

# Spatial Regression

EE464: Urban Economics

EE562: Selected Topics in Development Economics 2

Semester 1 / 2020

Faculty of Economics, Thammasat University

# Main Contents

- (1) The latest textbook on spatial analysis
- (2) Criteria for selecting model's specification
- (3) Spatial regression techniques

**(1) The latest textbook on spatial analysis**

# Spatial Analysis Methods and Practice

## Describe – Explore – Explain through GIS

*by George Grekousis*

Publisher: Cambridge University Press

Expected online publication date: May 2020

Print publication year: 2020

Online ISBN: 9781108614528

<https://www.cambridge.org/core/books/spatial-analysis-methods-and-practice/4C135005A621335D06CC63EFF17E3913>

# SPATIAL ANALYSIS METHODS AND PRACTICE

DESCRIBE – EXPLORE – EXPLAIN THROUGH GIS



CAMBRIDGE  
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

[www.cambridge.org](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781108498982](http://www.cambridge.org/9781108498982)

DOI: 10.1017/9781108614528

© George Grekousis 2020

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2020

Printed in the United Kingdom by TJ International, Padstow Cornwall

*A catalogue record for this publication is available from the British Library.*

*Library of Congress Cataloging-in-Publication Data*

Names: Grekousis, George, author.

Title: Spatial analysis methods and practice : describe - explore - explain through GIS / George Grekousis, Sun Yat-Sen University (SYSU), China.

Description: First edition. | New York, NY : Cambridge University Press, 2020. |

Includes bibliographical references and index.

Identifiers: LCCN 2019038257 (print) | LCCN 2019038258 (ebook) | ISBN 9781108498982

(hardback) | ISBN 9781108712934 (paperback) | ISBN 9781108614528 (epub)

Subjects: LCSH: Spatial analysis (Statistics) | Geographic information systems.

Classification: LCC QA278.2 .G737 2020 (print) | LCC QA278.2 (ebook) | DDC 910.285–dc23

LC record available at <https://lcn.loc.gov/2019038257>

LC ebook record available at <https://lcn.loc.gov/2019038258>

ISBN 978-1-108-49898-2 Hardback

ISBN 978-1-108-71293-4 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Preface	page xiii
<b>Think Spatially: Basic Concepts of Spatial Analysis and Space</b>	
<b>Conceptualization</b>	1
Learning Objectives	1
1.1 Introduction: Spatial Analysis	2
1.2 Basic Definitions	6
1.3 Spatial Data: What Makes Them Special?	10
1.4 Conceptualization of Spatial Relationships	15
1.5 Distance Measure	17
1.5.1 Fixed Distance Band (Sphere of Influence)	18
1.5.2 Distance Decay	20
1.6 Contiguity: Adjacency Matrix	23
1.6.1 Polygons Contiguity	23
1.6.2 Adjacency Matrix	25
1.7 Interaction	26
1.8 Neighborhood and Neighbors	27
1.8.1 <i>k</i> -Nearest Neighbors ( <i>k</i> -NN)	27
1.8.2 Space–Time Window	29
1.8.3 Proximity Polygons	30
1.8.4 Delaunay Triangulation and Triangular Irregular Networks (TIN)	31
1.9 Spatial Weights and Row Standardization	31
1.10 Chapter Concluding Remarks	33
Questions and Answers	34
Lab 1 The Project: Spatial Analysis for Real Estate Market Investments	39
Overall Progress	39
Scope of Analysis	39
Dataset Structure	44
Guidelines	45

v

<b>4</b>	<b>Spatial Autocorrelation</b>	207
	Learning Objectives	207
	4.1 Spatial Autocorrelation	208
	4.2 Global Spatial Autocorrelation	211
	4.2.1 Moran's <i>I</i> Index and Scatter Plot	211
	4.2.2 Geary's <i>C</i> Index	216
	4.2.3 General <i>G</i> -Statistic	217
	4.3 Incremental Spatial Autocorrelation	219
	4.4 Local Spatial Autocorrelation	222
	4.4.1 Local Moran's <i>I</i> (Cluster and Outlier Analysis)	222
	4.4.2 Optimized Outlier Analysis	226
	4.4.3 Getis-Ord $G_i^*$ and $G_i^*$ (Hot Spot Analysis)	227
	4.4.4 Optimized Hot Spot Analysis	229
	4.5 Space–Time Correlation Analysis	230
	4.5.1 Bivariate Moran's <i>I</i> for Space–Time Correlation	230
	4.5.2 Differential Moran's <i>I</i>	232
	4.5.3 Emerging Hot Spot Analysis	233
	4.6 Multiple Comparisons Problem and Spatial Dependence	234
	4.7 Chapter Concluding Remarks	237
	Questions and Answers	238
	Lab 4 Spatial Autocorrelation	241
	Overall Progress	241
	Scope of the Analysis	241
	Section A ArcGIS	242
	Exercise 4.1 Global Spatial Autocorrelation	242
	Exercise 4.2 Incremental Spatial Autocorrelation and Spatial Weights Matrix	246
	Exercise 4.3 Cluster and Outlier Analysis (Anselin Local Moran's <i>I</i> )	252
	Exercise 4.4 Hot Spot Analysis (Getis-Ord $G_i^*$ ) and Optimized Hot Spot Analysis	255
	Exercise 4.5 Optimized Hot Spot Analysis for Crime Events	261
	Section B GeoDa	265
	Exercises 4.1 and 4.2 Global Spatial Autocorrelation and Spatial Weights Matrix	265
	Exercise 4.3 Cluster and Outlier Analysis (Anselin Local Moran's <i>I</i> )	270
	Exercise 4.4 Hot Spot Analysis (Getis-Ord $G_i^*$ )	273
<b>5</b>	<b>Multivariate Data in Geography: Data Reduction and Clustering</b>	275
	Learning Objectives	275
	5.1 Multivariate Data Analysis	276

