PALGRAVE TEXTS IN ECONOMETRICS

A Primer for Spatial Econometrics

With Applications in R

Giuseppe Arbia

A Primer for Spatial Econometrics

PalgraveTexts in Econometrics

General Editors: **Kerry Patterson and Terence Mills**
Regional Editors: **Northern Europe: Niels Haldrup; Far East: In Choi; USA: William Greene; Southern Europe and Oceania: Tommaso Proietti**

The pace of developments in econometrics has made it increasingly difficult for students and professionals alike to be aware of what is impotent and likely to be lasting in theory and practical applications. This series addresses important key developments within a unified framework, with individual volumes organised thematically. The series is relevant for both students and professionals who need to keep up with the econometric developments, yet is written for a wide audience with a style that is designed to make econometric concepts available to economists who are not econometricians.

Titles include:
Simon P. Burke and John Hunter
MODELLING NON-STATIONARY TIME SERIES

Michael P. Clements
EVALUATING ECONOMETRIC FORECASTS OF ECONOMIC AND FINANCIAL VARIABLES

Lesley Godfrey
BOOTSTRAP TESTS FOR REGRESSION MODELS

Terence C. Mills
MODELLING TRENDS AND CYCLES IN ECONOMIC TIME SERIES

Kerry Patterson
A PRIMER FOR UNIT ROOT TESTING

Kerry Patterson
UNIT ROOTS TESTS IN TIME SERIES VOLUME 1
Key Concepts and Problems

Kerry Patterson
UNIT ROOTS TESTS IN TIME SERIES VOLUME 2
Extensions and Developments

Terence C. Mills
ANALYSING ECONOMIC DATA
A Concise Introduction

Giuseppe Arbia
A PRIMER FOR SPATIAL ECONOMETRICS
With Applications in R

Palgrave Texts in Econometrics
Series Standing Order ISBN 978–1–403–90172–9 (hardback)
978–1–403–90173–6 (paperback)
(*outside North America only*)

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the title of the series and the ISBN quoted above.

Customer Services Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England

---

A Primer for Spatial Econometrics

With Applications in R

Giuseppe Arbia

palgrave
macmillan

Softcover reprint of the hardcover 1st edition 2014 978-0-230-36038-9

First published 2014 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

ISBN 978-1-137-42816-5 ISBN 978-1-137-31794-0 (eBook)

DOI 10.1057/9781137317940

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

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

A catalog record for this book is available from the Library of Congress.

Typeset by MPS Limited, Chennai, India.

Transferred to Digital Printing in 2014

To P, E, F & E

Contents

List of Figures ix

List of Examples x

Foreword by William Greene xii

Preface and Acknowledgements xiii

1 The Classical Linear Regression Model **1**
1.1 The basic linear regression model 1
1.2 Non-sphericity of the disturbances 11
1.3 Endogeneity 16
1.4 `R Codes`: running a regression 20
Key Terms and Concepts Introduced 22
Questions 22
Exercises 23
References to the Chapter 25

2 Some Important Spatial Definitions **26**
2.1 The Spatial Weight Matrix *W* and the definition of Spatial Lag 26
2.2 Testing spatial autocorrelation among OLS residuals without an explicit alternative hypothesis 33
2.3 `R Codes` 39
Key Terms and Concepts Introduced 47
Questions 47
Exercises 47
References to the Chapter 50

3 Spatial Linear Regression Models **51**
3.1 Generalities 51
3.2 Pure spatial autoregression 52
3.3 The classical model with spatially lagged non-stochastic regressors 55
3.4 The Spatial Error Model (SEM) 55
3.5 The Spatial Lag Model (SLM) 66
3.6 The general SARAR(1,1) Model 74
3.7 Testing spatial autocorrelation among the residuals with an explicit alternative hypothesis 82

3.8 Interpretation of the parameters in spatial econometric models 87
3.9 `R Codes`: estimating linear spatial models 90
Key Terms and Concepts Introduced 93
Questions 93
Exercises 94
References to the Chapter 97

4 Further Topics in Spatial Econometrics 99
4.1 Heteroscedastic innovations 99
4.2 Spatial models for binary response variables 111
4.3 Spatial panel data models (Written by Giovanni Millo) 128
4.4 Non-stationary spatial econometric models 145
4.5 `R Codes` 153
Key Terms and Concepts Introduced 158
Questions 160
Exercises 160
References to the Chapter 162

5 Alternative Model Specifications for Big Datasets 167
5.1 Introduction 167
5.2 The MESS specification 169
5.3 The unilateral approximation approach 178
5.4 A composite likelihood approach 189
5.5 `R Codes` 196
Key Terms and Concepts Introduced 198
Questions 198
Exercises 199
References to the Chapter 200

6 Conclusions: What's Next? 202

Solutions to the Exercises 207

Index 225

List of Figures

2.1 Spatial autocorrelation and spatial heterogeneity 27
2.2 Contiguity criteria in a regular square lattice grid 27
2.3 Boundaries of the 20 Italian regions 41
5.1 Three regions arranged on a regular lattice grid 180
5.2 Two different W matrices incorporating two different directions of spatial dependence 181
5.3 Bivariate coding pattern in a regular square lattice grid 190
5.4 Four different bivariate coding schemes for the data in Figure 5.3 195

List of Examples

1.1	Barro and Sala-i-Martin model of regional convergence	9
1.2	Zellner–Revankar revised Cobb–Douglas production function	10
1.3	Barro and Sala-i-Martin model of regional convergence (continued)	15
1.4	Zellner–Revankar Production function (continued)	18
2.1	Some examples of W matrices	28
2.2	UK regions contiguity-based W matrix	30
2.3	Okun's law for the 20 Italian regions	35
2.4	Phillips curve for the 20 Italian regions	37
3.1	Spatial autoregression of price index in the 20 Italian regions	54
3.2	The relationship between used cars price and taxes in 48 US states	63
3.3	House price determinants in Boston	70
3.4	House price determinants in Boston (continued)	80
3.5	Phillips curve (continued)	86
3.6	Okun's law in Italy (continued)	90
4.1	An urban study: The determinants of crime in Columbus (Ohio)	106
4.2	Explaining luxury house prices in Baltimore	115
4.3	Luxury house prices in Baltimore (continued)	127
4.4	Munnell's model of public capital productivity	130
4.5	Munnell's model of public capital productivity (continued)	142
4.6	A spatial panel fixed effects analysis of convergence in Italian regions	143
4.7	The determinant of educational achievements in Georgia	151
5.1	Health planning in Mexico	176

5.2 Anisotropies in the Barro and Sala-i-Martin convergence model 183
5.3 Health planning in Mexico (continued) 187
5.4 Monte Carlo assessment of BML estimation method 195

Foreword

I have had the distinct pleasure to enjoy my own study of spatial econometrics in collaboration with Giuseppe Arbia during his work on this monograph. There is much to be learned in this rapidly growing field. This primer introduces the workhorse of the field, the spatially autoregressive and spatially autocorrelated linear regression model. A chapter is devoted to important terms of art in the field. We then progress through extensions of the linear model to heteroscedasticity and panel data treatments. Recent developments of spatial econometrics have extended the models to many non-linear cases, including binary and multinomial choice, stochastic frontiers, sample selection and models for count data and ordered choices. This primer provides a gateway to that literature through a presentation of a spatial binary choice model. Readers will appreciate the extensive presentation of real numerical examples in R, which has emerged as the software of choice for model builders in this area.

Spatial econometrics is enjoying widely dispersed growth spurt in many of the social sciences. Readers of this primer will find it to be a very approachable, informative springboard to their own ongoing study and research, as I have.

William Greene
Stern School of Business
New York, August, 2013

Preface and Acknowledgements

Spatial econometrics is a rapidly expanding topic with applications in so many diverse scientific fields that it is almost impossible to fully enumerate all the disciplines. Indeed, in recent years we can find applications in fields such as regional economics, criminology, public finance, industrial organization, political sciences, psychology, agricultural economics, health economics, demography, epidemiology, managerial economics, urban planning, education, land use, social sciences, economic development, innovation diffusion, environmental studies, history, labor, resources and energy economics, transportation, food security, real estate, marketing, and many others.

Given the widespread interest in the subject, this book aims to meet the growing demand by introducing basic spatial econometrics methodologies to a wide variety of researchers. It is specifically designed as a reference for applied researchers who want to receive a broad overview on the topic. In this sense, it is not intended to be a comprehensive textbook on the subject; rather the intention is to keep the details to a minimum and to provide a practical guide which illustrates the potential of spatial econometric modeling, discusses problems and solutions, and enables the reader to correctly interpret empirical results and to start working with the methods.

There are several features that distinguish this book from other existing texts on the subject. First, this book is self-contained and it does not assume any background knowledge apart from a working knowledge of elementary inferential statistics. Chapter 1 contains a summary of basic standard econometric results that are used throughout this text; as such, it can be omitted by the reader who is already knowledgeable on the subject. The treatment of the various topics is rigorous, but with formal proofs that are low level and reduced to a minimum. The book provides the minimum essential basics and intuitions on each topic and refers to other textbooks and to the literature for a more in-depth discussion. Furthermore, the text is integrated with examples, problem sets and practical exercises. To some extent, one may think of this book as an extended chapter of an econometrics textbook. It thus explicitly aims to bridge the gap between the standard textbooks, which still largely ignore the subject, and the more comprehensive and specialized textbooks. Although the book is application-oriented, I have taken care

to provide sound methodological developments and to use notations that are generally accepted for the topics being covered. Therefore, the reader will find that this text provides good preparation for the study of more advanced spatial econometrics material.

Secondly, this book only partially overlaps with existing textbooks on the subject. Even if it does not contain a comprehensive treatment of all the topics in spatial econometrics, it nonetheless includes some of the recent advances that are not discussed in other existing textbooks. For example, various estimation alternatives to the traditional Maximum Likelihood paradigm, the treatment of heteroscedastic innovations, spatial discrete choice models and non-stationary models. In addition, even if the bulk of the book deals with synchronic cross-sectional spatial models, section 4.3 contains an introduction to the treatment of spatial panel data, a fast-growing field in spatial econometrics.

Thirdly, people working in this field quickly learn that most of the procedures illustrated in this book encounter severe computational limitations when applied to very large datasets with sample sizes approaching thousands of observations. Computational issues can, indeed, represent a big limitation for many scholars engaged in spatial analysis in non-specialized fields that do not have access to large computer facilities. They can represent a limitation even with the availability of powerful computing machines in all those cases when a real-time decision has to be taken on the basis of the econometric analysis such as, for example, epidemiological and environmental surveillance or computer-assisted surgery, based on medical imaging. To overcome such limitations, this book also includes a chapter entirely devoted to discussing a series of alternative estimation techniques that can help in dramatically reducing the computational time and computer memory requirements.

Finally, the text introduces the reader to the procedures contained in the free statistical software "R". While spatial econometric methodologies are still not integrated in econometric software products (such as, for example, Eviews, Gauss, Gretl, Limdep, Microfit, Minitab, RATS, SAS, SPSS, TSP and many others), there are presently some packages that deal with most of the topics treated in this book (for example, STATA, Matlab) together with more specialized software in the subject (for example, GeoDa). In this book, however, we have decided to illustrate the various methods by making use of the statistical language R for three main reasons. Firstly, the package is freely available, so that the reader of the book can immediately replicate the techniques using available data. Secondly, the R language is intuitive and requires only a small initial investment. Thirdly, since spatial econometrics is a rapidly

expanding topic, the language R guarantees almost real-time updating when new procedures are introduced in the literature.

The material presented can be used as a textbook for an introductory course in spatial econometrics which assumes a working knowledge in econometrics at the level of, for example, the introductory chapters of the 7th edition of the textbook by Greene (see W. Greene, *Econometric Analysis* (2011)) or similar. In particular, chapters 1 to 3 could constitute the material for a two- or three-day course (10–12 lecture hours). Chapters 4 (3–4 hours) and 5 could be supplemental material covered on additional days. For chapter 4, the instructor would have the option of covering the entire chapter which could take an entire day or cover a portion of the chapter which would only take three or four hours. The examples, the questions and the exercises contained in the final part of each chapter may be used in combination with a set of computer laboratory sessions that could accompany the lectures.

