  1000   1001 2.236939
1006 0.374404   3.89 39.82 41.18  6.981   1   1  1  0  1000   1006 2.187547
1002 1.186944   3.70 36.50 40.52  2.908   1   1  0  0  1000   1002 1.427635
1007 0.624596   2.83 40.01 38.00 16.827   1   1  1  0  1000   1007 2.997133
1008 0.254130   3.78 43.75 39.28  8.929   1   1  1  0  1000   1008 2.335634
```

R code: OLS Regression

```
columbus.lm<- lm(CRIME ~ INC + HOVAL, data=columbus)
summary(columbus.lm)
```

```
Call:
lm(formula = CRIME ~ INC + HOVAL, data = columbus)

Residuals:
    Min      1Q  Median      3Q     Max 
-34.418  -6.388  -1.580   9.052  28.649 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  68.6190     4.7355  14.490  < 2e-16 ***
INC          -1.5973     0.3341  -4.780 1.83e-05 ***
HOVAL        -0.2739     0.1032  -2.654   0.0109 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.43 on 46 degrees of freedom
Multiple R-squared:  0.5524,	Adjusted R-squared:  0.5329 
F-statistic: 28.39 on 2 and 46 DF,  p-value: 9.341e-09
```

R code: Checking the regression residuals

```
col.listw <- nb2listw(col.gal.nb)

col.moran <-  lm.morantest(columbus.lm,
col.listw, alternative="two.sided")

col.moran
```

```
 Global Moran's I for regression residuals

data:
model: lm(formula = CRIME ~ INC + HOVAL, data = columbus)
weights: col.listw

Moran I statistic standard deviate = 2.681, p-value = 0.00734
alternative hypothesis: two.sided
sample estimates:
Observed Moran's I        Expectation           Variance
       0.212374153       -0.033268284        0.008394853
```

R code: What NOT to do


```
col.e <- resid(columbus.lm)

col.morane <- moran.test(col.e, col.listw,
randomisation=FALSE, alternative="two.sided")
```

```
 Moran's I test under normality

data:  col.e
weights: col.listw

Moran I statistic standard deviate = 2.4774, p-value = 0.01323
alternative hypothesis: two.sided
sample estimates:
Moran I statistic       Expectation          Variance
      0.212374153      -0.020833333       0.008860962
```

Spatial Regression Models

- Simultaneous Autoregressive Model (SAR)
- Spatial Lag Model (SLM), or

  Spatial autoregressive Model (SAR)
- Spatial Error Model (SEM)
- Spatial Durbin Model (SDM)

Simultaneous Autoregressive Model (SAR)

- The SAR specification uses a regression on the values from the other areas to account for the spatial dependence. This means that the error terms *ei* are modelled so that they depend on each other in the following way:

$$e_i = \sum_{i=1}^{m} b_{ij} e_i + \varepsilon_i .$$

residual errors

The *bij* values are used to represent spatial dependence between areas. *bii* must be set to zero so that each area is not regressed on itself.

Simultaneous Autoregressive Model (SAR)

Note that if we express the error terms as $e = B(Y - X^{\mathrm{T}}\beta) + \varepsilon$, the model can also be expressed as

$$Y = X^{\mathrm{T}}\beta + B(Y - X^{\mathrm{T}}\beta) + \varepsilon.$$

Hence, this model can be formulated in a matrix form as follows:

$$(I - B)(Y - X^{\mathrm{T}}\beta) = \varepsilon,$$

where B is a matrix that contains the dependence parameters b_{ij} and I is the identity matrix of the required dimension. It is important to point out that in order for this SAR model to be well defined, the matrix $I - B$ must be non-singular.

$$B = \lambda W$$

λ is a spatial autocorrelation parameter and
W is a matrix that represents spatial dependence

Simultaneous Autoregressive Model (SAR)

$$Y = X^{\mathrm{T}}\beta + B(Y - X^{\mathrm{T}}\beta) + \varepsilon.$$

$$B = \lambda W$$

$$Y = X^{\mathrm{T}}\beta + \lambda W(Y - X^{\mathrm{T}}\beta) + \varepsilon.$$

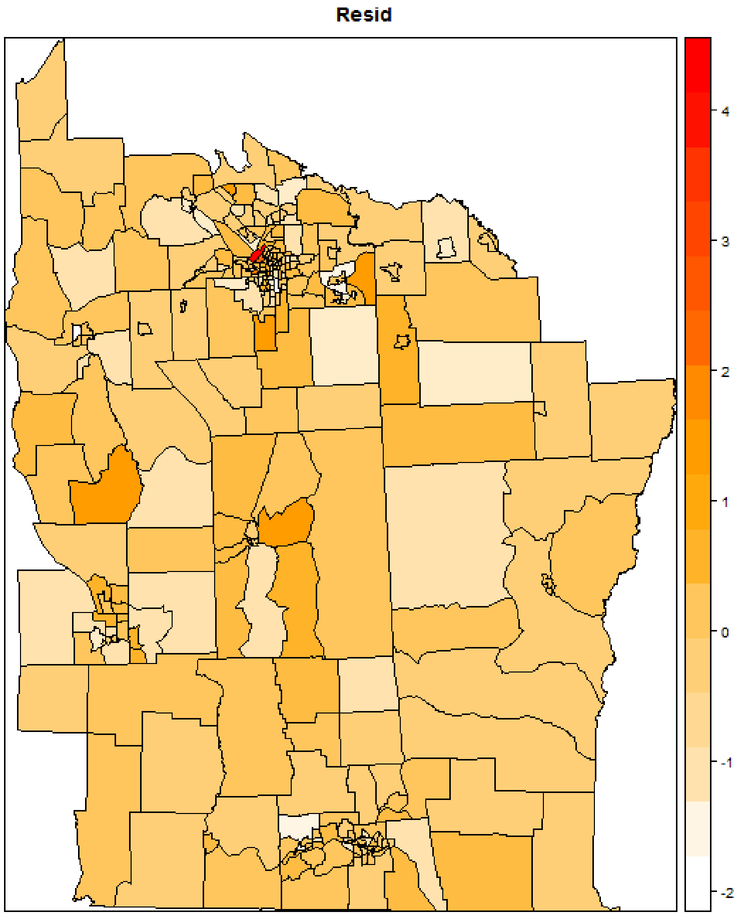
aspatial trend component

spatial stochastic component

R code: Preparation for SAR model

```
# STEP1 Mapping OLS regression residuals.

NY8 <- readOGR("Spatial.R", "NY8_utm18")

nylm <- lm(Z~PEXPOSURE+PCTAGE65P+PCTOWNHOME,
data=NY8)

NY8$lmresid <- residuals(nylm)

lm.palette <- colorRampPalette(c("white","orange",
"red"), space = "rgb")

spplot(NY8, zcol="lmresid", col.regions=lm.palette(20),
main="Resid"
```

R code: Results
STEP1. Mapping OLS regression residuals



R code:
STEP2. Checking the regression residuals

```
> NY_nb <- read.gal("Spatial.R/NY_nb.gal", region.id=row.names(NY8))
> NYlistw<-nb2listw(NY_nb, style = "B")
> lm.morantest(nylm, NYlistw)

 Global Moran's I for regression residuals

data:
model: lm(formula = Z ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME, data = NY8)
weights: NYlistw

Moran I statistic standard deviate = 2.638, p-value = 0.004169
alternative hypothesis: greater
sample estimates:
Observed Moran's I        Expectation           Variance
       0.083090278       -0.009891282        0.001242320
```

R code: STEP3. SAR model

$$Y = X^{\mathrm{T}}\beta + \lambda W(Y - X^{\mathrm{T}}\beta) + \varepsilon.$$