5.2	Principal Component Analysis (PCA)	279
5.3	Factor Analysis (FA)	289
5.4	Multidimensional Scaling (MDS)	290
5.5	Cluster Analysis	292
5.5.1	Hierarchical Clustering	293
5.5.2	k-Means Algorithm (Partitional Clustering)	298
5.6	Regionalization	304
5.6.1	SKATER Method	306
5.6.2	REDCAP Method	309
5.7	Density-Based Clustering: DBSCAN, HDBSCAN, OPTICS	310
5.8	Similarity Analysis: Cosine Similarity	311
5.9	Chapter Concluding Remarks	313
	Questions and Answers	314
	Lab 5 Multivariate Statistics: Clustering	317
	Overall Progress	317
	Scope of the Analysis	317
	Section A ArcGIS	318
	Exercise 5.1 k-Means Clustering	318
	Exercise 5.2 Spatial Clustering (Regionalization)	329
	Exercise 5.3 Similarity Analysis	332
	Exercise 5.4 Synthesis	337
	Section B GeoDa	345
	Exercise 5.1 k-Means Clustering	345
	Exercise 5.2 Spatial Clustering	348
<b>6</b>	<b>Modeling Relationships: Regression and Geographically Weighted Regression</b>	<b>351</b>
	Learning Objectives	351
6.1	Simple Linear Regression	352
6.1.1	Simple Linear Regression Assumptions	355
6.1.2	Ordinary Least Squares (Intercept and Slope by OLS)	355
6.2	Multiple Linear Regression (MLR)	356
6.2.1	Multiple Regression Basics	356
6.2.2	Model Overfit: Selecting the Number of Variables by Defining a Functional Relationship	359
6.2.3	Missing Values	360
6.2.4	Outliers and Leverage Points	360
6.2.5	Dummy Variables	361
6.2.6	Methods for Entering Variables in MLR: Explanatory Analysis; Identifying Causes and Effects	363

6.3	Evaluating Linear Regression Results: Metrics, Tests and Plots	365
6.3.1	Multiple $r$	365
6.3.2	Variation and Coefficient of Determination $R$ -Squared	365
6.3.3	Adjusted $R$ -Squared	369
6.3.4	Predicted $R$ -Squared	370
6.3.5	Standard Error (Deviation) of Regression (or Standard Error of the Estimate)	371
6.3.6	F-Test of the Overall Significance	373
6.3.7	t-Statistic (Coefficients' Test)	374
6.3.8	Wald Test (Coefficient's Test)	375
6.3.9	Standardized Coefficients (Beta)	376
6.3.10	Residuals, Residual Plots and Standardized Residuals	378
6.3.11	Influential Points: Outliers and High-Leverage Observations	381
6.4	Multiple Linear Regression Assumptions: Diagnose and Fix	383
6.5	Multicollinearity	387
6.6	Worked Example: Simple and Multiple Linear Regression	391
6.7	Exploratory Regression	400
6.8	Geographically Weighted Regression	403
6.8.1	Spatial Kernel Types	405
6.8.2	Bandwidth	406
6.8.3	Interpreting GWR Results and Practical Guidelines	406
6.9	Chapter Concluding Remarks	409
	Questions and Answers	411
	Lab 6 OLS, Explanatory Regression, GWR	415
	Overall Progress	415
	Scope of the Analysis	415
	Exercise 6.1 Exploratory Regression	416
	Exercise 6.2 OLS Regression	428
	Exercise 6.3 GWR	437
<b>7</b>	<b>Spatial Econometrics</b>	<b>451</b>
	Learning Objectives	451
7.1	Spatial Econometrics	451
7.2	Spatial Dependence: Spatial Regression Models and Diagnostics	453
7.2.1	Diagnostics for Spatial Dependence	454
7.2.2	Selecting between Spatial Lag or Spatial Error Model	457
7.2.3	Estimation Methods	459