The idea of an introductory, easy-to-use, textbook for applied researchers was developed during the many occasions throughout the last decade, in which I had the opportunity to teach introductory courses in spatial econometrics at different universities and institutions in Milan, Barcelona, Fortaleza, Salvador-Bahia and in the summer school, 'Spatial Econometrics Advanced Institute', held yearly in Rome since 2008.

The first plan of the book was written while I was visiting the Economics Department at the Stern School of Business, at New York University during the spring semester of 2011. I am grateful to Bill Greene for inviting me on that occasion and in the following two years in spring when I worked on this project. The final draft was completed during the period I spent as a *professeur invité* at ERMES, Université Pantéon Assas, Paris II for which I am indebted to Alain Pirotte, who was so kind in hosting me and providing me with a suitable environment to finish my work. The rest of the work was done at the Catholic University of the Sacred Heart in Rome, where I moved in late 2011. Section 4.3 is written by Giovanni Millo of Assicurazioni Generali, Trieste, Italy and I am grateful to him for his substantial contribution to the preparation of the book and for his patience in working with me. I am also grateful to Carrie Dolan of the College of William & Mary, Williamsburg, USA for carefully proofreading my draft. I am obviously entirely responsible for all the remaining errors. Carrie was also invaluable in providing and editing some of the maps used in the text and in the examples. Thanks are due also to Miguel Flores of Tecnologico de Monterrey, Mexico for providing the data and the maps on Mexico on which I based some of the examples in chapter 5, and to Diego

Giuliani, Francesca Petrarca and Myriam Tabasso for their rapid, online assistance and advice in respect of some of the R procedures. Giovanni, Carrie, Miguel, Diego, Francesca and Myriam are all former participants of the aforementioned 'Spatial Econometrics Advanced Institute' and I would also wish to express my gratitude to all the students who took part at the school in the last six years (more than 200!) because their active presence in class was certainly the greatest stimulus I received in writing this book.

This book is dedicated to my family. If I look back to the forewords I have written for my previous books, I find expressions of thanks to Paola and to my three children for their presence, patience, help and encouragement. Twenty-five years have now passed since the publication of my first book, and seven since my last one. The children have grown up and many things have changed, but my family is still here with me and to them my thoughts gratefully go on this rainy and gloomy late winter day, when I am here sitting in front of my computer writing the final words of this book.

Rome, Ash Wednesday, 2014

1
The Classical Linear Regression Model

1.1 The basic linear regression model

Let us consider the following linear model

$$ {}_{n}y_{1} = {}_{n}X_{kk}\beta_{1} + {}_{n}\varepsilon_{1} \tag{1.1} $$

where ${}_{n}y_{1} = \begin{bmatrix} y_1 \\ \dots \\ \dots \\ y_n \end{bmatrix}$ is a vector of n observations of the dependent variable y, ${}_{n}X_{k} = \begin{bmatrix} 1 & X_{11} & X_{1k-1} \\ \dots & & \dots \\ 1 & & \\ 1 & X_{n1} & X_{nk-1} \end{bmatrix}$ a matrix of n observations on $k-1$ non-stochastic exogenous regressors including a constant term, ${}_{k}\beta_{1} = \begin{bmatrix} \beta_1 \\ \\ \beta_k \end{bmatrix}$ a vector of k unknown parameters to be estimated and ${}_{n}\varepsilon_{1} = \begin{bmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{bmatrix}$ a vector of stochastic disturbances. We will assume throughout the book that the n observations refer to territorial units such as regions or countries.

The classical linear regression model assumes normality, identicity and independence of the stochastic disturbances conditional upon the k regressors. In short

$$\varepsilon \,|\, X \approx i.i.d.N(0, \sigma^2_{\varepsilon\, n} I_n) \tag{1.2}$$

${}_nI_n$ being an n-by-n identity matrix. Equation (1.2) can also be written as:

$$E(\varepsilon \,|\, X) = 0 \tag{1.3}$$

$$E(\varepsilon\varepsilon^T \,|\, X) = \sigma^2_{\varepsilon n} I_n \tag{1.4}$$

Equation (1.3) corresponds to the assumption of *exogeneity*, Equation (1.4) to the assumption of *spherical disturbances* (Greene, 2011).

Furthermore it is assumed that the k regressors are not perfectly dependent on one another (full rank of matrix X). Under this set of hypotheses the Ordinary Least Squares fitting criterion (OLS) leads to the best linear unbiased estimators (*BLUE*) of the vector of parameters β, say $\hat{\beta}_{OLS} = \hat{\beta}$. In fact the OLS criterion requires:

$$S(\beta) = e^T e = \min \tag{1.5}$$

where $e = y - X\hat{\beta}$ are the observed errors and e^T indicates the transpose of e.

From Equation (1.5) we have:

$$\frac{d}{d\beta} S(\beta) = \frac{d}{d\beta} e^T e = \frac{d}{d\beta} (y - X\hat{\beta})^T (y - X\hat{\beta}) = 2\left(X^T X\hat{\beta} - X^T y\right) = 0$$

whence:

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y \tag{1.6}$$

As said the OLS estimator is *unbiased*

$$E(\hat{\beta}_{OLS} \,|\, X) = \beta \tag{1.7}$$

with a variance

$$Var(\hat{\beta}_{OLS} \,|\, X) = (X^T X)^{-1} \sigma^2_{\varepsilon} \tag{1.8}$$

which achieves the minimum among all possible linear estimators (full *efficiency*) and tends to zero when n tends to infinity (*weak consistency*).

From the assumption of normality of the stochastic disturbances, normality of the estimators also follows:

$$\hat{\beta}_{OLS} \,|\, X \approx N[\beta;\ (X^T X)^{-1}\sigma_\varepsilon^2] \tag{1.9}$$

Furthermore, from the assumption of normality of the stochastic disturbances, it also follows that the alternative estimators, based on the Maximum Likelihood criterion (ML), coincide with the OLS solution.

In fact, the single stochastic disturbance is distributed as:

$$f_{\varepsilon_i}(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\varepsilon_i^2\right]$$

f being a density function, and consequently the likelihood of the observed sample is:

$$\begin{aligned} L(\beta, \sigma_\varepsilon^2) &= \prod_{i=1}^{n} f_{\varepsilon_i}(\varepsilon_i) = \prod_{i=1}^{n} \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_\varepsilon^2}\varepsilon_i^2\right] \\ &= \left(\sigma_\varepsilon^2\right)^{-\frac{n}{2}} (2\pi)^{-\frac{n}{2}} \exp\left[-\frac{\varepsilon^T\varepsilon}{2\sigma_\varepsilon^2}\right] \\ &= \cos t \left(\sigma_\varepsilon^2\right)^{-\frac{n}{2}} \exp\left[-\frac{\varepsilon^T\varepsilon}{2\sigma_\varepsilon^2}\right] \end{aligned} \tag{1.10}$$

from the assumption of independence of the disturbances. From (1.1) we have that

$$\varepsilon = y - X\beta \tag{1.11}$$

hence (1.10) can be written as:

$$L(\beta, \sigma_\varepsilon^2) = \cos t \left(\sigma_\varepsilon^2\right)^{-\frac{n}{2}} \exp\left[-\frac{(y - X\beta)^T (y - X\beta)}{2\sigma_\varepsilon^2}\right] \tag{1.12}$$

and the log-likelihood as:

$$l(\beta, \sigma_\varepsilon^2) = \ln\left[L(\beta, \sigma_\varepsilon^2)\right] = \cos t - \frac{n}{2}\ln\left(\sigma_\varepsilon^2\right) - \frac{(y - X\beta)^T (y - X\beta)}{2\sigma_\varepsilon^2} \tag{1.13}$$

The scores functions are defined as:

$$\begin{cases} s(\beta) = \dfrac{d\ln\left[L(\beta, \sigma_\varepsilon^2)\right]}{d\beta} = -\dfrac{1}{\hat{\sigma}_\varepsilon^2}\left(X^T y - X^T X\hat{\beta}\right) = 0 \\ s(\sigma_\varepsilon^2) = \dfrac{d\ln\left[L(\beta, \sigma_\varepsilon^2)\right]}{d\sigma_\varepsilon^2} = -\dfrac{n}{2\hat{\sigma}_\varepsilon^2} + \dfrac{(y - X\hat{\beta})^T (y - X\hat{\beta})}{2\hat{\sigma}_\varepsilon^4} = 0 \end{cases} \tag{1.14}$$

and solving the system of $k+1$ equations, we have:

$$\begin{aligned}\hat{\beta}_{ML} &= (X^T X)^{-1} X^T y \\ \hat{\sigma}^2_{\varepsilon,ML} &= \frac{e^T e}{n}\end{aligned} \tag{1.15}$$

Thus, under the hypothesis of normality of residuals, the ML estimator of β coincides with the OLS estimator. The ML estimator of σ^2_ε on the contrary differs from the unbiased estimator $s^2_\varepsilon = \frac{e^T e}{n-k}$ and it is biased, but asymptotically unbiased.

To ensure that the solution obtained is a maximum we consider the second derivatives:

$$\begin{cases} \dfrac{d^2 l(\beta, \sigma^2_\varepsilon)}{d\beta^2} = \dfrac{ds(\beta)}{d\beta^2} = \dfrac{1}{\hat{\sigma}^2_\varepsilon} X^T X \\ \dfrac{d^2 l(\beta, \sigma^2_\varepsilon)}{d\sigma^{2^2}_\varepsilon} = \dfrac{ds(\sigma^2_\varepsilon)}{d\sigma^2_\varepsilon} = \dfrac{n}{2\hat{\sigma}^4_\varepsilon} \\ \dfrac{d^2 l(\beta, \sigma^2_\varepsilon)}{d\beta d\sigma^2_\varepsilon} = 0 \end{cases} \tag{1.16}$$

which can be arranged in the Fisher's Information Matrix:

$${}_{k+1}I_{k+1}(\beta, \sigma^2_\varepsilon) = \begin{bmatrix} \dfrac{1}{\hat{\sigma}^2_\varepsilon} X^T X & 0 \\ 0 & \dfrac{n}{2\hat{\sigma}^4_\varepsilon} \end{bmatrix} \tag{1.17}$$

which is positive definite.

The equivalence between the ML and the OLS estimators ensures that the solution found enjoys all the large sample properties of the ML estimators, that is to say: asymptotic normality, consistency, asymptotic unbiasedness, full efficiency with respect to a larger class of estimators other than the linear estimators, and invariance.