```
nysar<-
spautolm(Z~PEXPOSURE+PCTAGE65P+PCTOWNHOME,
data=NY8, listw=NYlistw)

summary(nysar)
```

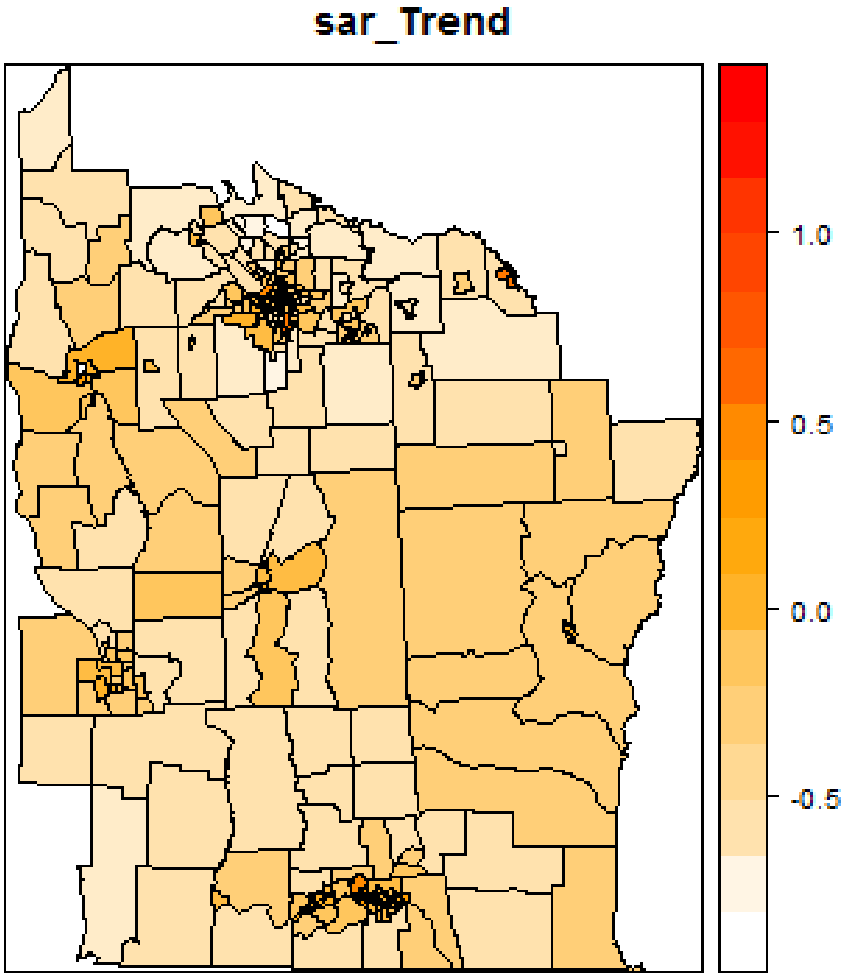
```
Call: spautolm(formula = Z ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME, data = NY8,
    listw = NYlistw)

Residuals:
     Min       1Q   Median       3Q      Max
-1.56754 -0.38239 -0.02643  0.33109  4.01219

Coefficients:
             Estimate Std. Error z value  Pr(>|z|)
(Intercept) -0.618193   0.176784 -3.4969 0.0004707
PEXPOSURE    0.071014   0.042051  1.6888 0.0912635
PCTAGE65P    3.754200   0.624722  6.0094 1.862e-09
PCTOWNHOME  -0.419890   0.191329 -2.1946 0.0281930

Lambda: 0.040487 LR test value: 5.2438 p-value: 0.022026
Numerical Hessian standard error of lambda: 0.017199

Log likelihood: -276.1069
ML residual variance (sigma squared): 0.41388, (sigma: 0.64333)
Number of observations: 281
Number of parameters estimated: 6
AIC: 564.21
```

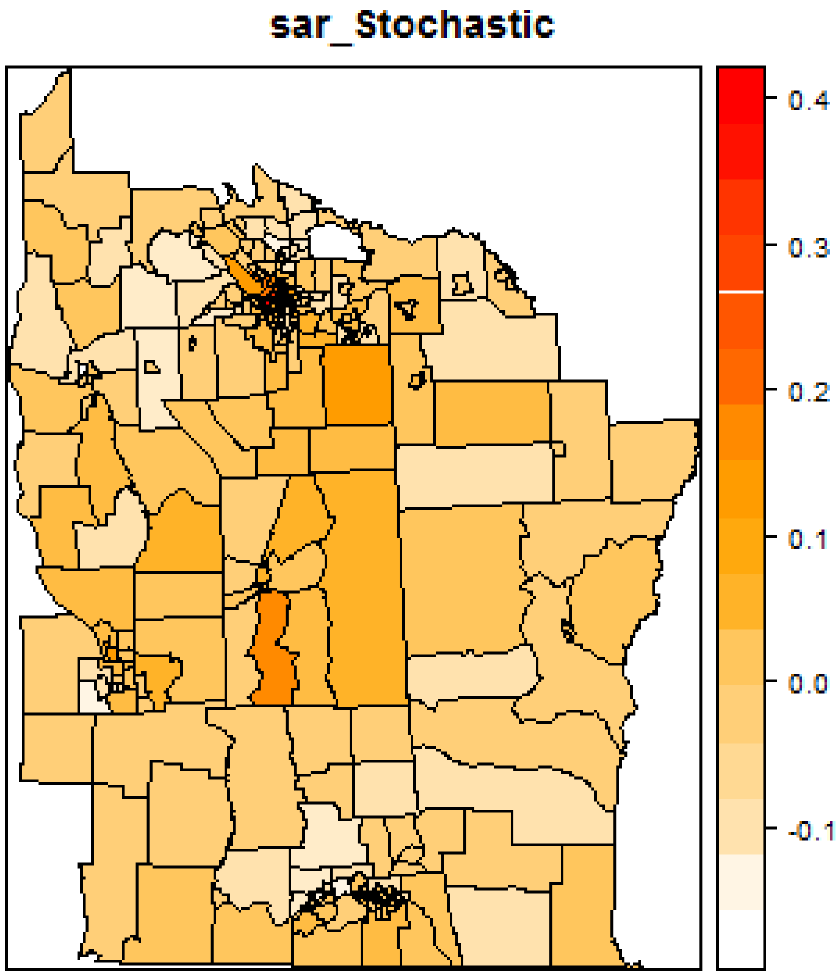
R code: Trend vs. Stochastic

```
NY8$sar_trend <- nysar$fit$signal_trend

NY8$sar_stochastic <- nysar$fit$signal_stochastic

lm.palette <- colorRampPalette(c("white","orange",
"red"), space = "rgb")

spplot(NY8, zcol="sar_trend", col.regions=lm.palette(20),
main="sar_Trend")