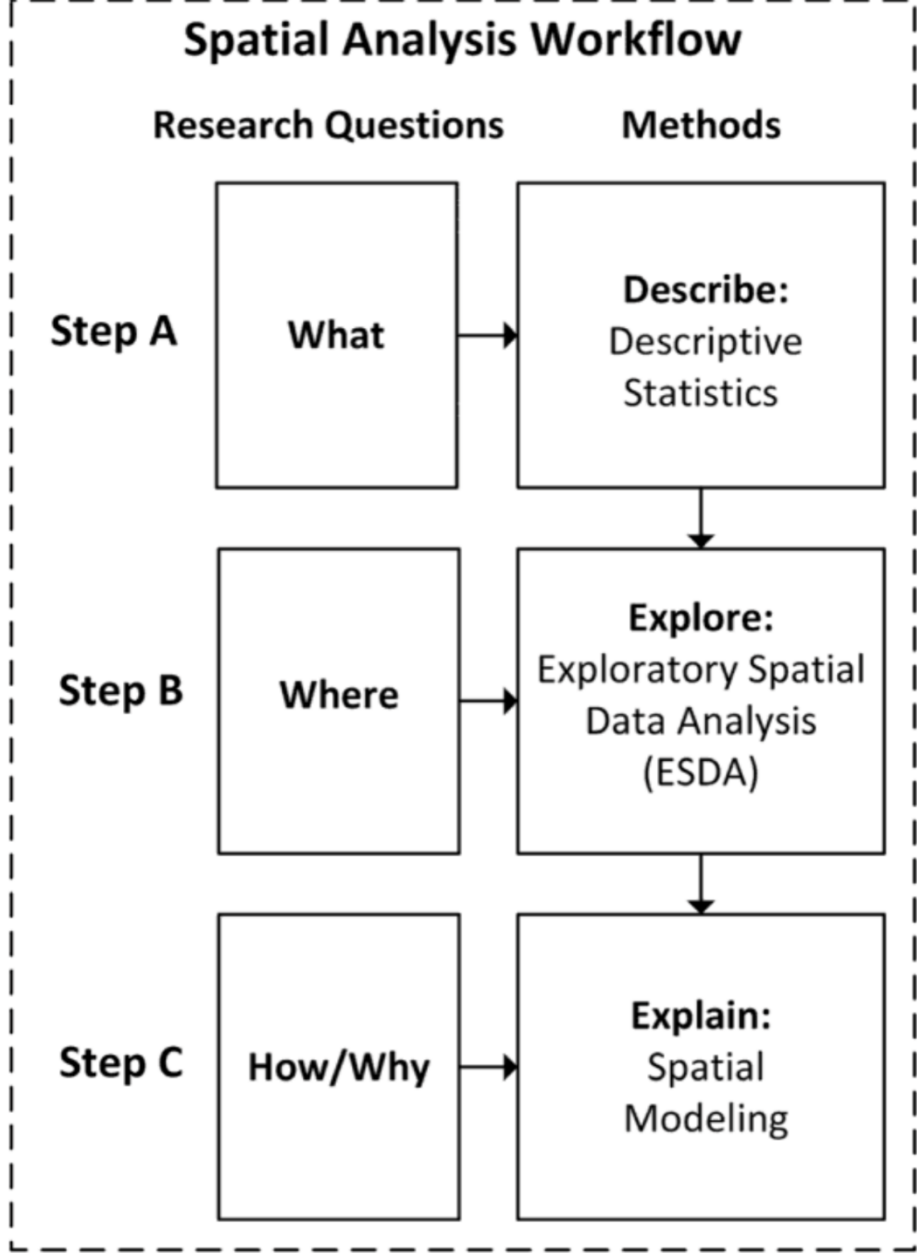


Figure 1.1 Spatial analysis workflow.