The OLS estimators also coincide with the Method of Moments estimators (MM). In fact consider the following moment condition:

$$\frac{1}{n} X^T e = E\left(X^T \varepsilon\right) = 0 \tag{1.18}$$

Solving Equation (1.18) gives:

$$\frac{1}{n}X^T(y - X\beta) = 0$$
$$\frac{1}{n}X^T y - \frac{1}{n}X^T X\beta = 0$$

and solving for β, we have:

$$\hat{\beta}_{MM} = \left(X^T X\right)^{-1} X^T y = \hat{\beta}_{OLS} = \hat{\beta}_{ML} \tag{1.19}$$

As for hypothesis testing, let us first consider the following system of hypotheses related to the single parameter β_i:

$$\begin{aligned} H_0&: \beta_i = 0 \\ H_1&: \beta_i \neq 0 \end{aligned} \tag{1.20}$$

where β_i is a generic element of the matrix β such that $\hat{\beta}_i \approx N[\beta_i, S^{ii}\sigma_\varepsilon^2]$, and S^{ii} is the i-th element in the main diagonal of matrix X^TX. A statistical test can be derived taking the difference between the value of β under the null and under the alternative hypotheses scaled by its standard deviation:

$$test = \frac{\hat{\beta}_i}{\sigma_\varepsilon\sqrt{S^{ii}}} \overset{H_0}{\approx} N(0,1) \tag{1.21}$$

This, however, is not a pivotal quantity unless we know the value of σ_ε^2. Since $\frac{(n-k)s_\varepsilon^2}{\sigma_\varepsilon^2} \approx \chi^2_{(n-k)}$ (with $\chi^2_{(n-k)}$ a chi-squared distribution with $n-k$ degrees of freedom) and using the independence between s_ε^2 and $\hat{\beta}$, we can build up the pivotal quantity:

$$test = \frac{\hat{\beta}_i}{s_\varepsilon\sqrt{S^{ii}}} \approx t_{(n-k)} \tag{1.22}$$

which can be used to test the null hypothesis. If we consider, instead, the multiple null hypothesis

$$\begin{aligned} H_0&: \beta_2 = \beta_3 = \cdots = \beta_k = 0 \\ H_1&: \beta_i \neq 0 \end{aligned} \tag{1.23}$$

we can test the significance of the model as a whole through the quantity:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \approx F(k-1; n-k) \tag{1.24}$$

which is distributed as an F with $k-1$ and $n-k$ degrees of freedom and where

$$R^2 = \frac{SSR}{SST} \tag{1.25}$$

In Equation (1.25), SSR represents the sum of squares of regression defined as $SSR = 1 - SSE = 1 - e^T e$ (SSE representing the sum of squares of errors), $SST = y^T y - n\bar{y}$ represents the total sum of squares and $\bar{y}$ the sample mean of y. R^2 is the so-called *coefficient of determination* ($0 < R^2 \leq 1$), a parameter that measures the degree of fit of the observed data to a linear function. The adjusted version of R^2, which takes into account the number of degrees of freedom of the regression, is given by:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k+1}(1-R^2) \tag{1.26}$$

Alternative measures of the degree of fit are the Akaike Information Criterion

$$AIC = \ln\left(\frac{e^T e}{n}\right) + \frac{2k}{n} \tag{1.27}$$

and the Schwartz (or Bayesian) Information Criterion

$$BIC = \ln\left(\frac{e^T e}{n}\right) + \frac{k \ln n}{n} \tag{1.28}$$

A further approach to hypothesis testing in regression (that could be applied to the system of hypotheses (1.20) and that will be employed later in this book) is based on the general testing procedure known as the *likelihood ratio*. The *likelihood ratio* of a parameter vector, say θ, is given by:

$$\lambda = \frac{L(\theta_0)}{L(\hat{\theta}_{ML})} = \frac{L(\theta_0)}{L(\hat{\theta})} \tag{1.29}$$

with the subscript *0* indicating the value of the parameters under the null. It represents the ratio between the likelihood function evaluated at the parameter's value under the null and the likelihood function at its maximum. A monotonic transformation of the test statistic λ does not change the inferential conclusions, so that it is more common to refer to the *likelihood ratio test* as the quantity:

$$LR = -2\ln(\lambda) = -2\left[l(\theta_0) - l(\hat{\theta})\right] \tag{1.30}$$

Expanding $l(\theta_0)$ as a Taylor series about $\hat{\theta}$ we obtain:

$$\begin{aligned} LR &= -2\left[\left(\theta_0 - \hat{\theta}\right)l'\left(\hat{\theta}\right) + \frac{1}{2}\left(\theta_0 - \hat{\theta}\right)^2 l''\left(\tilde{\theta}\right) + rest\right] \\ &= -2\left[\left(\theta_0 - \hat{\theta}\right)s\left(\hat{\theta}\right) + \frac{1}{2}\left(\theta_0 - \hat{\theta}\right)^2 ni\left(\tilde{\theta}\right) + rest\right] \end{aligned} \tag{1.31}$$

where $\tilde{\theta} \in (\hat{\theta}, \theta_0)$, s(.) is the score function and $ni(\tilde{\theta})$ the element of Fisher's Information Matrix. By definition $s(\hat{\theta}) = 0$, so that:

$$LR = n\left(\theta_0 - \hat{\theta}\right)^2 i\left(\theta_0\right) + o_p(1) \tag{1.32}$$

The approximation:

$$LR \approx W = n\left(\theta_0 - \hat{\theta}\right)^2 i\left(\theta_0\right) \tag{1.33}$$

is called the Wald test statistics. A further approximation of LR:

$$LR \approx LM = \frac{l'(\theta_0)^2}{ni\left(\theta_0\right)} \tag{1.34}$$

is called the "*Rao's score test statistics*" in statistics, but is better known in econometrics as the "*Lagrange multiplier test*". The three test statistics *LR*, *W* and *LM* are asymptotically equivalent and asymptotically distributed as a χ^2 with the number of degrees of freedom equal to the number of parameters to be estimated. With respect to the other two tests, the *LM* test has the advantage that it can be computed without previously obtaining the Maximum Likelihood estimation of the unknown parameters and that it does not require the specification of any explicit alternative hypothesis.

For a vector of parameters Equation (1.34) becomes:

$$LM = s(\theta_0)^T I(\theta)^{-1} s(\theta_0) \tag{1.35}$$

which is the general expression of the Lagrange Multiplier test that will be used later in this book. In the case of the linear regression, a simple way of testing hypotheses on the parameters is to define:

$${}_1s_{k+1}(\theta_0) = \left[{}_1s(\beta_0)_k \, , s\left(\sigma^2_{0,\varepsilon}\right) \right] \tag{1.36}$$

and

$${}_{k+1}I_{k+1}(\theta_0) = {}_{k+1}I_{k+1}\left(\beta_0, \sigma^2_{0,\varepsilon}\right) \tag{1.37}$$

and substituting (1.14) and (1.17) into (1.36) and (1.37) and both into (1.35), we obtain the *LM* test for hypotheses on the regression parameters.

Finally, a crucial hypothesis to be tested on the model is the hypothesis of the normality of the residuals on which all the previous testing strategies are grounded. A popular parametric procedure was introduced by Jarque and Bera (1987) who suggested building up a test of normality by testing jointly that the third and the fourth moments of the empirical distribution of residuals are not significantly different from those of the Gaussian distribution. The formal expression of the test is the following:

$$JB = \frac{n}{6}\left[SK^2 + \frac{(K-3)^2}{4} \right]$$

with $SK = \dfrac{1/n \sum_{i=1}^{n} (e_i - \bar{e})^3}{\left[1/n \sum_{i=1}^{n} (e_i - \bar{e})^2 \right]^{3/2}}$ the skewness and, respectively,

$K = \dfrac{1/n \sum_{i=1}^{n} (e_i - \bar{e})^4}{\left[1/n \sum_{i=1}^{n} (e_i - \bar{e})^2 \right]^2}$ the kurtosis of the residuals. Under the null of normality, this quantity can be shown to be distributed as a χ^2 with 2 degrees of freedom.

Example 1.1 Barro and Sala-i-Martin model of regional convergence

The Barro and Sala-i-Martin (1995) model of regional convergence expresses the growth rate of per capita GDP in one region in a certain moment of time (expressed as the logarithm of the ratio) as a linear function of the per capita GDP at the beginning of the period. If the slope in this linear model is negative, then those regions that are poorer at the beginning of the period experience higher growth rates and, conversely, those regions with the higher per capita GDP at the beginning of the period experience lower growth rates. This indicates convergence of the regions towards the same level of per capita GDP. We can express the model as:

$$\log \frac{y_{it}}{y_{i0}} = \alpha + \beta y_{i0} + \varepsilon_i \qquad t = 1, 2, \ldots, T$$

y_{it} = being the per capita GDP in year t and region i. The parameter $b = -\frac{\ln(1+\beta)}{T}$ represents the so-called "speed of convergence". The following table shows the per capita GDP in year 2000 and the growth of the real per capita GDP in the period 2000–08 as it was observed in the 20 Italian regions.

Region	Per capita GDP	Growth of GDP (2000–2008)	Region	Per capita GDP	Growth of GDP (2000–2008)
1. Piedmont	130	2.7	11. Marche	125	3.1
2. Aosta Valley	150	2.5	12. Latium	130	2.9
3. Lombardy	140	2.7	13. Abruzzo	100	4.0
4. Trentino-Alto Adige.	170	0.5	14. Molise	90	3.5
			15. Campania	110	2.1
5. Veneto	160	1.5	16. Puglia	95	3.0
6. Friuli Venezia Giulia	160	0.5	17. Basilicata	80	4.2
			18. Calabria	100	3.0
7. Liguria	135	2.0	19. Sicily	100	2.0
8. Emilia Romagna	145	1.6	20. Sardinia	110	2.4
9. Tuscany	135	2.2			
10. Umbria	130	3.2			

Source: http://sitis.istat.it/sitis/html/.

The estimation of the model using the OLS method leads to the following results

Parameter		Standard Error	t-test	p-value
Intercept	6.161369	0.731837	8.419	1.18e–07***
Slope	–0.029510	0.005752	–5.130	7.01e–05***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

R^2 = 0.5938 F-statistics = 26.32 (p-value=7.01e−05) = Wald-test
AIC = 42.95174 BIC = 45.93894 JB test = 1.0884 (p-value = 0.5803)

The t-tests show that both the intercept and the slope are significantly different from zero with the slope of the expected negative sign, while the F-test leads to the acceptance of the model with a significance of more than 99%. Furthermore, the coefficient of determination (R^2) indicates that the model explains 59.38% percent of the variability of the growth. Finally, the hypothesis of normality cannot be rejected with a probability of 0.5803.

Example 1.2 Zellner–Revankar revised Cobb–Douglas Production function

The popular Cobb–Douglas production function expresses the level of production as a function of the two inputs of production factors: capital and labor. The revised function introduced by Zellner and Revankar (1970) is based on the per capita values of both the output and the input variables obtained dividing the regional aggregates by the total number of establishments in each region. This specification allows economies of scale to be taken into account. The following table shows the data used for the empirical validation of the model referring to the US transportation equipment industry. They refer to "value added", "capital" and "labor" observed in 1957 in 25 US states. All values were originally expressed in millions of 1957 dollars.

State	Per capita Value Added	Per capita Capital	Per capita Labor	State	Per capita Value Added	Per capita Capital	Per capita Labor
Alabama	1855,118	55,941	463,985	Massachusetts	1404,244	89,227	229,163
California	2333,445	135,165	330,061	Michigan	7182,313	766,030	863,352
Connecticut	4484,870	257,870	805,675	Missouri	5216,680	262,720	678,648
Florida	192,795	22,421	65,688	New Jersey	2700,862	134,785	336,166
Georgia	4289,169	162,394	641,324	New York	2039,978	158,295	412,351
Illinois	2629,193	214,498	321,422	Ohio	4440,493	435,201	716,022
Indiana	3816,035	434,169	571,269	Pennsylvania	2650,554	147,313	421,253
Iowa	477,280	35,973	106,893	Texas	1712,380	73,818	356,260
Kansas	6506,776	136,316	1134,066	Virginia	2051,694	84,388	368,247
Kentucky	4030,581	168,161	387,097	Washington	3558,369	172,106	491,413
Louisiana	637,635	32,722	138,261	West Virginia	1513,333	102,867	270,867
Maine	363,790	24,284	79,877	Wisconsin	2462,754	154,937	371,958
Maryland	3219,085	136,016	537,535				

The generalized Cobb–Douglas function proposed by Zellner and Revankar (1970) can be specified in the following way:

$$\ln(y_i) = \beta_0 + \beta_1 \ln(k_i) + \beta_2 \ln(l_i) + \varepsilon_i$$

with y_i being the per capita value added in region i, k_i the per capita capital expenditure in region i and l_i the per capita labor expenditure in region i. The OLS estimation of the parameters leads to the results shown here below in the table:

Parameter		*Standard error*	*t-test*	*p-value*
β_0	*– 0.06707*	*0.21156*	*–0.317*	*0.754208*
β_1	*3.19780*	*0.83224*	*3.842*	*0.000885****
β_2	*5.35042*	*0.52584*	*10.175*	*8.8e-10****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

R^2 = 0.9243 Adjusted R^2 = 0.9174 F-statistics = 134.3 (p-value = 4.7678e – 13)

AIC = 43.4129 BIC = 48.28841 JB test = 5.6914 (p-value = 0.05809)

The p-values of the estimates show that the parameters are all highly significant with the only exception being the constant term. The model explains 92.43% of the total variability of the per capita value added. The F-test leads to the acceptance of the model and, finally, the null of normality can be accepted at the 5% significance level.