spplot(NY8, zcol="sar_stochastic",
col.regions=lm.palette(20), main="sar_Stochastic")
```

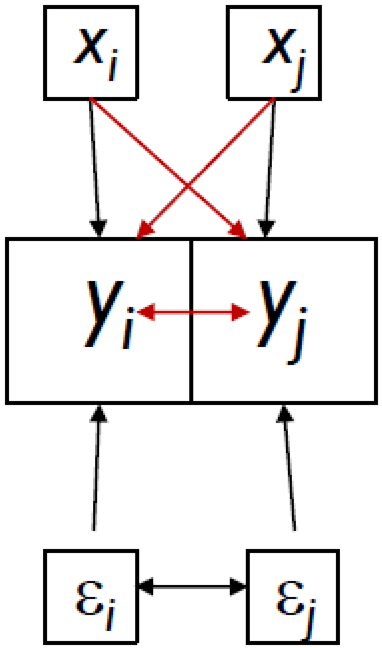
R code: Mapping Trend vs. Stochastic Components

$$Y = X^{\mathrm{T}}\beta + \lambda W(Y - X^{\mathrm{T}}\beta) + \varepsilon.$$





Spatial Lag Model vs. Spatial Error Model



Spatial Lag Model (SLM)

$$y = \boxed{\rho W y} + X\beta + \varepsilon$$

Spatial Error Model (SEM)

$$\mathbf{y} - \lambda \mathbf{W}\mathbf{y} = \mathbf{X}\beta - \lambda \mathbf{W}\mathbf{X}\beta + \mathbf{e},$$

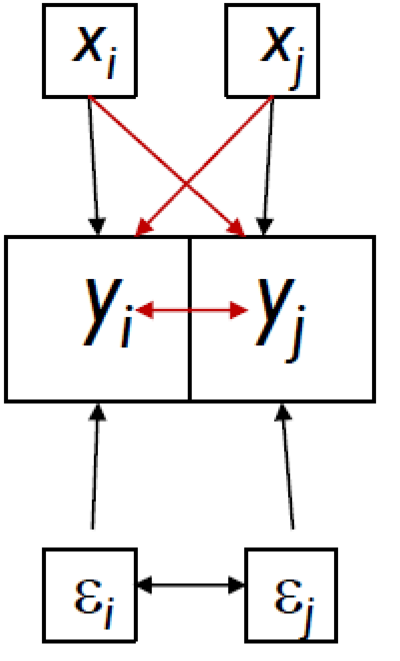
$$(\mathbf{I} - \lambda \mathbf{W})\mathbf{y} = (\mathbf{I} - \lambda \mathbf{W})\mathbf{X}\beta + \mathbf{e},$$

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u},$$

$$\mathbf{u} = \boxed{\lambda \mathbf{W}\mathbf{u}} + \mathbf{e}.$$

Spatial Durbin Model (SDM)

$$\mathbf{y} = \boxed{\rho \mathbf{W} \mathbf{y}} + \mathbf{X}\beta + \boxed{\mathbf{W}\mathbf{X}\gamma} + \mathbf{e},$$



Fitting spatial regression models (e.g. SLM): Maximum Likelihood Estimation, MLE

1st step: Regression of **y** only on **X**

OLS estimator: $\hat{\boldsymbol\beta}_O = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$

Residual vector: $\hat{\mathbf{e}}_O = \mathbf{y} - \mathbf{X}\hat{\boldsymbol\beta}_O$

2nd step: Regression of **Wy** only on **X**

OLS estimator: $\hat{\boldsymbol\beta}_L = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$

Residual vector: $\hat{\mathbf{e}}_L = \mathbf{W}\mathbf{y} - \mathbf{X}\hat{\boldsymbol\beta}_L$

3rd step: Maximisation of the concentrated likelihood function L_C given $\hat{\mathbf{e}}_O$ and $\hat{\mathbf{e}}_L$ to obtain an ML estimator $\hat\rho_{ML}$ for the autoregressive parameter ρ:

$$\max_{\rho} \quad \ln L_c(\rho) = C - \frac{n}{2} \cdot \boxed{\ln[\frac{1}{n}(\mathbf{e}_O - \rho\mathbf{e}_L)'(\mathbf{e}_O - \rho\mathbf{e}_L)]} + \ln|\mathbf{I} - \rho\mathbf{W}|$$

4rth step: ML estimates for β and σ^2

$$\hat{\boldsymbol\beta}_{ML} = \hat{\boldsymbol\beta}_O - \hat\rho_{ML} \cdot \hat{\boldsymbol\beta}_L \quad \text{and} \quad \hat\sigma^2_{ML} = \frac{1}{n}(\mathbf{e}_O - \hat\rho_{ML} \cdot \mathbf{e}_L)'(\mathbf{e}_O - \hat\rho_{ML} \cdot \mathbf{e}_L)$$

http://www.uni-kassel.de/~rkosfeld/

Example

We use the data on output growth (X) and productivity growth (Y) of the 5-region example in order to illustrate maximum likelihood (ML) estimation of the mixed regressive, spatial autoregressive model:

Region	1	2	3	4	5
Output growth (X)	0.6	1.0	1.6	2.6	2.2
Productivity growth (Y)	0.4	0.6	0.9	1.1	1.2

The extended spatial lag model presumes that regional productivity growth is determined by own region's output growth and productivity growth in neighbouring regions

(5.24)
$$y_i = \beta_1 + \beta_2 \cdot x_i + \rho \cdot \sum_{j=1}^{n} w_{ij} \cdot y_j + \varepsilon_i$$

with x_{i1}=1 for all i and $x_{i2} = x_i$. The endogenous spatial lag may capture regional productivity spilllovers suggested by endogenous growth theory.

ML estimation in the extended spatial lag model

1st step: Regression of **y** only on **X** (see section 4.1)

OLS estimator: $$\hat{\boldsymbol\beta}_O = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1.1412 & -0.5882 \\ -0.5882 & 0.3676 \end{bmatrix}\begin{bmatrix} 4.20 \\ 7.78 \end{bmatrix} = \begin{bmatrix} 0.2165 \\ 0.3897 \end{bmatrix}$$