**Overall Progress**

**Spatial Analysis/ Lab Workflow**

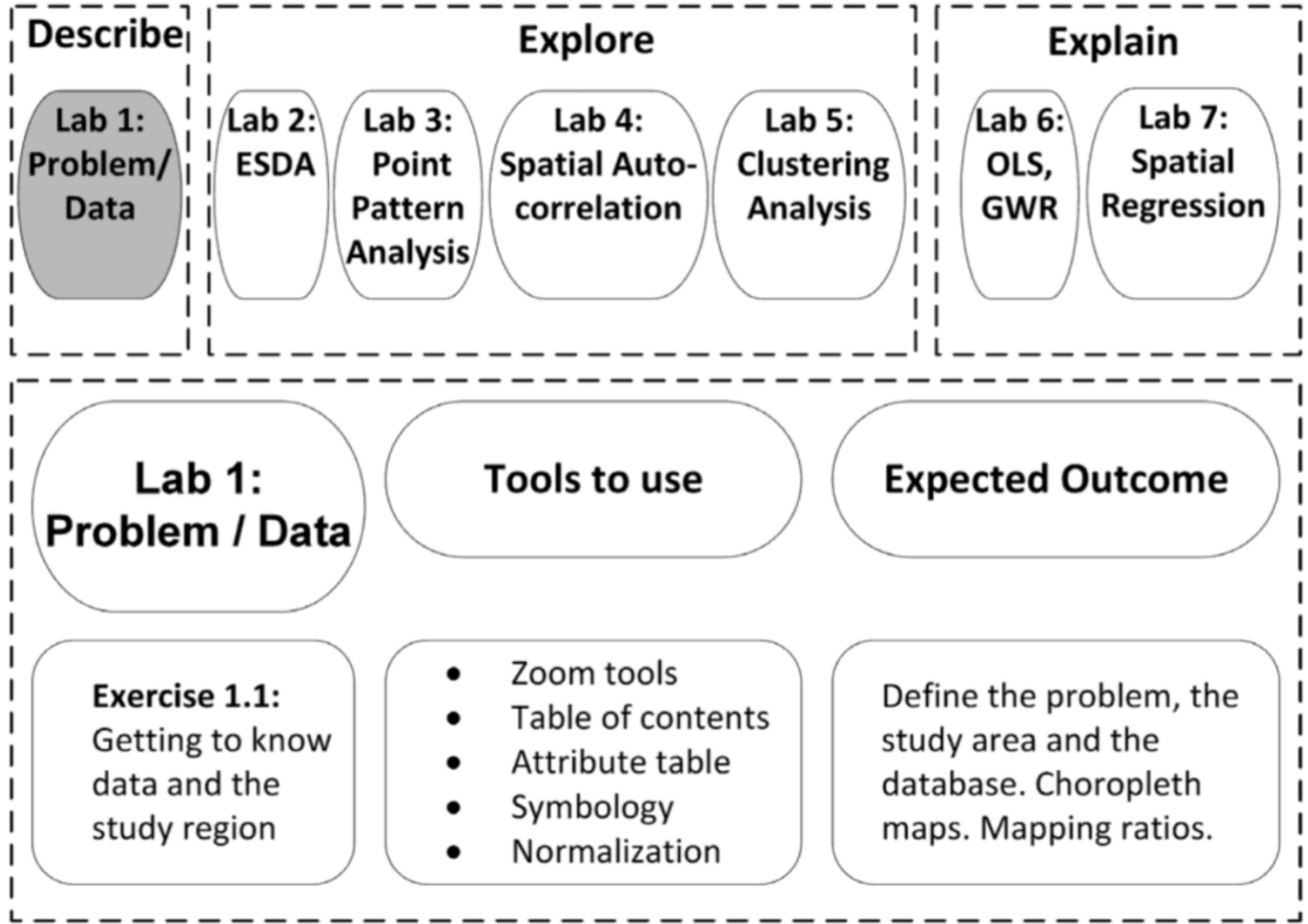
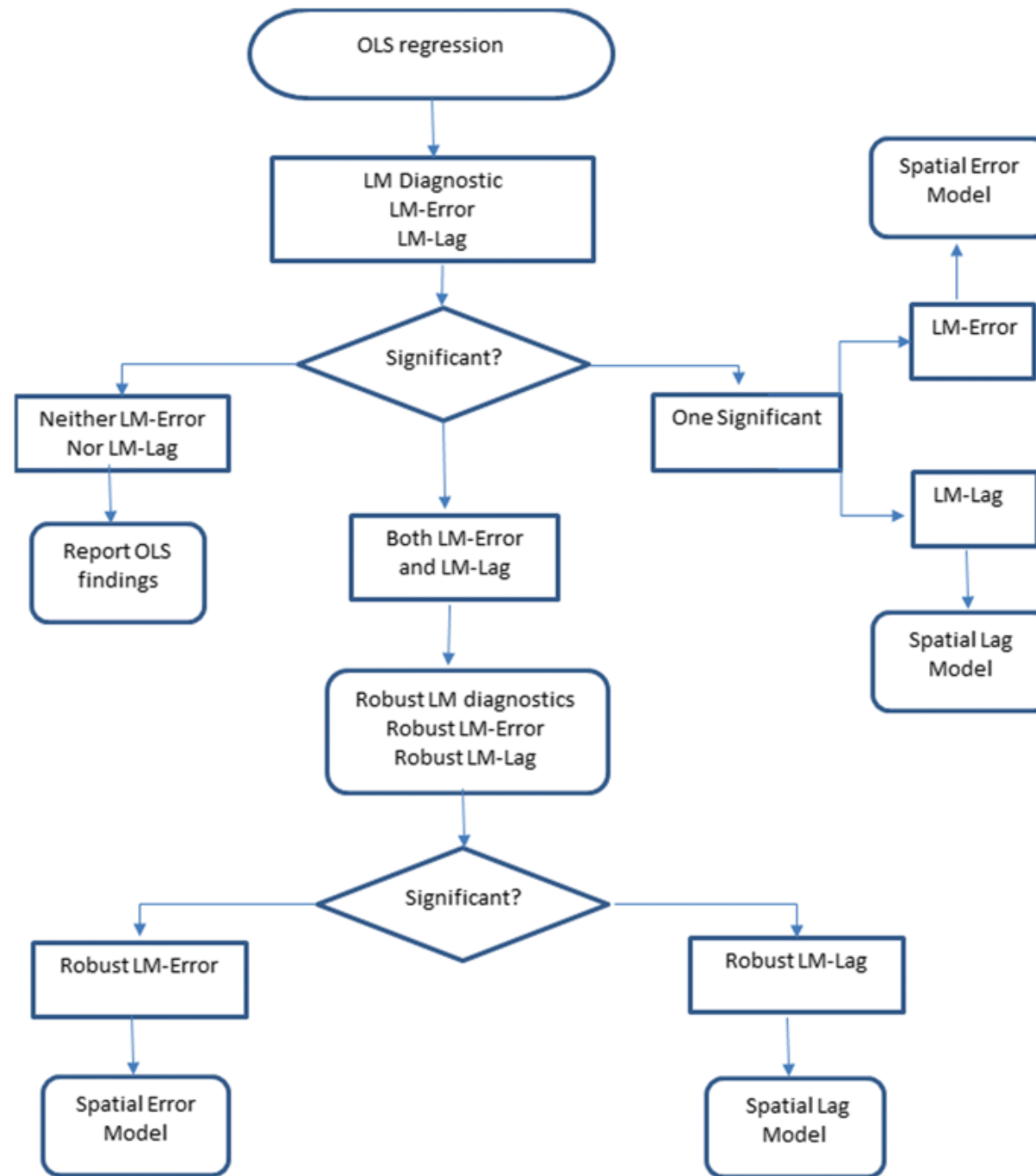