1.2 Non-sphericity of the disturbances

The fundamental hypothesis from which all previous results are derived is the one contained in Equation (1.4), referred to as the hypothesis of *sphericity of the errors*. It has a twofold implication:

(i) constant elements on the main diagonal of the variance-covariance matrix (*homoscedasticity*)
(ii) zero value of the off-diagonal elements of the variance-covariance matrix (absence of *autocorrelation*)

These are both unrealistic when dealing with spatial observations typically characterized by absence of homogeneity and presence of spatial correlation. It is easy to show that when one or both conditions are violated, the OLS estimators generally lose their optimality properties.

In this case, in fact, we have:

$$E(\varepsilon\varepsilon^T|X) = {}_nVC_n = \sigma_\varepsilon^2\Omega \tag{1.38}$$

with *VC* indicating the n-by-n variance-covariance matrix amongst the errors and Ω the corresponding correlation matrix.

In this case the OLS estimator is such that:

$$E\left(\hat{\beta}_{OLS}|X\right)=\beta \tag{1.39}$$

showing unbiasedness as before, but with variance now:

$$Var\left(\hat{\beta}_{OLS}|X\right)=\sigma_{\varepsilon}^{2}(X^{T}X)^{-1}X^{T}\Omega X(X^{T}X)^{-1}\neq\sigma_{\varepsilon}^{2}(X^{T}X)^{-1} \tag{1.40}$$

Depending on the values of the elements of the Ω matrix we may have that $Var(\hat{\beta}_{OLS})>\sigma_{\varepsilon}^{2}(X^{T}X)^{-1}$ (as generally happens) showing that OLS estimators are no longer of minimal variance.

In order to test the homoscedasticity, there are many possible alternatives that are all based on the ability of the researcher to propose plausible forms of violations. The Breusch and Pagan (1979) test proposes the generic form expressed through the equation $\sigma_i^2=\sigma^2 f(\alpha^T\mathbf{X}_i)$ with $\boldsymbol{\alpha}$ a vector of constants and $\mathbf{X}_i$ the vector of regressors for the *i-th* observation. In the case of homoscedasticity we have $\boldsymbol{\alpha}=0$ and the test can be obtained through a simple regression. In practice the *LM* test statistics can be expressed as:

$$BP=\frac{1}{2}\left[g^{T}X(X^{T}X)^{-1}X^{T}g\right] \tag{1.41}$$

with $\mathbf{g}$ the vector of the transformed disturbances defined as $g_i=\dfrac{e_i^2}{(\mathbf{e^T e}/n)-1}$. Under the null hypothesis of constant conditional variance, the test statistic is asymptotically distributed as a χ^2 with $k-1$ degrees of freedom. A popular alternative is the White test (White, 1980) that uses a consistent estimator of the variance-covariance matrix reported in Equation (1.40) to build up a test of departure from homescedasticity. Operationally, this implies calculating the test statistic, $WH=nR^2$ with R^2 the coefficient of determination of a regression where the square of the empirical disturbances, *e*, is explained in terms of all the independent variables and of all the cross-products between the independent variables. Under the null of homescedasticity *WH* is asymptotically distributed as a χ^2 with $k-1$ degrees of freedom. It should be remarked, however, that both the *BP* and the *WH* test require the independence of the residuals and so none of them can be used without simultaneously testing such a hypothesis.

When dealing with time series data, the hypothesis of absence of autocorrelation among the disturbances can be tested using the Durbin–Watson statistics (Greene, 2011), but testing it when using spatial data is

a more complicated issue that requires an ad hoc treatment which will be introduced in Chapter 2.

As previously mentioned, if the non-sphericity of the errors is detected, the *OLS* solution does not provide minimum variance estimators and the following alternative procedure should be adopted.

Let us consider a decomposition of the correlation matrix such that:

$$ {}_n\Omega_n = {}_nP_{nn}P_n^T \tag{1.42} $$

P being a non-singular matrix. From (1.42) we have

$$ \Omega^{-1} = P^{-T}P^{-1} \tag{1.43} $$

Let us now consider the following variable transformations:

$$ y^* = P^{-1}y \quad X^* = P^{-1}X \quad \varepsilon^* = P^{-1}\varepsilon \tag{1.44} $$

If we premultiply Equation (1.1) by P^{-1} we have:

$$ P^{-1}y = P^{-1}X\beta + P^{-1}\varepsilon $$

or

$$ y^* = X^*\ \beta + \varepsilon^* \tag{1.45} $$

Notice that the variance-covariance matrix of the errors is now:

$$ E(\varepsilon^*\varepsilon^{*T}) = E(P^{-1}\varepsilon\varepsilon^T P^{-T}) = \sigma_\varepsilon^2 P^{-1}\Omega P^{-T} \tag{1.46} $$

from (1.38) and, finally:

$$ E(\varepsilon^*\ \varepsilon^{*T}) = \sigma_\varepsilon^2 I $$

from (1.42). The errors in Equation (1.44), therefore, now satisfy the assumptions for the applicability of the OLS.

The OLS estimators of the parameters in the previous equation can thus be derived as:

$$ \begin{aligned} \hat{\beta}^*_{OLS} &= (X^{*T}\ X^*)^{-1}X^{*T}\ y^* \\ &= (X^T P^{-T}PX)^{-1}X^T P^{-T}P y \\ &= (X^T\Omega^{-1}X)^{-1}X^T\Omega^{-1}y = \hat{\beta}_{GLS} \end{aligned} \tag{1.47} $$

whose variance-covariance matrix is now given by:

$$Var(\hat{\beta}_{GLS}) = \sigma_\varepsilon^2 (X^T \Omega^{-1} X)^{-1} \tag{1.48}$$

The estimator $\hat{\beta}_{GLS}$ is called the Generalized Least Squares estimator (GLS) or Aitken estimator. Such an estimator enjoys all optimality properties of the OLS.

The GLS estimator can be shown to be a Maximum Likelihood estimator. In fact, let us consider the likelihood function in this situation:

$$L(\beta, \sigma_\varepsilon^2) = cost \left(\sigma_\varepsilon^2\right)^{-\frac{n}{2}} |\Omega|^{1/2} \exp\left[-\frac{(y - X\beta)^T \Omega^{-1} (y - X\beta)}{2\sigma_\varepsilon^2}\right] \tag{1.49}$$

$|\Omega|$ being the determinant of the correlation matrix. The log-likelihood consequently is:

$$\begin{aligned} l(\beta, \sigma_\varepsilon^2) &= \ln\left[L(\beta, \sigma_\varepsilon^2)\right] \\ &= cost - \frac{n}{2}\ln\left(\sigma_\varepsilon^2\right) - \frac{1}{2}\log|\Omega| - \frac{(y - X\beta)^T \Omega^{-1} (y - X\beta)}{2\sigma_\varepsilon^2} \end{aligned} \tag{1.50}$$

Maximizing Equation (1.50) with respect to the parameter β corresponds to maximizing only its last term since all other terms are constant with respect to β. Thus the ML estimators are the solution of the equation:

$$(y - X\beta)^T \Omega^{-1} (y - X\beta) = \min$$

By taking the derivatives and equating to 0 we obtain again:

$$\hat{\beta}_{ML} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} y = \hat{\beta}_{GLS} \tag{1.51}$$

as in Equation (1.47).

An unbiased estimator of σ_ε^2 can be obtained applying the OLS criterion to equation (1.45) producing:

$$s_\varepsilon^2 = \frac{e^{*T} e^*}{n - k} \tag{1.52}$$

with $e^* = P^{-1}y - P^{-1}X\beta^*$. Hence we have:

$$s_\varepsilon^2 = \frac{\left(P^{-1}y - P^{-1}X\beta^*\right)^T \left(P^{-1}y - P^{-1}X\beta^*\right)}{n-k} \tag{1.53}$$

which, after some algebra gives:

$$s_\varepsilon^2 = \frac{y^T\Omega^{-1}y - \beta^{*T}X^T\Omega^{-1}y}{n-k} \tag{1.54}$$

Under the previous assumptions, all the testing procedures on the single parameters and on the whole model are still valid. Notice, however, that all the expressions derived for the estimators and for the testing procedures give rise to computable expressions only if:

a) either the *VC* matrix of the errors is fully specified in such a way that various forms of heteroscedasticity and autocorrelation can be taken into account, or
b) all the elements of the VC matrix are known.

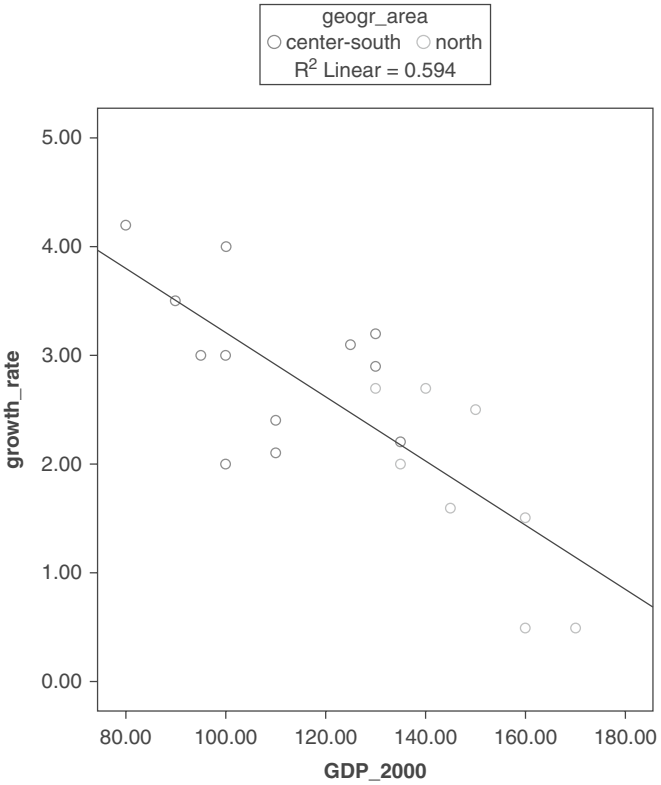
To satisfy condition (a) we can consider different specifications of the VC matrix that are able to capture heteroscedasticity and autocorrelation effects using a parsimonious number of parameters. This aim is achieved in the following chapters of this book. Condition (b) is only rarely realized in practical instances so that, in order to use the GLS criterion, we have to substitute the unknown parameters in the VC matrix with some consistent estimators and the consequences in terms of the properties of the estimators have to be considered in each particular circumstance.

Example 1.3 Barro and Sala-i-Martin model of regional convergence (continued)

Referring again to the data considered in Example 1.1 calculating the BP heteroscedasticity test we obtain

BP = 0.0045, p-value = 0.9462

which leads to the acceptance of the null hypothesis of homoscedasticity. Remember, however, that this test leads to a correct decision only if the residuals are also uncorrelated. Regarding the residuals correlation, as previously mentioned, we will need to discuss the topic in greater detail in Chapter 2. For the time being, the following figure shows the scatter diagram and the regression residuals of the 20 observations distinguishing northern and southern Italian regions. The graph does not depict any evident North–South geographical pattern, but it is obvious that further analysis is required.



Source: The author's creation using the data sourced from the ISTAT database called "Territorial indicators" that can be downloaded at the webpage: http://sitis.istat.it/sitis/html/.

1.3 Endogeneity

The OLS estimators possess their optimality properties only if the assumption expressed in Equation (1.3) is satisfied, that is when $E(\varepsilon|X)=0$. This condition can also be expressed as $E(\varepsilon X)=0$ (given that $E(X)$ is a constant) which indicates that the regressors X are uncorrelated with the disturbances, ε. This happens when either the disurbances are predetermined exogenously or they are innovations. Failing the condition expressed in Equation (1.3) some or all the regressors are said to be endogenous and the OLS estimators will, in general, be biased and inconsistent. Endogenous variables can be present in various cases in econometrics like the case of *errors in variables* or the case of *simultaneity*

(Greene, 2011). They also occur in some particular instances when dealing with spatial regressions as we will see later.