Residual vector: $$\mathbf{e}_O = \mathbf{y} - \mathbf{X}\hat{\boldsymbol\beta}_O = \begin{bmatrix} 0.4 \\ 0.6 \\ 0.9 \\ 1.1 \\ 1.2 \end{bmatrix} - \begin{bmatrix} 0.4503 \\ 0.6062 \\ 0.8400 \\ 1.2297 \\ 1.0738 \end{bmatrix} = \begin{bmatrix} -0.0503 \\ -0.0062 \\ 0.0600 \\ -0.1297 \\ 0.1262 \end{bmatrix}$$

2nd step: Regression of **Wy** only on **X**

$$\underset{5\text{x}1}{\mathbf{Wy}} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0.4 \\ 0.6 \\ 0.9 \\ 1.1 \\ 1.2 \end{bmatrix} = \begin{bmatrix} 0.75 \\ 0.8 \\ 0.7 \\ 0.9 \\ 1.1 \end{bmatrix} \quad \text{and} \quad \underset{5\text{x}2}{\mathbf{X}} = \begin{bmatrix} 1 & 0.6 \\ 1 & 1.0 \\ 1 & 1.6 \\ 1 & 2.6 \\ 1 & 2.2 \end{bmatrix}$$

$$\mathbf{X'Wy} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0.6 & 1.0 & 1.6 & 2.6 & 2.2 \end{bmatrix} \begin{bmatrix} 0.75 \\ 0.8 \\ 0.7 \\ 0.9 \\ 1.1 \end{bmatrix} = \begin{bmatrix} 4.25 \\ 7.13 \end{bmatrix}$$

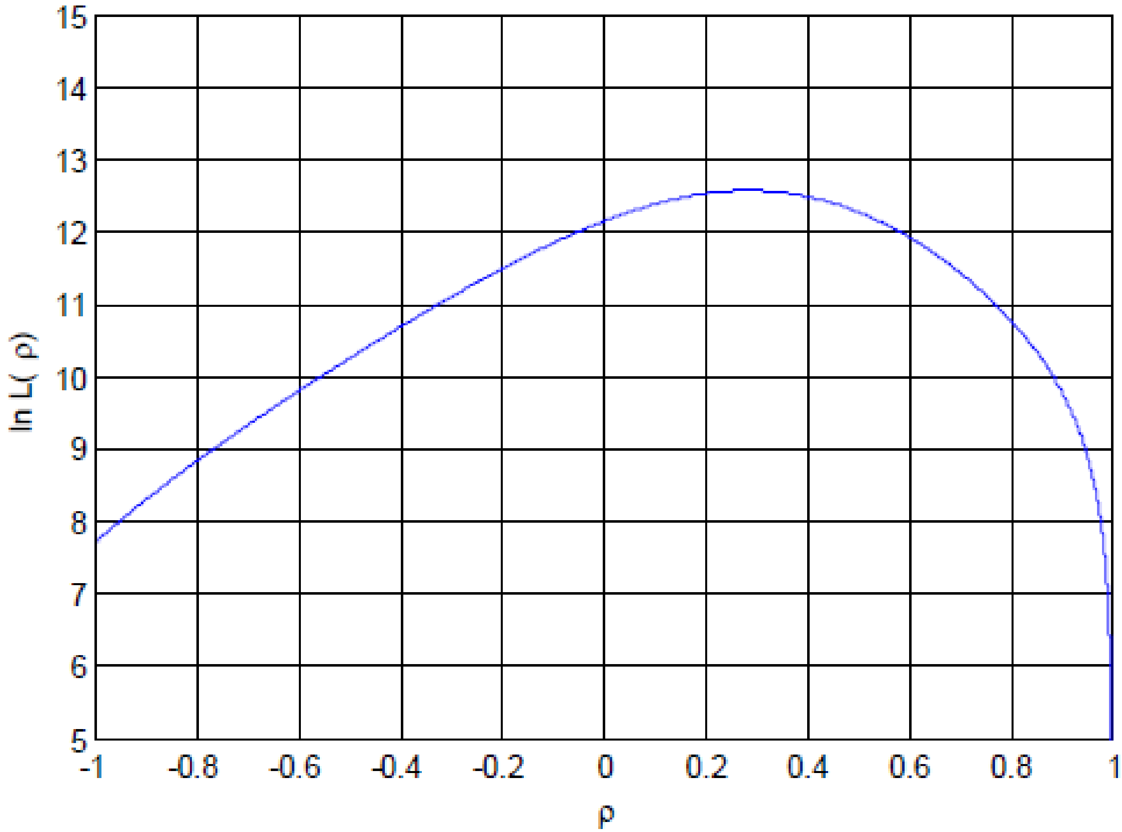
OLS estimator: $\hat{\boldsymbol{\beta}}_L = (\mathbf{X'X})^{-1}\mathbf{X'Wy} = \begin{bmatrix} 1.1412 & -0.5882 \\ -0.5882 & 0.3676 \end{bmatrix} \begin{bmatrix} 4.25 \\ 7.13 \end{bmatrix} = \begin{bmatrix} 0.6562 \\ 0.1211 \end{bmatrix}$

$$\mathbf{X}\hat{\boldsymbol{\beta}}_L = \begin{bmatrix} 1 & 0.6 \\ 1 & 1.0 \\ 1 & 1.6 \\ 1 & 2.6 \\ 1 & 2.2 \end{bmatrix} \begin{bmatrix} 0.6562 \\ 0.1211 \end{bmatrix} = \begin{bmatrix} 0.7289 \\ 0.7773 \\ 0.8500 \\ 0.9711 \\ 0.9226 \end{bmatrix}$$

Residual vector: $\mathbf{e}_L = \mathbf{Wy} - \mathbf{X}\hat{\boldsymbol{\beta}}_L = \begin{bmatrix} 0.75 \\ 0.8 \\ 0.7 \\ 0.9 \\ 1.1 \end{bmatrix} - \begin{bmatrix} 0.7289 \\ 0.7773 \\ 0.8500 \\ 0.9711 \\ 0.9226 \end{bmatrix} = \begin{bmatrix} 0.0211 \\ 0.0227 \\ -0.1500 \\ -0.0711 \\ 0.1774 \end{bmatrix}$

3rd step: Maximisation of the concentrated likelihood function L_C given $\hat{e}_O$ and $\hat{e}_L$ to obtain an ML estimator $\hat{\rho}_{ML}$ for the autoregressive parameter ρ

Figure: Log L_c function of the extended spatial lag model



ML estimator for ρ:

$\hat{\rho}_{ML} = 0.2872$

Log Lc function at $\hat{\rho}_{ML} = 0.2872$:

$\ln L_c^* = 12.5729$

(neglecting the constant C)

4rth step: ML estimates for β and σ^2

ML estimator for β:

$$\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_O - \hat{\rho}_{ML} \cdot \hat{\boldsymbol{\beta}}_L = \begin{bmatrix} 0.2165 \\ 0.3897 \end{bmatrix} - 0.2872 \cdot \begin{bmatrix} 0.6562 \\ 0.1211 \end{bmatrix} = \begin{bmatrix} 0.0280 \\ 0.3549 \end{bmatrix}$$

ML estimator of the error variance:

Error (residual)

$$\mathbf{e}_{ML} = \mathbf{e}_O - \hat{\rho}_{ML} \cdot \mathbf{e}_L = \begin{bmatrix} -0.0503 \\ -0.0062 \\ 0.0600 \\ -0.1297 \\ 0.1262 \end{bmatrix} - 0.2872 \cdot \begin{bmatrix} 0.0211 \\ 0.0227 \\ -0.1500 \\ -0.0711 \\ 0.1774 \end{bmatrix} = \begin{bmatrix} -0.0564 \\ -0.0127 \\ 0.1031 \\ -0.1093 \\ 0.0753 \end{bmatrix}$$

Error Variance

$$\hat{\sigma}^2_{ML} = \frac{1}{n}(\mathbf{e}_O - \hat{\rho}_{ML} \cdot \mathbf{e}_L)'(\mathbf{e}_O - \hat{\rho}_{ML} \cdot \mathbf{e}_L)$$

$$= \frac{1}{5} \cdot \begin{bmatrix} -0.0564 & -0.0127 & 0.1031 & -0.1093 & 0.0753 \end{bmatrix} \begin{bmatrix} -0.0564 \\ -0.0127 \\ 0.1031 \\ -0.1093 \\ 0.0753 \end{bmatrix}$$

$$= \frac{1}{5} \cdot 0.0316 = 0.0063$$

http://www.uni-kassel.de/~rkosfeld/

R code: Spatial Lag Mode, SLM

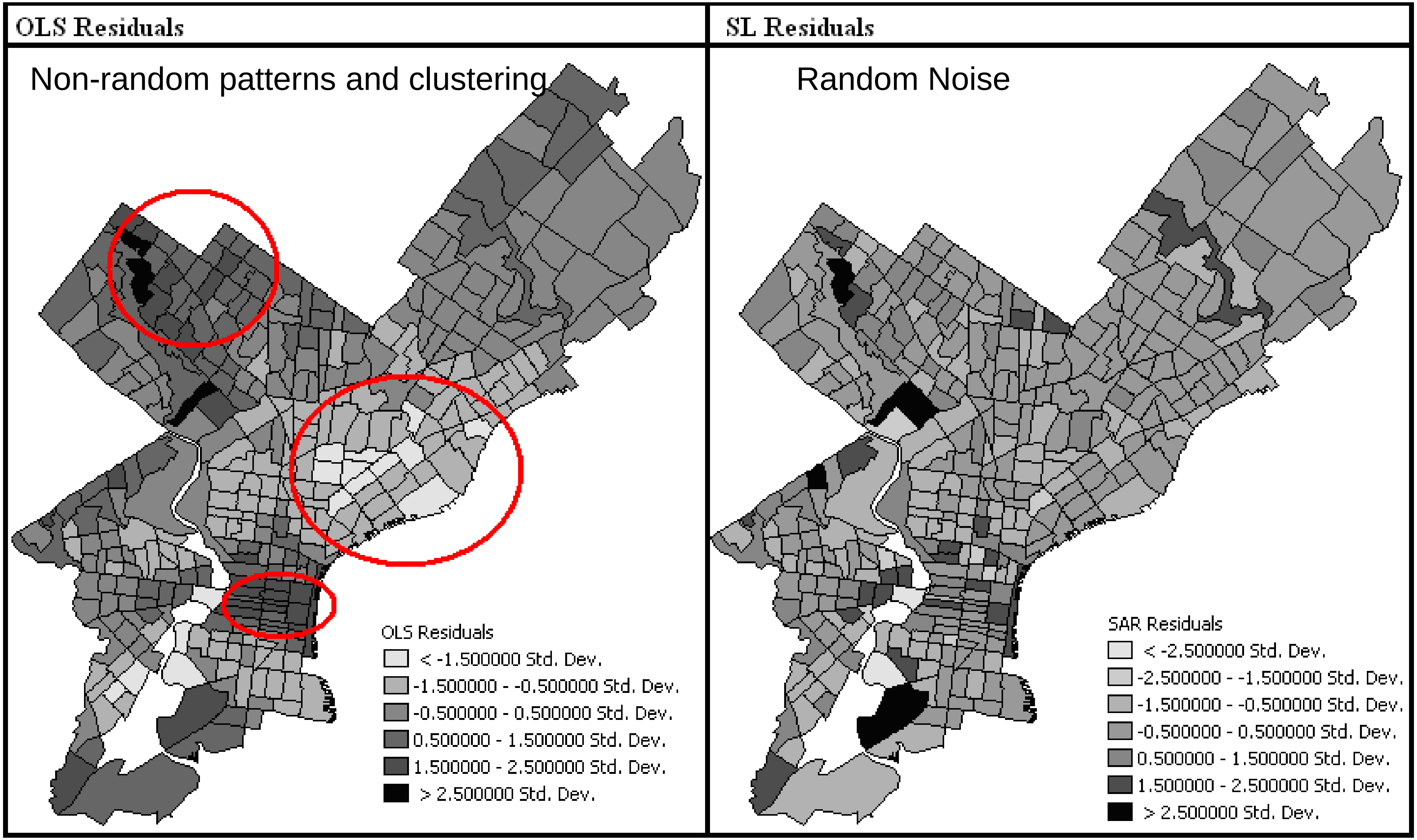
$$y = \rho W y + X\beta + \varepsilon$$