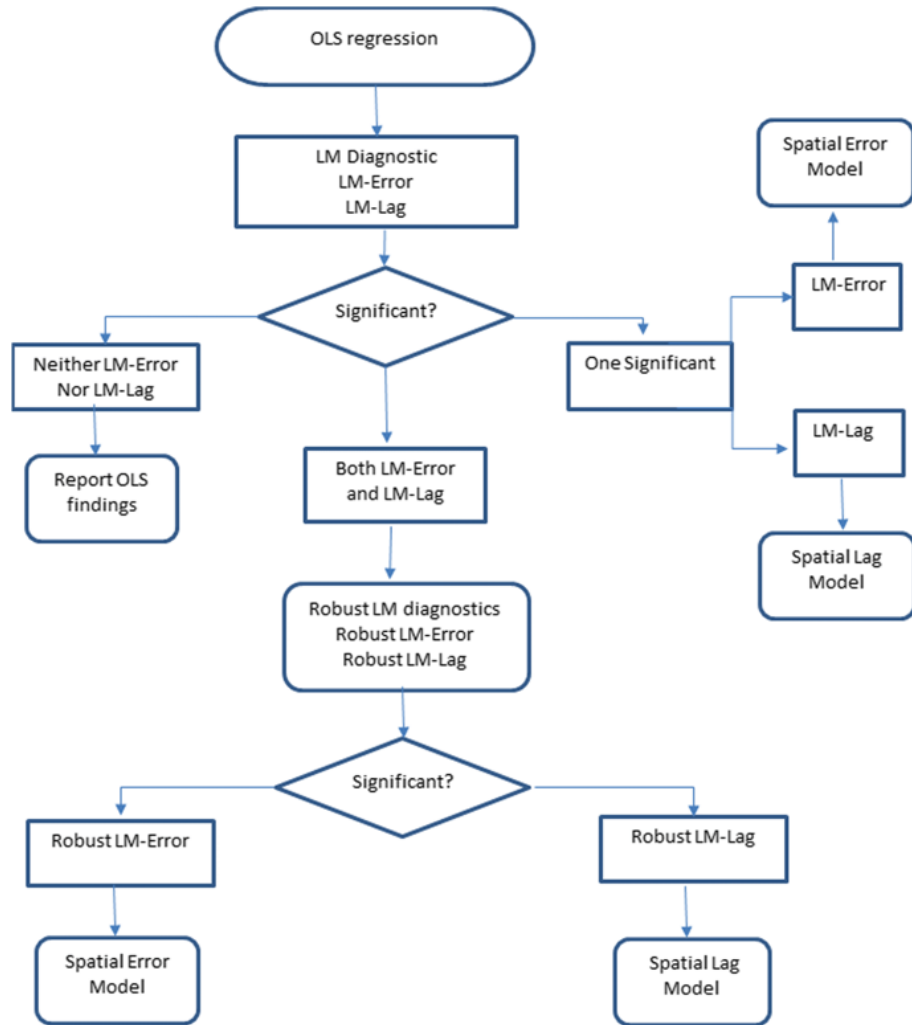
Figure 1.7 Lab 1 workflow and overall progress.

**Source: Grekousis (2020)**

## (2) Criteria for selecting model's specification



# Example 1



>>01/12/20 11:43:01

## REGRESSION

### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : TRF\_Thailand data 2008  
 Dependent Variable : ln\_gpp Number of Observations: 75  
 Mean dependent var : 10.6273 Number of Variables : 6  
 S.D. dependent var : 1.10065 Degrees of Freedom : 69

R-squared : 0.475942 F-statistic : 12.5329  
 Adjusted R-squared : 0.437966 Prob(F-statistic) : 1.15138e-008  
 Sum squared residual: 47.6149 Log likelihood : -89.3826  
 Sigma-square : 0.690071 Akaike info criterion : 190.765  
 S.E. of regression : 0.830705 Schwarz criterion : 204.67  
 Sigma-square ML : 0.634866  
 S.E of regression ML: 0.796785

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	-11.4482	12.2164	-0.937113	0.35197
Ln_Pop_D	-0.240969	0.22103	-1.09021	0.27941
Ln_NTL	1.24261	0.221595	5.6076	0.00000
Ln_NDVI_M	-1.09656	1.27192	-0.862127	0.39160
Ln_Rain_M	0.957134	0.46944	2.03889	0.04529
Ln_LSTD_M	4.2676	3.44137	1.24009	0.21914

### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 423.656107

### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	4.8372	0.08905

### DIAGNOSTICS FOR HETEROSKEDASTICITY

#### RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	16.8776	0.00474
Koenker-Bassett test	5	10.5731	0.06053

### DIAGNOSTICS FOR SPATIAL DEPENDENCE

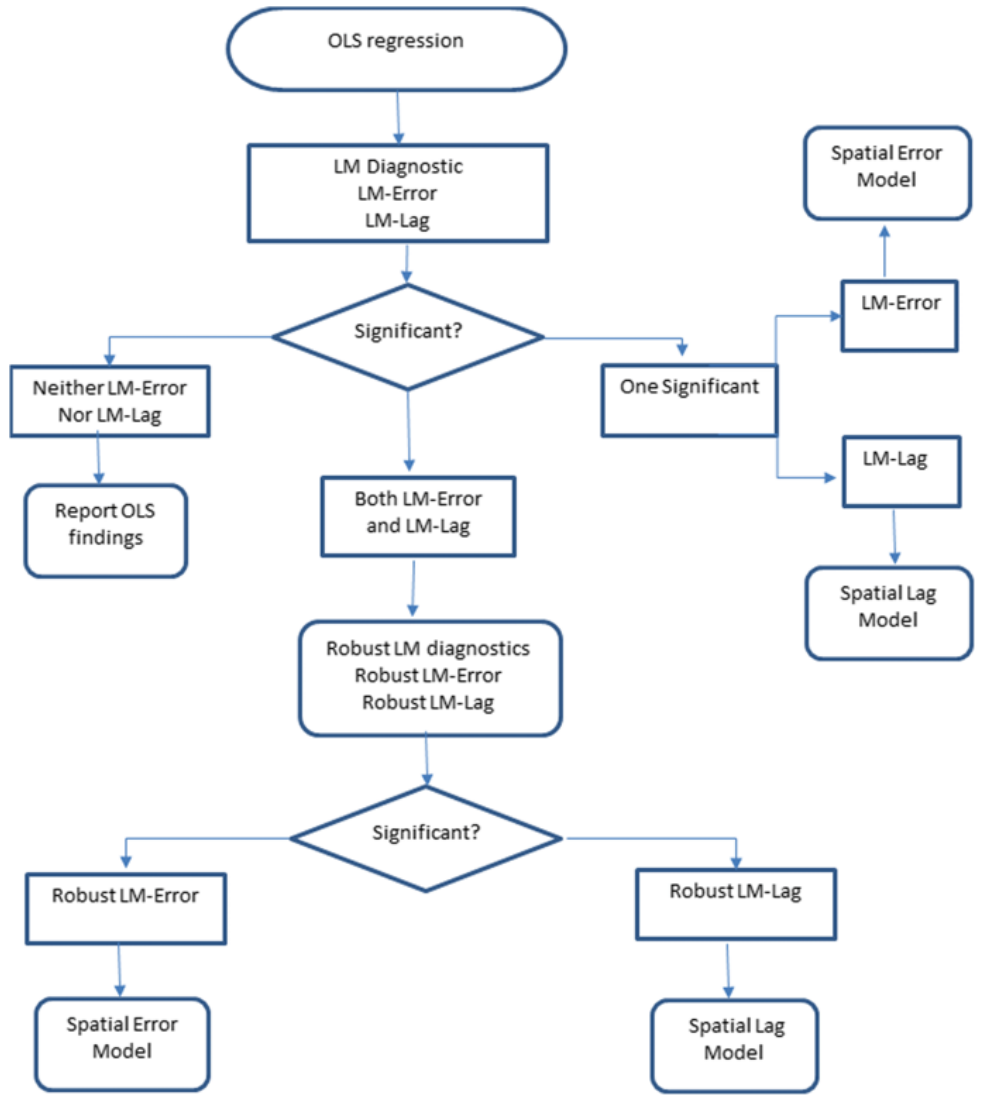
FOR WEIGHT MATRIX : TRF\_Thailand data 2008