In this case an optimal estimator is derived with the procedure termed *Two-Stage Least Squares* introduced by Theil (1953) and Basmann (1957). The procedure assumes that it is possible to identify a set of, say h, variables, called *instruments*. Let us define the matrix ${}_nH_h$ containing the n observations of the h instrumental variables. A *valid* instrument possesses the requisites of being uncorrelated with the errors (it is *exogeneous*) and correlated with the regressors (it is *relevant*). At the ***first stage*** of regression, each column of the X matrix is regressed on the instruments H through the equation

$$ {}_nX_1 = {}_nH_{h\,h}\gamma_1 + {}_n\eta_1 \tag{1.55} $$

with η assumed to be i.i.d. N(0, σ_η^2). The OLS estimators of the model's parameters are derived as:

$$ \hat{\gamma} = \hat{\gamma}_{OLS} = (H^TH)^{-1}H^TX \tag{1.56} $$

At the **second stage** the dependent variable y is regressed not directly on X, but on X *instrumented* (as we say) with H. More precisely, y is regressed on the estimated value of X on the basis of model (1.55), that is to say on $\hat{X} = H\hat{\gamma} = H(H^TH)^{-1}H^TX = P_HX$, with $P_H = H(H^TH)^{-1}H^T$ the idempotent projection matrix of H.

Such a regression is given by the model:

$$ y = \hat{X}\beta + \varepsilon \tag{1.57} $$

and the OLS estimator of γ is given by:

$$ \hat{\beta}_{2SLS} = \left(\hat{X}^T\hat{X}\right)^{-1}\hat{X}^Ty \tag{1.58} $$

From the properties of the idempotent matrix, Equation (1.58) can be written as:

$$ \begin{aligned} \hat{\beta}_{2SLS} &= \left(\hat{X}^T\hat{X}\right)^{-1}\hat{X}^Ty = \left(X^TP_H^TP_HX\right)^{-1}X^TP_H^Ty \\ &= \left(X^TP_HX\right)^{-1}X^TP_Hy \\ &= \left(X^TH\left(H^TH\right)^{-1}H^TX\right)^{-1}X^TH\left(H^TH\right)^{-1}H^Ty \end{aligned} \tag{1.59} $$

Such an estimator is known as the Two-Stage Least Squares estimator (2SLS), the Instrumental Variable estimator (IV), or the Generalized

Instrumental Variable Estimator (GIVE). Equation (1.59) also emphasizes that such an estimator can also be viewed as a GLS estimator with a matrix of weights equal to P_H. At the time when it was introduced (in the 1950s) the estimation proceeded in the two steps described, but nowadays (with much higher computing power) the computation is achieved in one single step imposing exogeneity on the instruments. At a population level this leads to:

$$H^T \varepsilon = 0 \tag{1.60}$$

and at the sample level the condition becomes:

$$H^T(y - X\hat{\beta}) = 0 \tag{1.61}$$

and, solving this system of equations, we obtain:

$$\hat{\beta}_{MM} = \left(H^T X\right)^{-1} H^T y \tag{1.62}$$

which coincides with (1.59). Since Equation (1.60) can be viewed as a set of moments conditions, the estimator derived from its solution can also be seen as a Method of Moments estimator.

It can be shown that, under general conditions, the Two-stage Least Squares estimator is asymptotically normally distributed with

$$E\left(\hat{\beta}_{2SLS}\right) = \beta \tag{1.63}$$

and

$$Var(\hat{\beta}_{2SLS}) = \sigma_\varepsilon^2 (H^T X)^{-1} H^T H (H^T Z)^{-1} \tag{1.64}$$

and these expressions can be used for inference and hypothesis testing.

Example 1.4 Zellner–Revankar Production function (continued)

The following table shows the original data on which the Zellner–Revankar (1970) production function (discussed in Example 1.2) was estimated.

Using these data, only to provide an example and without the aim of contributing substantially to understanding the phenomenon, we estimate the production function model where the variable "Value added" is expressed as a function of the variable "labor input" only using the variable "number of establishments" as an instrument. The instrument is "relevant" in the sense that the variable "number of establishments" has a correlation of 0.839001 with the regressor "labor".

State	*Value Added*	*Capital Input*	*Labour Input*	*Number of Establishments*
Alabama	*126148*	*3804*	*31551*	*68*
California	*3201486*	*185446*	*452844*	*1372*
Connecticut	*690670*	*39712*	*124074*	*154*
Florida	*56296*	*6547*	*19181*	*292*
Georgia	*304531*	*11530*	*45534*	*71*
Illinois	*723028*	*58987*	*88391*	*275*
Indiana	*992169*	*112884*	*148530*	*260*
Iowa	*35796*	*2698*	*8017*	*75*
Kansas	*494515*	*10360*	*86189*	*76*
Kentucky	*124948*	*5213*	*12000*	*31*
Louisiana	*73328*	*3763*	*15900*	*115*
Maine	*29467*	*1967*	*6470*	*81*
Maryland	*415262*	*17546*	*69342*	*129*
Massachusetts	*241530*	*15347*	*39416*	*172*
Michigan	*4079554*	*435105*	*490384*	*568*
Missouri	*652085*	*32840*	*84831*	*125*
NewJersey	*667113*	*33292*	*83033*	*247*
New York	*940430*	*72974*	*190094*	*461*
Ohio	*1611899*	*157978*	*259916*	*363*
Pennsylvania	*617579*	*34324*	*98152*	*233*
Texas	*527413*	*22736*	*109728*	*308*
Virginia	*174394*	*7173*	*31301*	*85*
Washington	*636948*	*30807*	*87963*	*179*
West Virginia	*22700*	*1543*	*4063*	*15*
Wisconsin	*349711*	*22001*	*52818*	*142*

The result of the estimation procedure is shown in the following table

Parameter		*Standard Error*	*t-test*	*p-value*
β_0	*–55486.25*	*0.21156*	*48669.26*	*–1.140068*
β_2	*7.264460*	*0.329093*	*22.07415*	*0.000****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

R^2 = 0.970192 Adjusted R^2 = 0.968896 F-statistics = 487.2683 (p-value=0.00000)

AIC = BIC = JB test=26.31817 (p-value=0.000002)

BP test = 20.18672 (p-value=0.0002) Second stage SSR=8.25e+12

The estimation shows a good fit of the model to the empirical data. However, there is strong evidence of non-normality and heteroscedasticity in the residuals (JB and BP tests) which suggest that the model should be re-evaluated.

1.4 R Codes: running a regression

To begin, download the free *R* software from the website http://cran.r-project.org/ by choosing your appropriate operating system (Windows, Mac OSX or Linux).

In the *R* environment, we can input the data using the keyboard, defining the variables x, y and z as vectors in the following way:

x=c(x_1,x_2,…,x_n)
y=c(y_1,y_2,…,y_n)
z=c(z_1,z_2,…,z_n)

or, alternatively, we can read the data from an external file using one of the following commands according to the format of the source data file:

```
>read.table("filename.txt", header=T, dec=".")
>read.txt("filename.txt", header=T, dec=".")
>read.csv("filename.csv", header=T, dec=".")
```

in the case, respectively, of a text (`.txt`) or a comma separated values (`.csv`) input file.

In the case of simple linear regression, we can plot the data on a scatter diagram with the command:

```
> plot(x,y)
```

and we can estimate a simple regression model through the command

```
>model1 <-lm(y ~x)
```

"`model1`" being an arbitrarily assigned name and "`lm`" being the command corresponding to the acronym of Linear Modeling.

We can also add the fit line to the scatter plot with command:

```
>abline(model1)
```

In the case of a multiple linear regression, similarly, we estimate the model with the command:

```
>model2 <-lm(y ~x+z)
```

The following command:

```
>summary(model2)
```

returns the point estimates of the parameters, their significance, the R-squared, the adjusted R-squared and the F-test.

Similarly, the commands:

```
>AIC(model2)
>BIC(model2)
```

calculate the model *AIC* and *BIC* tests that are not included in the previous output.

Given the model estimation, we can calculate the confidence intervals of the parameters through the command:

```
>confint(model2)
```

For any further reference, the residuals are automatically stored with the name:

```
model2$residuals
```

For more diagnostics on the model, we need to install the packages "`lmtest`" and "`tseries`". To do so, type the commands:

```
> install.packages("lmtest")
> install.packages("tseries")
```

Once the two packages are installed, at the beginning of each session, we need to invoke them through the commands:

```
> library(lmtest)
> library(tseries)
```

In the packages "`lmtest`" and "`tseries`" we find several useful model diagnostics like, for example, the Breush–Pagan test of homoscedasticity (which uses all regressors to explain heteroscedasticity):

```
> bptest(model2)
```

and the Jarque–Bera normality test:

```
>jarque.bera.test(model2$residuals)
```

Key Terms and Concepts Introduced

- Linear model with independent disturbances
- Ordinary Least Squares estimation method (OLS)
- Best Linear Unbiased Estimator (BLUE)
- Properties of the parameter estimators: unbiasedness, efficiency, consistency, normality
- Maximum Likelihood estimation method (ML)
- Score function
- Fisher's information matrix
- Method of Moments estimation method (MM)
- Hypothesis testing on the model's parameters
- F-test on the model
- Coefficient of determination. R^2
- Adjusted R^2
- Akaike information criterion (AIC)
- Schwartz (or Bayesian) information criterion (BIC)
- Likelihood ratio test
- Wald test
- Rao's score test (or Lagrange multiplier)
- Jarque–Bera test
- Heteroscedasticity
- Breusch–Pagan test of homoscedasticity
- Autocorrelation
- Generalized Least Squares (or Aitken) estimation criterion (GLS)
- Exogeneity
- Two-Stage Least Squares estimation method (Generalized Instrumental Variable Estimator) (2SLS)
- Relevant instruments
- Exogenous instruments

Questions

1. Under what conditions are the OLS estimators equivalent to the ML estimators? Are these conditions likely to be encountered when dealing with regressions on spatial data? What are the implications of such equivalence? What are the implications of the lack of this equivalence?
2. Suppose we have a variable y = regional consumption, and variable x = regional income, both expressed in thousands of dollars and assume further the regression model $y = \beta x + \varepsilon$. What are the consequences on the OLS estimator of β (say $\hat{\beta}$) of expressing the income

in millions of dollars instead? What happens if the consumption is also expressed in millions of dollars?

3. What are the crucial elements to consider if we want to judge the goodness of fit of the empirically observed data to our theoretical model?
4. In Instrumental Variable (IV) estimation how can we prove that an instrument is *relevant*?
5. In time series regression modeling we test autocorrelation in the residuals using the Durbin-Watson test statistics defined by $d = \sum_{t=2}^{T}[e_t - e_{t-1}]^2 \Big/ \sum_{t=1}^{T} e_t^2 \;\; t=1,\ldots, T$. Such a test exploits the natural ordering of the time observations which is evident in the expression observing that the time index, t ranges between 1 and T. What elements do we have to take into account when testing residuals autocorrelation on spatial data?
6. Why do we have to use the GLS estimation procedure instead of the OLS procedure when estimating a regression model on data that do not satisfy the hypotheses of homoscedasticity or/and residuals uncorrelation?

Exercises

Exercise 1.1 The data shown in the table below refer to some regional economic data in the United Kingdom. They report the regional Gross Value Added (GVA) as a percentage of the country total, labor productivity (reported to a country total of 100) and the business birth rate.

Country	Region	GVA (% of UK)	Labor Productivity (UK=100)	Business Birth Rate (%)
Wales		3,6	81,5	9,3
Scotland		8,3	96,9	10,9
Northern Ireland		2,3	82,9	6,5
England	North of England	3,2	86,2	11,2
England	North West England	9,5	88,6	11,1
England	Yorkshire & Humberside	6,9	84,7	10,5
England	East Midlands	6,2	89,2	10,3
England	West Midlands	7,3	89,1	10,5
England	East Anglia	8,7	96,8	10,5
England	Greater London	21,6	139,7	14,6
England	South East England	14,7	108,3	10,8
England	South West England	7,7	89,8	9,6

Source: http://www.statistics.gov.uk/hub/index.html.