```
# row-standardized matrix

NYlistwW <- nb2listw(NY_nb, style = "W")


nylag <-
lagsarlm(Z~PEXPOSURE+PCTAGE65P+PCTOWNHOME,
data=NY8, listw=NYlistwW)


summary(nylag)
```

R code: Spatial Lag Mode, SLM

```
Call:lagsarlm(formula = Z ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME, data = NY8,
    listw = NYlistwW)

Residuals:
      Min        1Q    Median        3Q       Max
-1.626029 -0.393321 -0.018767  0.326616  4.058315

Type: lag
Coefficients: (asymptotic standard errors)
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.505343   0.155850 -3.2425 0.001185
PEXPOSURE    0.045543   0.034433  1.3227 0.185943
PCTAGE65P    3.650055   0.599219  6.0914 1.12e-09
PCTOWNHOME  -0.411829   0.169095 -2.4355 0.014872

Rho: 0.22518, LR test value: 7.7503, p-value: 0.0053703
Asymptotic standard error: 0.079538
    z-value: 2.8312, p-value: 0.0046378
Wald statistic: 8.0155, p-value: 0.0046378

Log likelihood: -274.8536 for lag model
ML residual variance (sigma squared): 0.40998, (sigma: 0.64029)
Number of observations: 281
Number of parameters estimated: 6
AIC: 561.71, (AIC for lm: 567.46)
LM test for residual autocorrelation
test value: 0.6627, p-value: 0.41561
```

Example: OLS Residuals vs. SL Residuals



R code: Spatial Error Model, SEM

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u},$$

$$\mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \mathbf{e}.$$

```
nyerr <-
errorsarlm(Z~PEXPOSURE+PCTAGE65P+PCTOWNHOME,
data=NY8, listw=NYlistwW)


summary(nyerr)
```

R code: Spatial Error Model, SEM

```
Call:errorsarlm(formula = Z ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME,
    data = NY8, listw = NYlistwW)

Residuals:
      Min        1Q    Median        3Q       Max
-1.628589 -0.384745 -0.030234  0.324747  4.047906

Type: error
Coefficients: (asymptotic standard errors)
            Estimate Std. Error z value  Pr(>|z|)
(Intercept) -0.58662    0.17471 -3.3577  0.000786
PEXPOSURE    0.05933    0.04226  1.4039  0.160335
PCTAGE65P    3.83746    0.62345  6.1552 7.496e-10
PCTOWNHOME  -0.44428    0.18897 -2.3510  0.018721

Lambda: 0.21693, LR test value: 5.4248, p-value: 0.019853
Asymptotic standard error: 0.085044
    z-value: 2.5507, p-value: 0.010749
Wald statistic: 6.5063, p-value: 0.010749

Log likelihood: -276.0164 for error model
ML residual variance (sigma squared): 0.41369, (sigma: 0.64319)
Number of observations: 281
Number of parameters estimated: 6
AIC: 564.03, (AIC for lm: 567.46)
```

Spatial Error Model =
Simultaneous Autoregressive Model (SAR)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u},$$

$$\mathbf{u} = \lambda \mathbf{W}\mathbf{u} + \mathbf{e}.$$

$$Y = X^{\mathrm{T}}\beta + \lambda W(Y - X^{\mathrm{T}}\beta) + \varepsilon.$$

errorsarlm()

spautolm()

Compare errorsarlm() and spautolm()

```
library(rgdal);library (spdep)
data(columbus)

columbus.err <- errorsarlm(CRIME ~ INC + HOVAL,data=columbus,col.listw)
summary(columbus.err)

columbus.sar<-spautolm(CRIME ~ INC + HOVAL, data=columbus, listw=col.listw)
summary(columbus.sar)
```

```
> columbus.err <- errorsarlm(CRIME ~ INC + HOVAL,data=columbus,col.listw)
> summary(columbus.err)

Call:errorsarlm(formula = CRIME ~ INC + HOVAL, data = columbus, listw = col.listw)

Residuals:
      Min        1Q    Median        3Q       Max
-34.45950  -6.21730  -0.69775   7.65256  24.23631

Type: error
Coefficients: (asymptotic standard errors)
             Estimate Std. Error z value  Pr(>|z|)
(Intercept) 61.053618   5.314875 11.4873 < 2.2e-16
INC         -0.995473   0.337025 -2.9537 0.0031398
HOVAL       -0.307979   0.092584 -3.3265 0.0008794

Lambda: 0.52089, LR test value: 6.4441, p-value: 0.011132
```

```
> summary(columbus.sar)

Call: spautolm(formula = CRIME ~ INC + HOVAL, data = columbus, listw = col.listw)

Residuals:
      Min        1Q    Median        3Q       Max
-34.45950  -6.21730  -0.69775   7.65256  24.23631

Coefficients:
             Estimate Std. Error z value  Pr(>|z|)
(Intercept) 61.053618   5.314875 11.4873 < 2.2e-16
INC         -0.995473   0.337025 -2.9537 0.0031398
HOVAL       -0.307979   0.092584 -3.3265 0.0008794

Lambda: 0.52089 LR test value: 6.4441 p-value: 0.011132
Numerical Hessian standard error of lambda: 0.1638
```

R code: Spatial Durbin Mode, SDM

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \beta + \mathbf{W} \mathbf{X} \gamma + \mathbf{e},$$

```
nymix <-
lagsarlm(Z~PEXPOSURE+PCTAGE65P+PCTOWNHOME,
data=NY8, listw=NYlistwW, type="mixed“ )


summary(nymix)
```

R code: Spatial Durbin Mode, SDM

```
Call:lagsarlm(formula = Z ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME, data = NY8,
    listw = NYlistwW, type = "mixed")

Residuals:
      Min        1Q    Median        3Q       Max
-1.632286 -0.400142  0.011403  0.325858  4.056743