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.0280	0.8742	0.38199
Lagrange Multiplier (lag)	1	0.0050	0.94362
Robust LM (lag)	1	0.3243	0.56904
Lagrange Multiplier (error)	1	0.0867	0.76838
Robust LM (error)	1	0.4060	0.52400
Lagrange Multiplier (SARMA)	2	0.4110	0.81423

END OF REPORT

# Example 2



SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : TRF\_Thailand data 2016

Dependent Variable : All\_Avg Number of Observations: 77

Mean dependent var : 8.73412 Number of Variables : 6

S.D. dependent var : 0.236003 Degrees of Freedom : 71

R-squared : 0.687733 F-statistic : 31.274

Adjusted R-squared : 0.665743 Prob(F-statistic) : 1.10685e-016

Sum squared residual: 1.33922 Log likelihood : 46.7328

Sigma-square : 0.0188623 Akaike info criterion : -81.4656

S.E. of regression : 0.13734 Schwarz criterion : -67.4028

Sigma-square ML : 0.0173925

S.E of regression ML: 0.131881

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	6.76934	1.95872	3.456	0.00093
Ln_Pop_D	-0.0259939	0.0344682	-0.75414	0.45326
Ln_NTL	0.184405	0.0192027	9.60305	0.00000
Ln_NDVI_M	0.0697841	0.234638	0.297412	0.76702
Ln_Rain_M	0.176842	0.0857219	2.06297	0.04275
Ln_LSTD_M	0.321049	0.529664	0.606137	0.54636

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 379.980969

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	0.6733	0.71416

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	6.5858	0.25331
Koenker-Bassett test	5	7.8637	0.16391

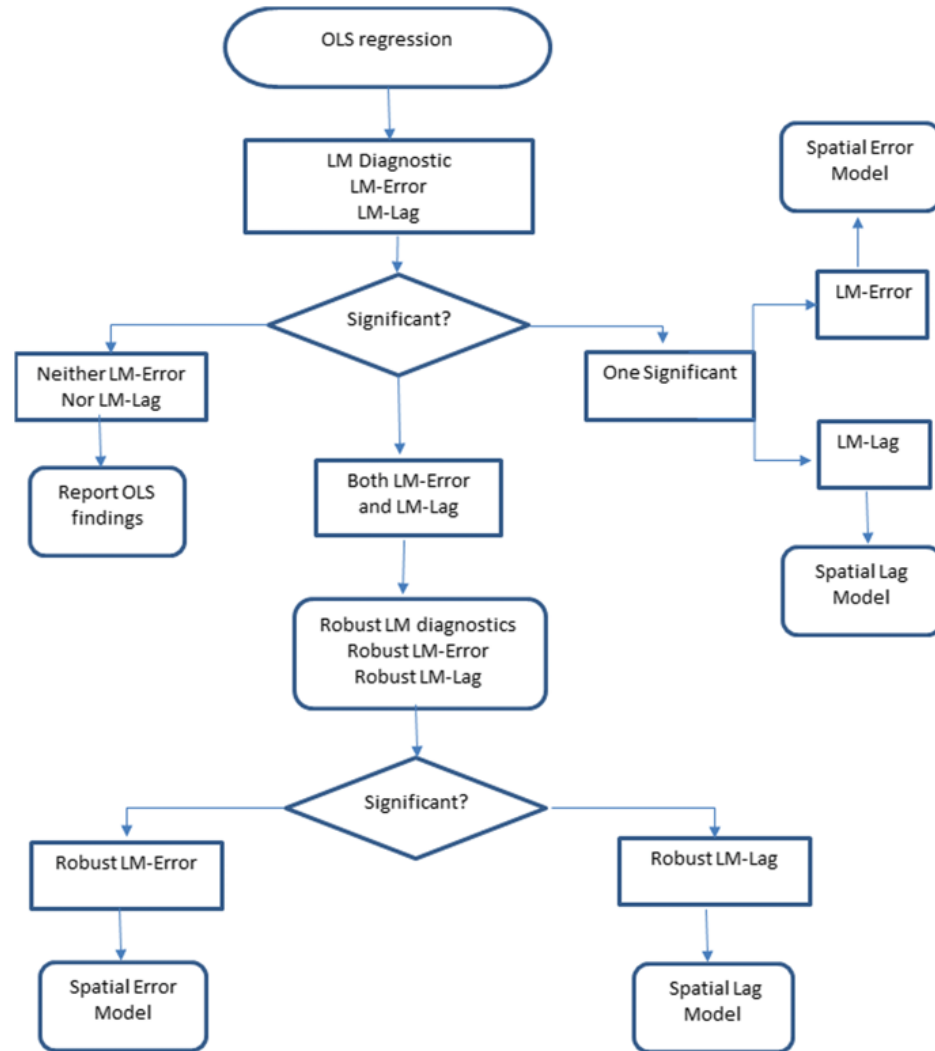
DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : TRF\_Thailand data 2016  
(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.0964	1.6592	0.09707
Lagrange Multiplier (lag)	1	3.3019	0.06920
Robust LM (lag)	1	2.2860	0.13054
Lagrange Multiplier (error)	1	1.0895	0.29659
Robust LM (error)	1	0.0736	0.78622
Lagrange Multiplier (SARMA)	2	3.3755	0.18494