Using the *R* codes reported in section 1.4 do the following:

1. Estimate the model that explains the GVA as a function of both labor productivity and business birth rate (model 1).
2. Estimate two models that explain respectively the GVA as a function of labor productivity and business birth rate (model 2 and model 3).
3. Compare the results obtained for model 1, model 2 and model 3. Which is the preferred model in terms of fit to the empirical data? What elements did you take into consideration when choosing the preferred model?
4. Regress labor productivity on business birth rate (model 4).
5. Calculate the residuals of model 1.
6. Test the hypothesis of normality and of homoscedasticity of model 1 under the hypothesis that the innovations are not correlated.
7. Observe the geographical distribution of the residuals of model 1 with reference to the map shown in the figure here below. Do you notice any interesting geographical pattern?

Map of the 12 UK regions. (Courtesy of Carrie Dolan)

References to the Chapter

Barro, R. J. and Sala-i-Martin, X. (1995) *Economic Growth*. New York: McGraw Hill.
Basmann, R.L. (1957) 'A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation', *Econometrica,* 25, 1, 77–83.
Breusch, T.S., Pagan, A.R. (1979) 'A Simple Test for Heteroscedasticity and Random Coefficient Variation', *Econometrica,* 47, 5, 1287–94.
Greene, W. (2011) *Econometric Analysis*, 7th edition, Upper Saddle River, NJ: Pearson Education.
Jarque, C.M. and Bera, A.K. (1987) 'A Test for Normality of Observations and Regression Residuals', *International Statistical Review,* 55, 163–72.
Theil, H. (1953) *Repeated Least Squares Applied to Complete Equation Systems*, The Hague: Central Planning Bureau.
White, H. (1980) 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity', *Econometrica,* 48, 4, 817–38.
Zellner, A. and Revankar, N. (1970) 'Generalized Production Functions', *Review of Economic Studies*, 37, 241–50.

2
Some Important Spatial Definitions

2.1 The Spatial Weight Matrix *W* and the definition of Spatial Lag

If some autocorrelation is present in the stochastic disturbances, some or all the off-diagonal elements of the variance-covariance matrix are non-zero. In such a situation, as stated earlier, the optimal properties of the OLS are not valid and the GLS criterion can only be applied if we are able to specify a plausible form of autocorrelation. In view of this, the following chapters will consider various alternatives to model non-diagonal variance-covariance matrices when data are observed in geographical units such as countries or regions. In this section we will introduce some preliminary concepts.

When we observe a phenomenon in say, $i=1,\ldots,n$ regions, non-diagonal variance-covariance matrices arise from the presence of spatial autocorrelation among the stochastic terms. Positive spatial autocorrelation arises when units that are *close* to one another are more similar than units that are far apart. Similarly, *VC* matrices can also display spatial heterogeneity when some areas present more variability than others. As an example see Figure 2.1.

In the definition of spatial autocorrelation we mentioned the concept of *closeness* which requires some further specification. Indeed, the major difference between standard econometrics and spatial econometrics lies in the fact that, in order to treat spatial data, we need to use two different sets of information.

The first set of information relates to the observed values of the economic variables whereas the second set of information relates to the particular location where those variables are observed and to the various links of proximity between all spatial observations. The presence of

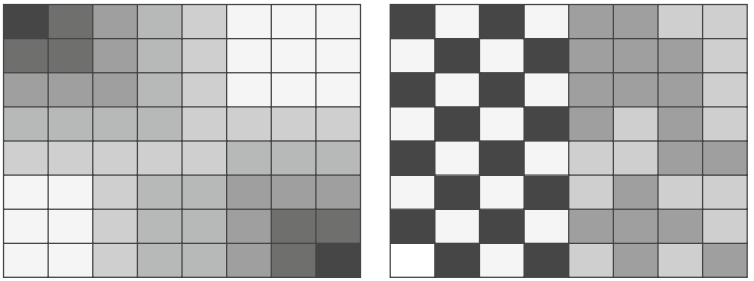


Figure 2.1 *Spatial autocorrelation* and *spatial heterogeneity* among 64 spatial units arranged in an 8-by-8 regular square lattice grid. Different greytones refer to different values of the variables under study ranging from low values (white) to high values (black). (a) Spatial autocorrelation. (b) Spatial heterogeneity. Left pane: high variability. Right pane: low variability

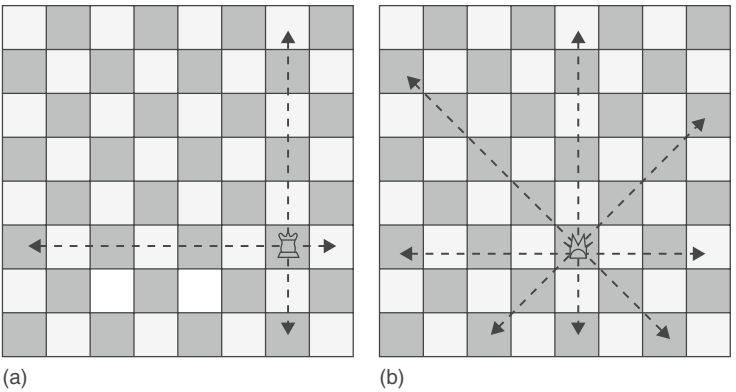


Figure 2.2 Contiguity criteria in a regular square lattice grid. (a) Rook's move and (b) Queen's move

this extra set of information related to space is also the reason why the standard econometric and statistical packages (for example, Eviews or SPSS) are so reluctant to introduce dedicated modules to spatial econometrics and spatial statistics which require extra capabilities in order to deal with spatial maps. If data are observed on a regular square lattice grid, like the one shown in Figure 2.1, *closeness* can be straightforwardly defined by choosing between the so-called *rook criterion* (two units are close to one another if they share a side) or the *queen criterion* (two units are close to one another if they share a side or an edge), drawing on the chess move analogues illustrated in Figure 2.2.

However, in spatial econometrics, almost invariably, we have to deal with irregularly spaced administrative units, such as regions or countries, so that further definitions are required.

At the heart of spatial econometrics methods is the definition of the so-called *weights matrix* (or *connectivity matrix*). The simplest of all definitions is the following:

$$ {}_nW_n = \begin{bmatrix} w_{11} & \dots & w_{n1} \\ \dots & w_{ij} & \\ & & \\ w_{1n} & & w_{nn} \end{bmatrix} \tag{2.1} $$

in which each generic element is defined as

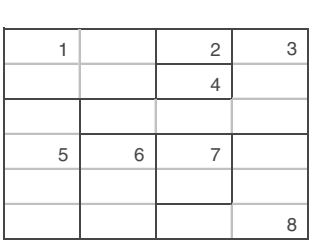
$$ w_{ij} = \begin{cases} 1 & \text{if } j \in N(i) \\ 0 & \text{otherwise} \end{cases} \tag{2.2} $$

$N(i)$ being the set of neighbors of location j. By definition we have that $w_{ii} = 0$.

Different concepts of the neighboring set $N(i)$ are possible, ranging from the one based on mere adjacency between the two territorial units illustrated in Figure 2.2, to those based on a *maximum distance* (that is $j \in N(i)$ if $d_{ij} < d_{\max}$, d_{ij} being the distance between location i and location j), to those based on the *nearest neighbor criterion*. More general $\boldsymbol{W}$ matrices can also be specified by considering the entries w_{ij} as (negative) functions of geographical, economic or social distances between areas rather than simply characterized by dichotomous entries like in Equation (2.2).

Example 2.1 Some examples of W matrices

Some simple examples of $\boldsymbol{W}$ *matrices for a system of irregular areas are shown here below. We consider a system of eight irregular regions (a) and the corresponding W matrix calculated with various criteria: (b) adjacency, (c) nearest neighbor, (d) distance < 2. Distances are measured between the centroids of the regions. A cell's side is conventionally set equal to 1. Notice that W matrices do not necessarily have to be symmetrical as happens, e. g., in case d). Notice, further, that conventionally we always have a 0 in the main diagonal in that each area is not considered a neighbor to itself. In the nearest neighbor criterion when more than one unit satisfies the condition, we select one randomly.*



(*a*)

	1	2	3	4	5	6	7	8
1	0	1	1	1	1	0	0	0
2	1	0	1	1	0	0	0	0
3	1	1	0	1	1	1	1	1
4	1	1	1	0	0	0	0	0
5	1	0	1	0	0	1	0	0
6	0	0	1	0	1	0	1	1
7	0	0	1	0	0	1	0	1
8	0	0	1	0	0	1	1	0

(*b*)

	1	2	3	4	5	6	7	8
1	0	1	0	0	0	0	0	0
2	0	0	0	1	0	0	0	0
3	0	0	0	1	0	0	0	0
4	0	1	0	0	0	0	0	0
5	0	0	0	0	0	1	0	0
6	0	0	0	0	0	0	1	0
7	0	0	0	0	0	1	0	0
8	0	0	0	0	0	0	1	0

(*c*)

	1	2	3	4	5	6	7	8
1	0	1	0	1	0	0	0	0
2	1	0	1	1	0	0	0	0
3	0	1	0	1	0	0	0	0
4	1	1	1	0	0	0	0	0
5	0	0	0	0	0	1	1	0
6	0	0	0	0	1	0	1	0
7	0	0	0	0	1	1	0	1
8	0	0	0	0	0	0	1	0

(*d*)

Quite often the W matrices are standardized to sum unity in each row. In this case we have:

$$w_{ij}^{*} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}; \quad w_{ij}^{*} \in W^{*} \tag{2.3}$$

This standardization may be very useful in some instances. For example, by using the standardized weights, we can define the matrix product,

$$L(y) = W^{*}y \tag{2.4}$$

in which each single element is equal to:

$$L(y_i) = \sum_{j=1}^{n} w_{ij}^{*} y_j = \sum_{j=1}^{n} \frac{w_{ij} y_j}{\sum_{j=1}^{n} w_{ij}} = \frac{\sum_{j \in N(i)} y_j}{\#N(i)} \tag{2.5}$$

with $\#N(i)$ representing the cardinality of the set $N(i)$. The term in Equation (2.5) represents the average of variable y observed in all the locations that are neighbors to location i (according to the criterion chosen in defining W). It therefore assumes the meaning of the *spatially lagged value* of y_i and for this reason is often indicated by the symbol $L(y)$ by analogy with the lag operator in time series analysis.

Example 2.2 UK regions contiguity-based W matrix

As an example, we can derive the W matrix for the 12 UK regions shown in Exercise 1.1. A contiguity-based version is shown here below.

	1	2	3	4	5	6	7	8	9	10	11	12	ROW SUM
	Scotland	N Ireland	Wales	N of England	NW England	Yorksh & Humber	W Midlands	E Midlands	E Anglia	SW England	SE England	G London	
1 Scotland	0	0	0	1	0	0	0	0	0	0	0	0	1
2 N Ireland	0	0	0	0	0	0	0	0	0	0	0	0	0
3 Wales	0	0	0	0	1	0	1	0	0	1	0	0	3
4 N of England	0	0	0	0	1	1	0	0	0	0	0	0	2
5 NW England	0	0	1	1	0	1	1	1	0	0	0	0	5
6 Yorksh & Humber	0	0	0	1	1	0	0	1	0	0	0	0	3
7 W Midlands	0	0	1	0	1	0	0	1	0	1	1	0	5
8 E Midlands	0	0	0	0	1	1	1	0	1	0	1	0	5
9 E Anglia	0	0	0	0	0	0	0	1	0	0	1	0	2
10 SW England	0	0	0	0	0	0	1	0	0	0	1	0	2
11 SE England	0	0	0	0	0	0	1	1	1	1	0	1	5
12 G London	0	0	0	0	0	0	0	0	0	0	1	0	1

Notice that an island (namely Northern Ireland) does not have any neighbors. The last column shows the number of neighbors for each region. The row-standardized version of the W matrix is shown here below. Notice that the row summation is now equal to 1.