Type: mixed
Coefficients: (asymptotic standard errors)
                Estimate Std. Error z value  Pr(>|z|)
(Intercept)    -0.322597   0.235599 -1.3693   0.17092
PEXPOSURE       0.090394   0.116765  0.7742   0.43884
PCTAGE65P       3.613563   0.657768  5.4937 3.937e-08
PCTOWNHOME     -0.026866   0.252887 -0.1062   0.91539
lag.PEXPOSURE  -0.051880   0.127429 -0.4071   0.68391
lag.PCTAGE65P   0.131232   1.208395  0.1086   0.91352
lag.PCTOWNHOME -0.699499   0.334331 -2.0922   0.03642

Rho: 0.17578, LR test value: 3.6967, p-value: 0.054521
Asymptotic standard error: 0.086624
    z-value: 2.0293, p-value: 0.042431
Wald statistic: 4.1179, p-value: 0.042431

Log likelihood: -272.6698 for mixed model
ML residual variance (sigma squared): 0.40527, (sigma: 0.63661)
Number of observations: 281
Number of parameters estimated: 9
AIC: 563.34, (AIC for lm: 565.04)
LM test for residual autocorrelation
test value: 1.0337, p-value: 0.30929
```

鄰近y的效果

Comparing SLM and SDM

$$y = \rho W y + X\beta + \varepsilon,$$

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\gamma + \mathbf{e},$$

```
> anova(nymix, nylag)
      Model df    AIC  logLik Test L.Ratio p-value
nymix     1  9 563.34 -272.67    1
nylag     2  6 561.71 -274.85    2  4.3678 0.22439
```

Model Evaluation (spatial lag or error?)

- Data-driven approach
  - tests for lack of spatial effects after fitting a spatial lag/error model
- Theory-based approach
  - based on substantive grounds

Model Fitting

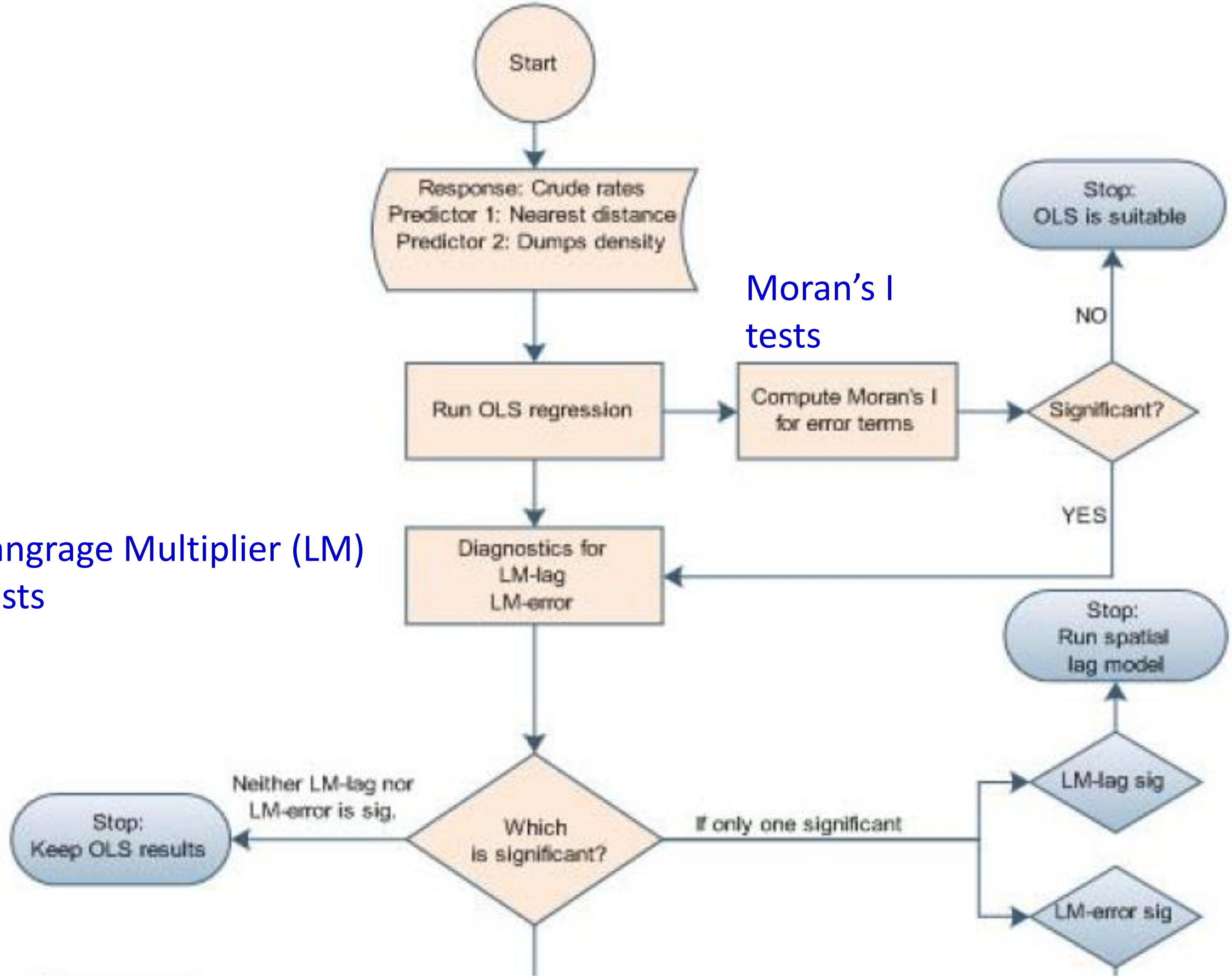
- Akaike's Information Criterion (AIC) and Schwartz's Bayesian Information Criterion (BIC) are often used, which measure the fit of the model to the data.
- Models having a smaller AIC or a smaller BIC are considered the better models in the sense of model fitting balanced with model parsimony.

Data-driven approach

- Moran's I for OLS residuals
- Langrage Multiplier (LM) tests
- Robust Langrage Multiplier (LM) tests

OLS residuals

$$LM(Lag) = \frac{(e'Wy / s^2)^2}{RJ_{\rho-\alpha}} \qquad LM(error) = \frac{(e'We / s^2)}{T}.$$

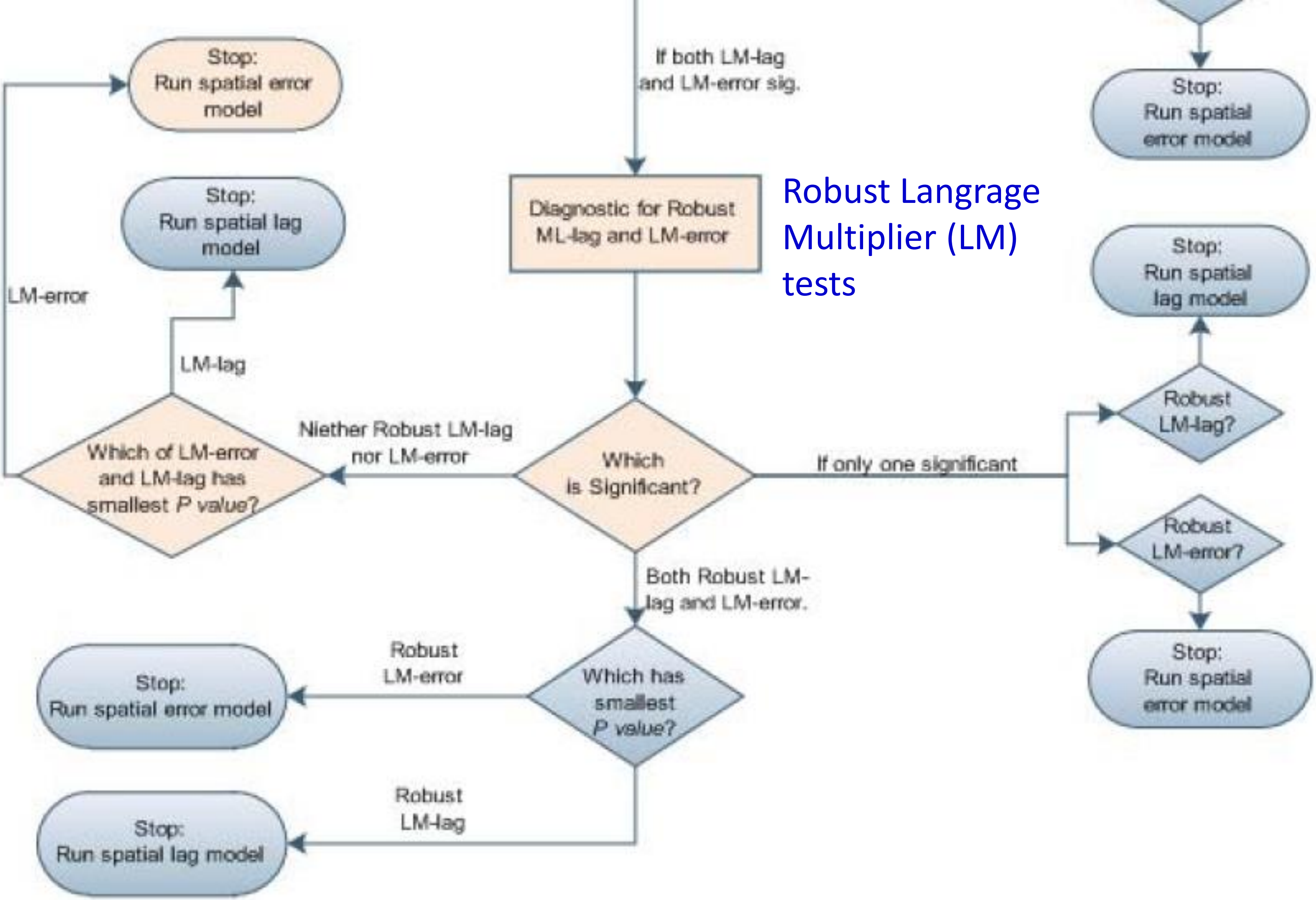


Moran's I tests

Langrage Multiplier (LM) tests

Source : International Journal of Health Geographics 2008, **7**:62

Robust Langrage Multiplier (LM) tests



Source : *International Journal of Health Geographics* 2008, **7**:62

R code: Langrage Multiplier (LM) tests