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

# Example 3



>>03/13/20 21:54:22

## REGRESSION

### SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

Data set : TH\_JETRO\_IndusSpillovr  
 Dependent Variable : TH\_lnVA Number of Observations: 76  
 Mean dependent var : 23.4965 Number of Variables : 6  
 S.D. dependent var : 1.7812 Degrees of Freedom : 70

R-squared : 0.882644 F-statistic : 105.295  
 Adjusted R-squared : 0.874261 Prob(F-statistic) : 3.70775e-031  
 Sum squared residual: 28.2975 Log likelihood : -70.2968  
 Sigma-square : 0.404249 Akaike info criterion : 152.594  
 S.E. of regression : 0.635806 Schwarz criterion : 166.578  
 Sigma-square ML : 0.372335  
 S.E of regression ML: 0.610193

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	6.87966	1.61102	4.27038	0.00006
TH_lnL	0.344717	0.128305	2.6867	0.00901
TH_lnK	0.421677	0.107584	3.9195	0.00020
lnNTL	0.68795	0.192754	3.56906	0.00065
TH_InIM	-0.0176677	0.0197637	-0.893948	0.37441
TH_lnEX	0.164259	0.0335603	4.89446	0.00001

### REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 93.190925

### TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	1.7363	0.41974

### DIAGNOSTICS FOR HETEROSKEDASTICITY

#### RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	5	11.3263	0.04528
Koenker-Bassett test	5	9.5991	0.08743

### DIAGNOSTICS FOR SPATIAL DEPENDENCE

FOR WEIGHT MATRIX : TH\_JETRO\_IndusSpillovr

(row-standardized weights)

TEST	MI/DF	VALUE	PROB
Moran's I (error)	0.2793	3.5698	0.00036
Lagrange Multiplier (lag)	1	14.7833	0.00012
Robust LM (lag)	1	7.3401	0.00674
Lagrange Multiplier (error)	1	9.1910	0.00243
Robust LM (error)	1	1.7478	0.18615
Lagrange Multiplier (SARMA)	2	16.5311	0.00026