		1	2	3	4	5	6	7	8	9	10	11	12	ROW SUM
		Scotland	**N Ireland**	**Wales**	**N of England**	**NW England**	**Yorksh & Humber**	**W Midlands**	**E Midlands**	**E Anglia**	**SW England**	**SE England**	**G London**	
1	*Scotland*	0	0	0	1	0	0	0	0	0	0	0	0	1
2	*N Ireland*	0	0	0	0	0	0	0	0	0	0	0	0	0
3	*Wales*	0	0	0	0	0,33333333	0	0,333333333	0	0	0,333333333	0	0	1
4	*N of England*	0	0	0	0	0,5	0,5	0	0	0	0	0	0	1
5	*NW England*	0	0	0,2	0,2	0	0,2	0,2	0,2	0	0	0	0	1
6	*Yorksh & Humber*	0	0	0	0,333333333	0,33333333	0	0	0,3333333	0	0	0	0	1
7	*W Midlands*	0	0	0,2	0	0,2	0	0	0,2	0	0,2	0,2	0	1
8	*E Midlands*	0	0	0	0	0,2	0,2	0,2	0	0,2	0	0,2	0	1
9	*E Anglia*	0	0	0	0	0	0	0	0,5	0	0	0,5	0	1
10	*SW England*	0	0	0	0	0	0	0,5	0	0	0	0,5	0	1
11	*SE England*	0	0	0	0	0	0	0,2	0,2	0,2	0,2	0	0,2	1
12	*G London*	0	0	0	0	0	0	0	0	0	0	1	0	1

The fact that Northern Ireland does not have any neighbors constitutes a problem if we want to compute a spatially lagged variable because, in this case, the value of the "spatially lagged" variable for Northern Ireland would always be zero. To tackle this problem we can conventionally consider the closest region (Scotland) as a neighbor of Northern Ireland even if, strictly speaking, the two regions do not share a common boundary. Having made this correction to the row-standardized ***W*** *matrix, it becomes:*

$$W = \begin{bmatrix} 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.33 & 0 & 0.33 & 0 & 0 & 0.33 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0.2 & 0 & 0.2 & 0.2 & 0.2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.33 & 0.33 & 0 & 0 & 0.33 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0.2 & 0 & 0 & 0.2 & 0 & 0.2 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0.2 & 0.2 & 0.2 & 0 & 0.2 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.2 & 0.2 & 0.2 & 0.2 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Now consider again the set of data shown in Exercise 1.1. If we premultiply the vector of the variable "labor productivity" (say variable y) by the (row-standardized) ***W**** *matrix we obtain the spatially lagged variable shown in the second column of the following table.*

y	***W*y***
81,5	*91,55*
96,9	*81,50*
82,9	*105,83*
86,2	*86,65*
88,6	*86,42*
84,7	*87,96*
89,2	*101,72*
89,1	*93,52*
96,8	*98,70*
139,7	*98,75*
108,3	*100,92*
89,8	*108,3*

2.2 Testing spatial autocorrelation among OLS residuals without an explicit alternative hypothesis

The most widely used test for spatial autocorrelation amongst OLS regression residuals is based on a general measure of spatial correlation introduced by Moran (1950) and proposed by Cliff and Ord (1972) as a test statistics for the null of uncorrelation among regression residuals. Notice that this statistics was introduced in the literature simultaneously to the analogous measure for time series regression residuals: the celebrated Durbin–Watson statistics (Durbin and Watson, 1950) even if, as already mentioned, its extension to deal with regression residuals was published only later (Cliff and Ord, 1972). Indeed the Durbin–Watson statistic can be defined as a special case of the Moran statistics by simply defining an appropriate $\boldsymbol{W}$ matrix (see, for example, Arbia, 2006). In its essence Moran's statistic takes the form of a correlation between the regression residuals and their spatially lagged values, that is:

$$Corr(\varepsilon, L\varepsilon) = \frac{Cov(\varepsilon, L\varepsilon)}{\sqrt{Var(\varepsilon)Var(L\varepsilon)}} \tag{2.6}$$

From Equation (2.6), by using the definition of spatial lag given in Equation (2.4) and assuming (by analogy with what happens with stationary time series) that:

$$Var(\varepsilon) = Var(L\varepsilon) \tag{2.7}$$

we have

$$Corr(\varepsilon, L\varepsilon) = \frac{Cov(\varepsilon, L\varepsilon)}{Var(\varepsilon)} = \frac{\varepsilon^T W\varepsilon}{\varepsilon^T\varepsilon} \tag{2.8}$$

It can be shown that, due to the nature of the spatial lag definition, equality (2.7) does not hold for spatial data where we have instead

$Var(\varepsilon) \geq Var(L\varepsilon)$ (see Arbia, 1989). One of the effects of this inequality is that the measure introduced in Equation (2.8) is not limited by 1 in absolute value, but possesses narrower limits given by $|I| \leq \sqrt{\frac{Var(L\varepsilon)}{Var(\varepsilon)}}$. However, partly for historical reasons, and more substantially for the equivalence that can be demonstrated with a Lagrange Multiplier test (see section 3.7), this is the definition that currently prevails in the literature and the one implemented in the software routines (alternatives are discussed in Whittle, 1954, in Cliff and Ord, 1972 and more recently in Li et al., 2007). In its original definition, Moran's I statistic considers the *biased* estimator of the variance in the denominator of Expression (2.8) and a normalizing factor for the numerator equal to the sum of the weights. As a consequence the empirical counterpart of (2.8) can be expressed as:

$$I = \frac{ne^T We}{e^T e\left[\sum_i \sum_j w_{ij}\right]} \tag{2.9}$$

When the weight matrix is row-standardized then $\sum_i \sum_j w_{ij} = n$ and the previous expression simplifies as:

$$I = \frac{e^T We}{e^T e} \tag{2.10}$$

Cliff and Ord (1972) derived the sampling distribution of the I statistic under two different hypotheses: (i) randomization and (ii) normality of residuals. In the first case the sampling distribution is obtained by considering all possible permutations of the observed data on the boundary system and calculating the Moran I statistic in each of them. They also proved that the asymptotical distribution is normal with an expected value that does not depend on the particular hypothesis chosen and that it is always expressed by:

$$E(I) = \frac{n\ tr(M_x W)}{S_0 (n-k)} \tag{2.11}$$

with $S_0 = \sum_i \sum_j w_{ij}$, $M_x = I - P_x$ and $P_x = X(X^TX)^{-1}X^T$. In contrast, its variance depends on the hypothesis selected. In particular, if we assume normality of the residuals, it can be expressed as:

$$Var(I) = \left(\frac{n}{S_0}\right)^2 \frac{tr(M_xWM_xW^T) + tr(M_xW)^2 + [tr(M_xW)]^2}{(n-k)(n-k+2)} - E(I)^2 \quad (2.12)$$

Notice that the Moran *I* test suffers from the limitation of not being based on an explicit alternative hypothesis. However, due to the already mentioned equivalence of the test (proved by Burridge, 1980) to an *LM* test, this is not a major drawback. The presentation of alternative test statistics for the hypothesis of residual correlation cannot be treated in more detail until we present some explicit formulations for an alternative hypothesis. This aim is accomplished in Chapter 3 and so we will revisit this issue again in section 3.7.

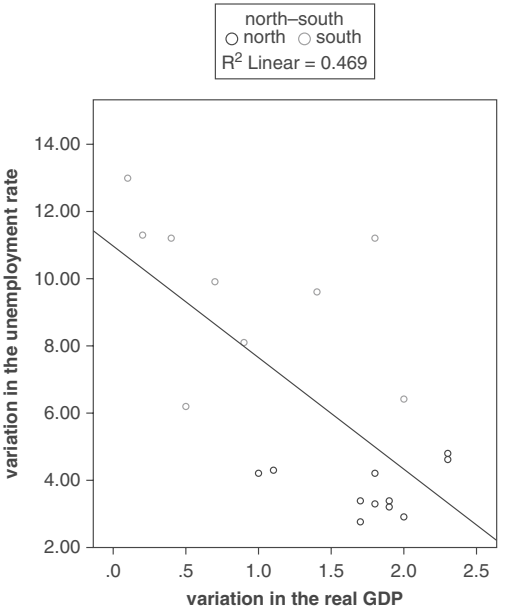
Example 2.3 Okun's law for the 20 Italian regions

Okun's Law (Okun, 1962) is an inverse relationship between the variation of the unemployment rate and the variation of the real GDP. The following table shows the data necessary to test Okun's Law on the 20 Italian regions. The variations of the two variables are observed in the period 1990–2010.

	Variation of Unemployment Rate	***Variation of Real GDP***		***Variation of Unemployment Rate***	***Variation of Real GDP***
1. *Piedmont*	*4.2*	*1*	11. *Marche*	*4.2*	*1.8*
2. *Aosta Valley*	*3.2*	*1.9*	12. *Latium*	*6.4*	*2*
3. *Lombardy*	*3.4*	*1.7*	13. *Abruzzo*	*6.2*	*0.5*
4. *Trentino Alto Adige*	*2.75*	*1.7*	14. *Molise*	*8.1*	*0.9*
			15. *Campania*	*11.2*	*0.4*
5. *Veneto*	*3.3*	*1.8*	16. *Puglia*	*11.2*	*1.8*
6. *Friuli Venezia Giulia*	*3.4*	*1.9*	17. *Basilicata*	*9.6*	*1.4*
			18. *Calabria*	*11.3*	*0.2*
7. *Liguria*	*4.8*	*2.3*	19. *Sicily*	*13*	*0.1*
8. *Emilia Romagna*	*2.9*	*2*	20. *Sardinia*	*9.9*	*0.7*
9. *Tuscany*	*4.3*	*1.1*			
10. *Umbria*	*4.6*	*2.3*			

Source: *http://sitis.istat.it/sitis/html/.*

The scatter diagram below shows the negative relationship expected from theory.



Source: The author's creation using the data sourced from the ISTAT database called "Territorial indicators" that can be downloaded at the webpage: http://sitis.istat.it/sitis/html/.

The graph shows that the variation in the unemployment rate is systematically higher than expected in the Southern Italian regions (light circles, corresponding to positive residuals) and lower in the Northern Italian regions (dark circles, corresponding to negative residuals), which could be interpreted as a possible model miss-specification and as a clear symptom of residual spatial autocorrelation.

The results of the estimation of the model with the OLS are shown here below together with the main test statistics.

Parameter		***Standard Error***	***t-test***	***p-value***
β_0	*10.971*	*1.283*	*8.8551*	*9.38e–08****
β_1	*–3.326*	*0.835*	*–3.984*	*0.000871****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

R^2 = 0.4686 Adjusted R^2 = 0.4391 F-statistics = 15.87 (p-value=0.0008705)

AIC = 98.28693 BIC = 101.2741 JB test =1.2331 (p-value=0.5398)

BP test = 0.0225 (p-value= 0.8808)

The F-test is highly significant and leads to the acceptance of the model. Furthermore both parameters are significantly different from zero at the usual confidence level. The JB and the BP tests lead to the acceptance of, respectively, the hypothesis of normality and homoscedasticity. The table here below summarizes the calculation of the Moran I test statistics for the hypothesis of spatial correlation of the residuals. Notice that the W matrix was specified by contiguity (the two islands are considered neighboring to the closest region).

Moran's I Test

	Observed Value	*Expected Value*	*Variance*	*z-test*	*p-value*
Moran's I	*0.40857021*	*–0.06968484*	*0.02737778*	*2.8904*	*0.001924*

Moran's I test reveals the presence of a significant positive spatial autocorrelation among the regression residuals, thus imposing a re-evaluation of all the results previously obtained and a redefinition of the model. In this particular case, in the presence of positive spatial autocorrelation, both the t-tests and the F-test will be inflated, leading us to accept as good models that should be rejected. Furthermore both JB and BP tests were not significant, thus leading to the acceptance of the hypothesis of homoscedasticity and normality. However, since we have detected significant spatial autocorrelation in the residuals both tests may lead to misleading conclusions.

Example 2.4 Phillips curve for the 20 Italian regions

The Phillips curve (Phillips, 1958) is an inverse relationship between the rate of unemployment and the rate of inflation. It states that lower unemployment is associated with a higher rate of inflation. Even if it was originally proposed to explain the historical behavior of the two variables, it was also considered to explain their spatial variations (Anselin, 1988). The following table shows the data necessary to test the Phillips model on the 20 Italian regions.