```
NYlistwW <- nb2listw(NY_nb, style = "W")
res <- lm.Lmtests(nylm, listw=NYlistwW, test="all“ )
summary(res)
```

```
> summary(res)
 Lagrange multiplier diagnostics for spatial dependence
data:
model: lm(formula = Z ~ PEXPOSURE + PCTAGE65P + PCTOWNHOME, data = NY8)
weights: NYlistwW

       statistic parameter  p.value
LMerr     5.1674         1 0.023015 *
LMlag     8.5430         1 0.003468 **
RLMerr    1.6789         1 0.195068
RLMlag    5.0546         1 0.024561 *
SARMA    10.2220         2 0.006030 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Example: Modeling Global Democratization

OLS Regression: $$\text{Polity score} = \beta_0 + \beta_1 \ln \text{GDP per capita} + \epsilon.$$

Country	Democracy	GDP	Country	Democracy	GDP
Guinea	−1	51	Iran	3	1776
Ethiopia	1	114	Macedonia	6	1801
Burundi	0	120	Namibia	6	1870
Zaire	0	135	Romania	8	1941
Sierra Leone	−10	172	Algeria	−3	2036
Eritrea	−7	175	Bosnia & Herzegovina	0	2108
Malawi	5	178	Thailand	9	2215
Iraq	−9	181	Suriname	9	2224
Guinea−Bissau	5	187	Guatemala	8	2257
Liberia	0	194	Russia	7	2279
Rwanda	−4	216	Ecuador	6	2305
Mozambique	6	217	Peru	9	2306
Tajikistan	−1	221	Colombia	7	2342
Niger	4	247	Jordan	−2	2375
Nepal	6	276	Fiji	5	2397
Burkina Faso	0	315	Tunisia	−4	2436
Chad	−2	317	El Salvador	7	2486
Uganda	−4	320	South Africa	9	2607
Tanzania	2	330	Dominican Republic	8	2745
C. African Rep.	5	333	Cuba	−7	2891
⋮	⋮	⋮	⋮	⋮	⋮
Turkmenistan	−9	1241	Canada	10	25139
Morocco	−6	1300	Finland	10	26235
Congo	−5	1303	Austria	10	26304
Djibouti	2	1313	Netherlands	10	27059
Byelarus	−7	1359	Sweden	10	27497
North Korea	−9	1361	United Kingdom	10	27650
Swaziland	−9	1412	Japan	10	31731
Albania	5	1416	United Arab Emirates	−8	34436
Syria	−7	1417	Qatar	−10	36611
Kazakhstan	−6	1437	Denmark	10	37063
Serbia	7	1573	Switzerland	10	39769
Egypt	−6	1602	United States	10	40180
Myanmar (Burma)	−7	1729	Norway	10	43895
Bulgaria	9	1744	Luxembourg	10	54255

Source:
Ward (2008). Spatial Regression Models, Sage Publications, Inc

OLS Results

$$\text{Polity score} = \beta_0 + \beta_1 \ln \text{GDP per capita} + \epsilon.$$

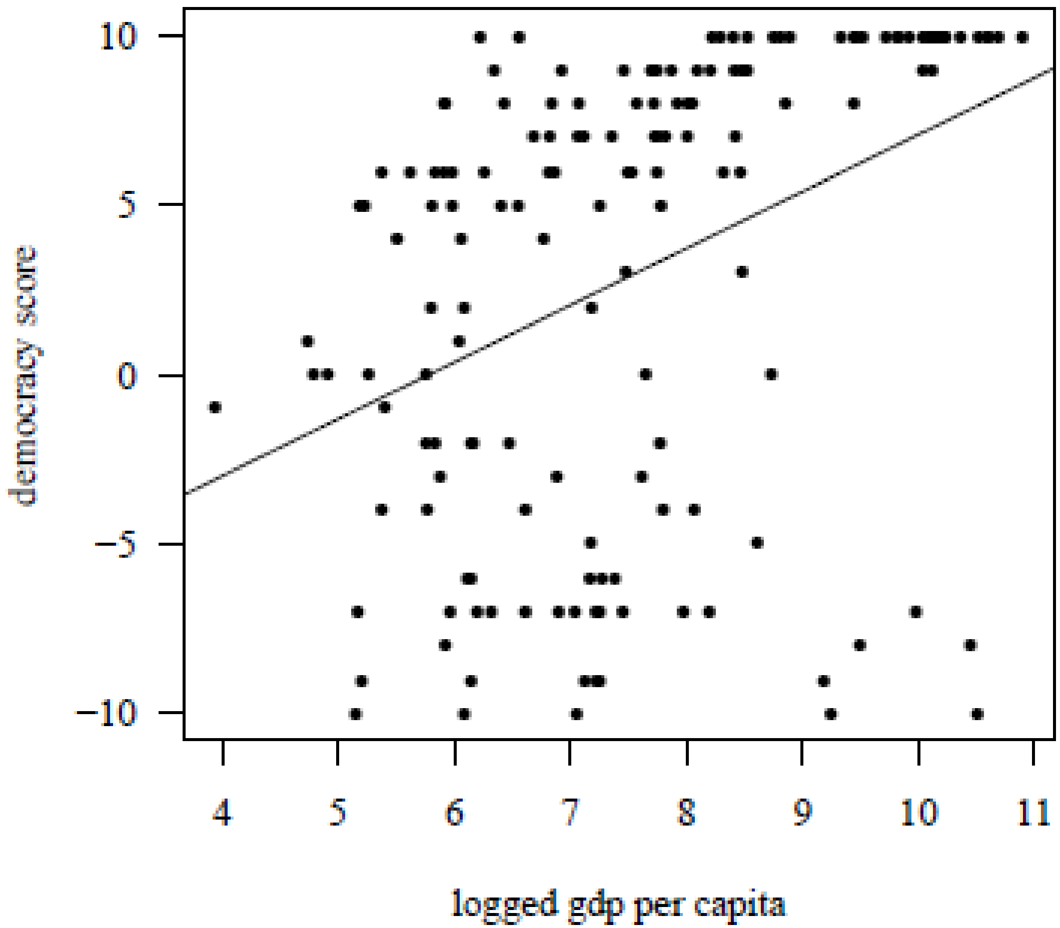
TABLE 1.2. *OLS Estimates of democracy as a linear function of logged GDP per capita, using 2002 data from the Polity project and the World Bank.*