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

## **(3) Spatial regression techniques**

# OLS

$x_i$



$y_i$



$u_i$

$$Y_i = a + \beta X_i + u_i$$

$x_j$



$y_j$



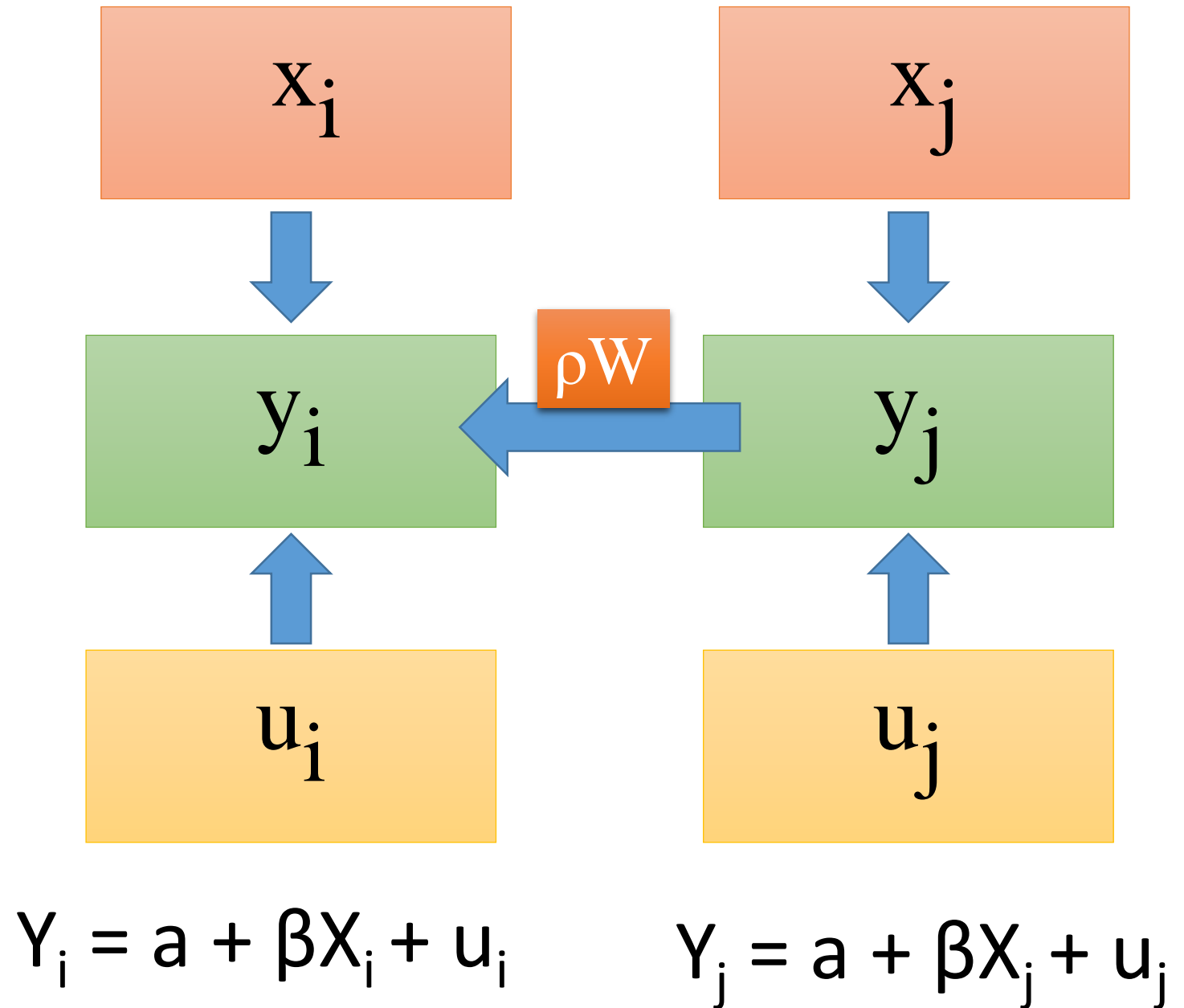
$u_j$

$$Y_j = a + \beta X_j + u_j$$

# SLM

Spatial Lag Model

$$Y_i = \rho W_{ij} Y_j + a + \beta X_i + u_i$$

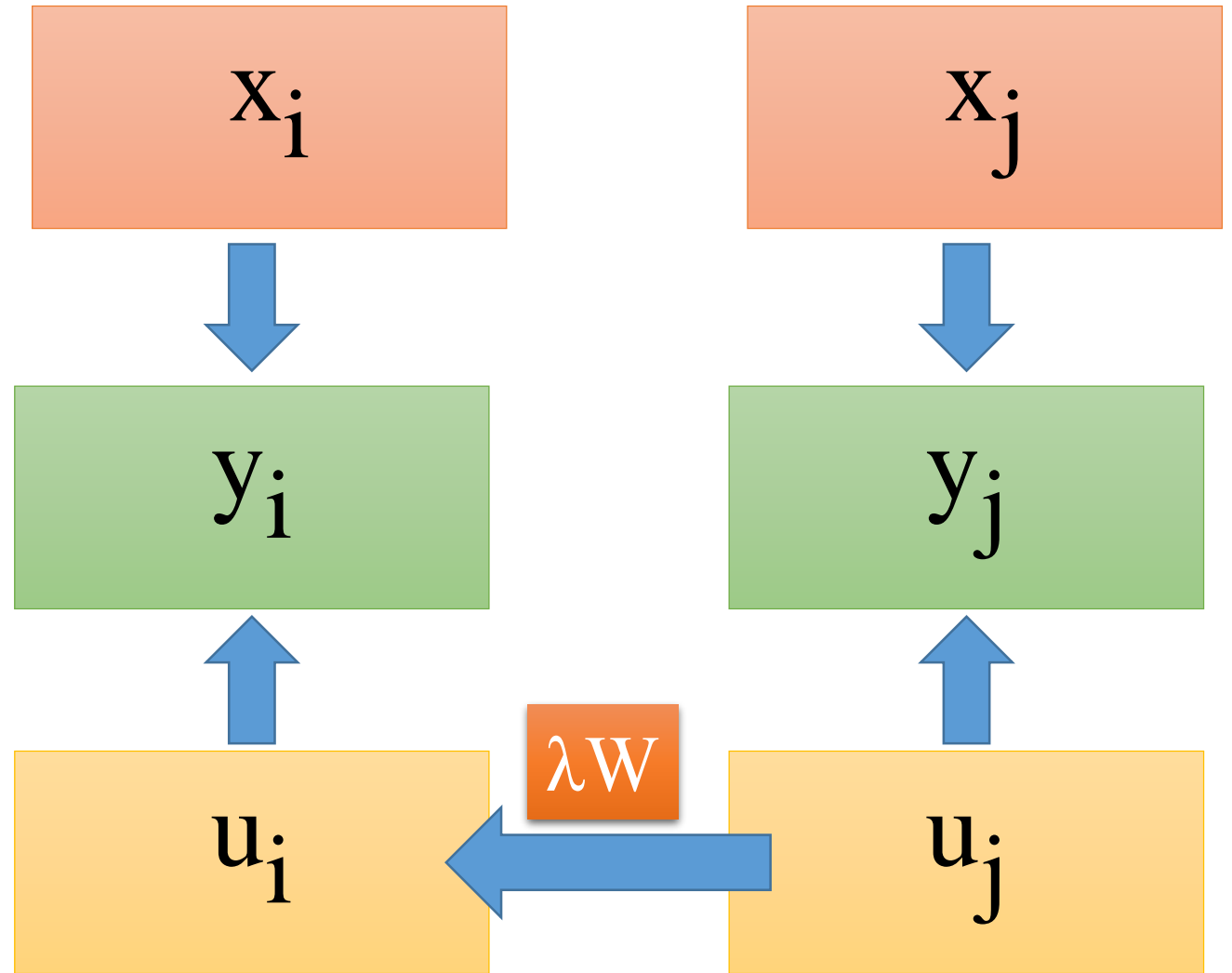


# SEM

Spatial Error Model

$$Y_i = a + \beta X_i + u_i$$

$$u_i = \lambda W_{ij} u_j + \varepsilon_i$$



$$Y_i = a + \beta X_i + u_i$$

$$Y_j = a + \beta X_j + u_j$$

# SDM

Spatial Durbin Model

$$Y_i = \rho W_{ij} Y_j + a + \beta X_i + W_{ij} X_j \theta + u_i$$