	$\hat{\beta}$	SE($\hat{\beta}$)	t-value
Intercept	−9.69	2.43	−3.99
Ln GDP per capita	1.69	0.31	5.36

N = 158
Log likelihood (df=3) = -513.62
$F = 28.77$ ($df_1 = 1$, $df_2 = 156$)

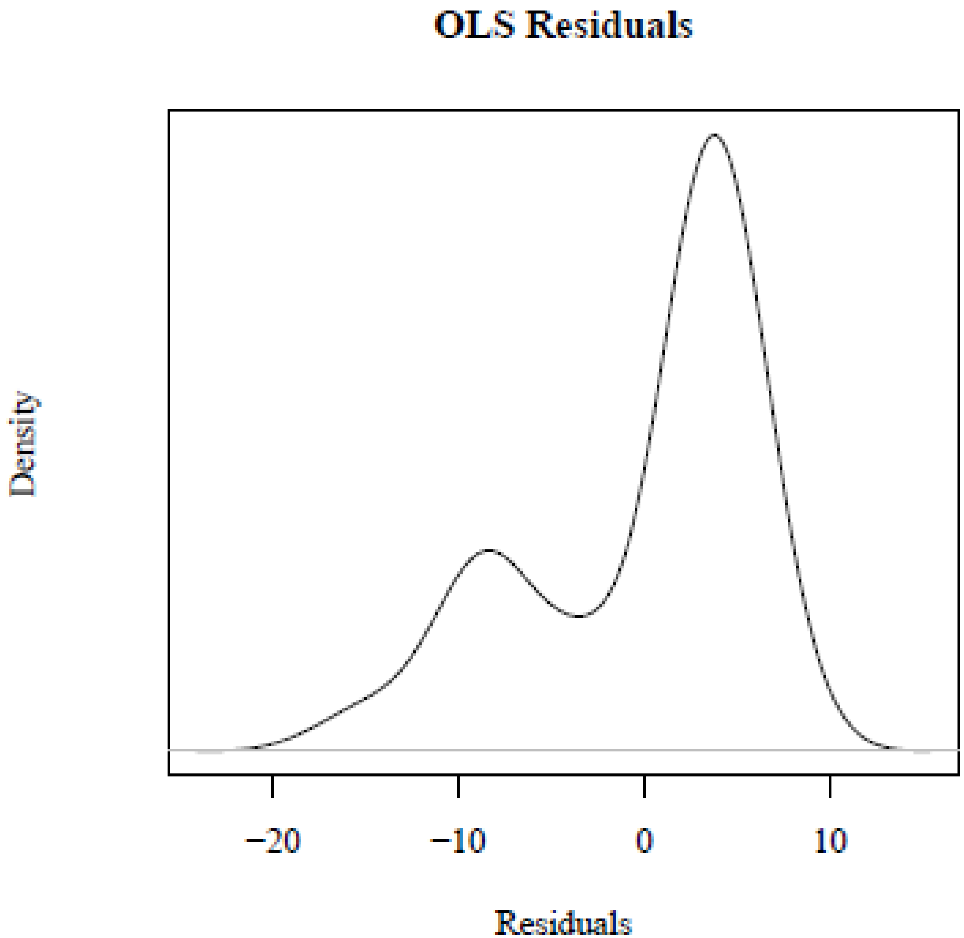
Source:
Ward (2008). Spatial Regression Models, Sage Publications, Inc

Democracy score vs. log (GDP)



Source:
Ward (2008). Spatial Regression Models, Sage Publications, Inc

OLS Regression Residuals



Source:
Ward (2008). Spatial Regression Models, Sage Publications, Inc

Spatial Autocorrelation Test for OLS Residuals

- The computed Moran's I statistic for these OLS residuals is 0.40, with a variance of 0.0028 and has an associated p-value that is ~ 0.
- This tells us that the OLS results, which assume independent observations, are strongly affected by the spatial clustering in the dependent and independent variables. As a result, they are likely to be ***misleading*** for both the statistical and substantive inferences that we may wish to draw about the relationship between democracy and its social requisite of wealth, as captured in GDP per capita.

Source:
Ward (2008). Spatial Regression Models, Sage Publications, Inc

Spatial Lag Model

$$y = \rho W y + X\beta + \varepsilon$$

- Using Maximum Likelihood Estimation (MLE) to estimate rho (ρ) and beta (β).

TABLE 2.3. *MLE estimates of the spatially lagged y model.*

	$\hat{\beta}$	SE($\hat{\beta}$)	z-value
Intercept	-6.20	2.08	-2.98
Ln GDP per capita	0.99	0.28	3.59
ρ	0.56	0.08	7.43
N = 158			
Log likelihood (df=4) = -491.10			

Equilibrium (Spillover) Effects in Spatial Lag Model

$$y = \mathbf{X}\beta + \rho \mathbf{W} y + \epsilon.$$

$$\Rightarrow \quad (\mathbf{I} - \rho \mathbf{W}) y = \mathbf{X}\beta + \epsilon.$$

spatial multiplier

$$\Rightarrow \quad E(y) = \boxed{(\mathbf{I} - \rho \mathbf{W})^{-1}} \mathbf{X}\beta.$$

This multiplier tells us how much of the change in x_i will "spill over" onto other states j and in turn affect y_i through the impact of y in the spatial lag.

Measuring Spillover Effects

- To understand how one state's GDP per capita affects the expected value of democracy in other states

$$(\mathbf{I} - \rho\mathbf{W})^{-1}\beta\Delta x(i)$$

Measuring Spillover Effects

Equilibrium impacts of log GDP per capita (X) for Russia

Country	Impact
Russia	1.09
People's Republic of Korea	0.24
Japan	0.24
Mongolia	0.24
Finland	0.22
Estonia	0.21
Norway	0.20
Lithuania	0.20
Latvia	0.120
Armenia	0.18

Effects on predicted democracy (Y) if China had a POLITY score of 10

Country	impact
Taiwan	1.88
North Korea	1.88
Mongolia	1.88
Nepal	1.41
Bhutan	1.41
Pakistan	1.13
Laos	1.13
Kyrgyzstan	1.13
Bangladesh	1.13
Uzbekistan	0.94
Thailand	0.94
Myanmar/Burma	0.94
Tajikistan	0.80
India	0.80
Vietnam	0.80
Afghanistan	0.80
Kazakhstan	0.70
Russia	0.28

回顧：空間迴歸模式的R函數

- OLS: lm (y~x1+x2+..., data=)
- Moran's I for Regression Residuals: lm.morantest (lm, listw=)
- SAR: spautolm (y~x1+x2+..., data= , listw=)
- SLM: lagsarlm (y~x1+x2+..., data= , listw=)
- SDM: lagsarlm (y~x1+x2+..., data= , listw= , type="mixed")
- SEM: errorsarlm (y~x1+x2+..., data= , listw=)
- Model Evaluation: lm.Lmtests (lm, listw= , test="all")
- Model Comparison: anova (lm1, lm2)