	% variation unemployment rate	*% variation price index*		*% variation unemployment rate*	*% variation price index*
1. Piedmont	*4.2*	*2.1*	*6. Friuli Venezia Giulia*	*3.4*	*1.8*
2. Aosta Valley	*3.2*	*1.4*	*7. Liguria*	*4.8*	*1.7*
3. Lombardy	*3.4*	*1.7*	*8. Emilia Romagna*	*2.9*	*1.9*
4. Trentino Alto Adige	*2.75*	*1.8*	*9. Tuscany*	*4.3*	*1.6*
5. Veneto	*3.3*	*1.5*	*10. Umbria*	*4.6*	*1.7*

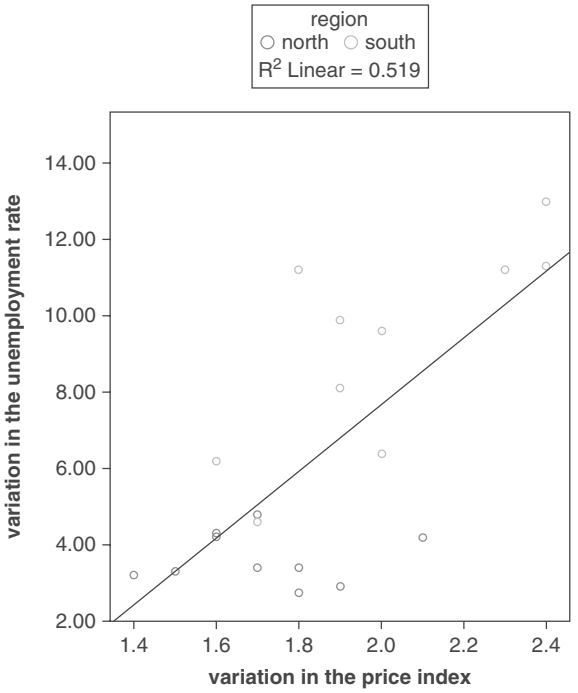
(*continued*)

Continued

	% variation unemployment rate	*% variation price index*		*% variation unemployment rate*	*% variation price index*
11. Marche	*4.2*	*1.6*	*16. Puglia*	*11.2*	*2.3*
12. Latium	*6.4*	*2*	*17. Basilicata*	*9.6*	*2*
13. Abruzzo	*6.2*	*1.6*	*18. Calabria*	*11.3*	*2.4*
14. Molise	*8.1*	*1.9*	*19. Sicily*	*13*	*2.4*
15. Campania	*11.2*	*1.8*	*20. Sardinia*	*9.9*	*1.9*

Source: *http://sitis.istat.it/sitis/html/.*

The scatter diagram below shows a positive relationship



Source: The author's creation using the data sourced from the ISTAT database called "Territorial indicators" that can be downloaded at the webpage: http://sitis.istat.it/sitis/html/.

However, the graph also shows that the variation in the unemployment rate is systematically higher than expected in the Southern Italian regions (light

circles, corresponding to positive residuals) and lower in the Northern Italian regions (dark circles, corresponding to negative residuals), which could possibly be an indication of residual spatial autocorrelation.

The result of the estimation of the model with the OLS are shown here below:

Parameter		*Standard Error*	*t-test*	*p-value*
β_0	–9.827	3.720	–2.642	0.016568*
β_2	8.746	1.984	4.409	0.000338***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

$R^2 = 0.5193$ Adjusted $R^2 = 0.4926$ F-statistics = 19.44 (p-value=0.000338)

AIC = 96.28177 BIC = 99.26897 JB test=0.0128 (p-value=0.9936)

BP test = 0.2556 (p-value=0.6131)

All tests of significance of the parameters lead to rejection of the null hypothesis, thus suggesting acceptance of the model. Similar indications come from the analysis of the F-test and of the tests of normality and homoscedasticity. By specifying a W matrix by contiguity (as in Example 2.3), the calculation of the Moran's I test of spatial autocorrelation among the residuals produce the following results:

Moran's I test

	Observed Value	*Expected Value*	*Variance*	*z-test*	*p-value*
Moran's I	0.3212607	–0.06938126	0.02711169	1.5297287	0.063042

In this case, although a positive spatial correlation is detected, this is not significant at the 5% level of confidence.

2.3 R Codes

The creation and the management of a ***W*** matrix is the trickiest part in running a spatial regression in any software. It is also what distinguishes software with spatial capabilities from standard econometric software. For this reason we will devote a significant part of the present chapter to discuss some of the most important steps needed for its creation, its import from external resources and its management. All the `R` procedures that will be illustrated in the present section, and most of those

presented in the rest of the book, are contained in the package `spdep`. To install the package, type the command:

```
>install.packages("spdep")
```

for the first time and then at the beginning of each new session, call it back by typing:

```
>library(spdep)
```

2.3.1 Creation of a *W* matrix for regular grid data

Consider, to start, the case of a regular square lattice grid of dimension – for example, 3-by-3. The software *R* generates the list of neighbors automatically with the command:

```
> Wnb<-cell2nb(3,3,type=" ")
```

where the type can be specified as either "`rook`" or "`queen`" according to the typology of neighborhood chosen.

The command indicates that we want to change our data from a cell system (`cell`) to (`2`) a list of neighbors (`nb`).

The object `Wnb` is just a list of neighbors. If we type `Wnb` we obtain a summary of the information contained in it (the number of regions, the number of proximity links, the average number of links and the number and the percentage of non-zero links). Once this object is created, we have to transform it into an actual matrix, say *W*, through the command:

```
> W<-nb2listw(Wnb)
```

The command indicates that we want to change our data from a list of neighbors (`nb`) to (`2`) a weight matrix (`listw`). In order to visualize the actual neighbors, type:

```
> W$weights
```

Once the weight matrix ***W*** is created, the spatially lagged variable of a variable X (say *WX*) can be easily obtained through the command

```
> WX<-lag.listw(W,X)
```

WX being just a conventional name assigned to this new variable.

2.3.2 Creation of a W matrix for irregular data

Consider now the case of an irregular set of regions and let us consider, as an example, the 20 Italian regions, already described in Examples 2.3 and 2.4, whose boundaries are shown in the following Figure 2.3.

Figure 2.3 Boundaries of the 20 Italian regions
Source: http://www.istat.it/it/archivio/44523.

In order to create a ***W*** matrix we need two different objects, namely:

1. an external file containing the list of neighbors for each region. This file has to be saved as a text file and named with the extension `.GAL`.
2. an internal variable containing only the region identifier (e.g. 1, 2, …, 20)

For the 20 Italian regions shown in the previous Figure 2.3, the first of the two files will have the following format:

First line

0 20 Italy ita_regions (this is the header: it starts with a mandatory 0 followed by the number of regions (20), by the name of the regional system (Italy) and by a variable containing the polygon identifier (ita_regions))

Second line

2 1 (the region identifier and the number of neigbors. This lines states that region 2 has only one neighbor)
1 (the polygon identifier of the neighbor. This line states that region 2 has only region 1 as a neighbor)

1, 4 (1 region has 4 neighbors)
2, 3, 7, 8 (the polygon identifiers of the 4 neighbors of region 1)

and so on. For the full file see below.

File*: Italy.GAL*

```
0 20 italy ita_regions
1 1
2
2 4
1 3 7 8
3 4
2 4 5 8
4 2
3 5
5 4
3 4 6 8
6 1
5
7 3
2 8 9
8 5
2 3 5 7 9
9 5
7 8 10 11 12
10 5
8 9 11 12 13
11 4
9 10 12 13
12 6
9 10 11 13 14 15
13 3
10 12 14
14 4
12 13 15 16
15 4
12 14 16 17
16 3
14 15 17
17 3
15 16 18
18 2
17 19
19 2
18 20
20 1
19
```

end of file

Regarding the second of the two files needed for the procedure, the polygon identifier is represented by the internal variable `ita_regions` which, in this case, can be created with command:

```
ita_regions<-c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,
15,16,17,18,19,20)
```

or read from an external file. Once we have created these two objects, they can be read in *R* through the command:

```
nbitaly<-read.gal("Italy.GAL", region.id=ita_regions)
```

Notice that the whole path of the `.GAL` file has to be fully specified so that it can be univocally identified by the procedure.

The object `nbitaly` thus obtained has to be transformed into an actual *W* matrix through the command (see section 2.3.1)

```
> witaly<-nb2listw(nbitaly)
```

If we further require a row-standardization of the *W* matrix, the previous command needs to be modified as follows:

```
> witaly<-nb2listw(nbitaly, style="W")
```

2.3.3 Reading a *W* matrix from external files

When the number of units becomes large, the method illustrated in section 2.3.2 of inputting the neighborhood information directly becomes rapidly unfeasible. In this case a *W* matrix can be obtained by reading external files generated by Geographical Information Systems (GIS) (see, for example, Burrough et al., 2014) and publicly available for many regional systems (see, for example, http://www.census.gov/geo/maps-data/data/tiger-line.html, for the US states and other US boundary systems or http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/popups/references/administrative_units_statistical_units_1 for the EU States and for some other world boundary systems).

In particular, the regional boundaries and their relationships can appear in the form of *shapefiles* which consist of three mandatory files identified by the extensions `.shp`, `.shx` and `.dbf`. The `.shp` file stores feature geometry such as the coordinates of polygon centroids and their

boundaries. The `.shx` file stores an index of the feature geometry and, finally, the `.dbf` file stores the attribute information of the features. To acquire the necessary information from external files, we need to develop three steps. In the first step we simply read the shapefiles from the external source into the *R* system. In step two we create a list of neighbors from the shapefiles. Finally, in step three we derive the *W* matrix from the list of neighbors using the same procedure illustrated in sections 2.3.1 and 2.3.2. These steps are outlined below.

Step 1: To import the shapefiles, (e.g. the shapefile, Italy consisting of the three files Italy.shp, Italy.shx and Italy.dbf), we use the command:

```
> italy<-readShapePoly("Italy", IDvar="ID")
```

where ID is the name of the variable containing the regional identifier code. Once the data are read into the *R* system, we can list the variables contained in the dataset with he command

```
>names(Italy)
```

we can see the plot of the borders with the command

```
>plot(Italy)
```

we can identify the centroids of each region with the command:

```
>coords<-coordinates(Italy)
```

and we can show the regional id's on the map with the command

```
>text (coords, label=sapply(slot(italy,"polygons"),
function(i) slot(i,"ID")))
```

Step 2: To calculate the **contiguity-based** neighbors' list, we use the command:

```
>contnb<-poly2nb(Italy,queen=T)
```

The queen criterion specified (see section 2.1) ensures that two regions are considered neighbors if they have a common boundary. The

command indicates that we want to change our data from a list of polygons (`poly`) to (`2`) a list of neighbors (`nb`).

However, this command does not tolerate the presence of isolated areas (like, for example, the case of the map shown in Figure 2.3) without any neighbor. We can force the command to include these areas by adding the option

```
>contnb<-poly2nb(Italy,queen=T,zero.policy=TRUE)
```

but in this case a *W* matrix will be generated with one or more areas with all zero in the corresponding line. The problem may be eliminated by generating a list of neighbors based on **minimum threshold distance** using the command:

```
> contnb <- dnearneigh(coordinates(Italy), 0, 380000,
longlat=F)
```

where 380000 is a conventional threshold distance (to be identified empirically in all practical cases) as measured in Great Circle kilometers (that is the distance along a path on a sphere).

Step 3: Finally, in the third step we can obtain the ***W*** matrix from the neighbors' list generated in the previous steps, by using the following command (already illustrated in Sections 2.3.1 and 2.3.2):

```
> W<-nb2listw(contnb, glist=NULL)
```

Again, as shown in the previous sections, if we wish to row-standardize the weights we will have to add the extra option:

```
> W<-nb2listw(contnb, glist=NULL, style ="W")
```

2.3.4 Computation of Moran's I for the residuals of an OLS regression

To compute the Moran I test on the residuals of a model previously estimated (say `model1`), use the command

```
> lm.morantest(model1, W)
```

which uses a *W* matrix contained in the object `W`, obtained, for example, through the procedures described in sections 2.3.1 to 2.3.3. By default

the randomization option and the one-sided test are considered. To change the default, introduce the option:

```
> lm.morantest(model1, W, randomization=FALSE, alternative
"two-sided")
```

which considers the hypothesis of normality and a two-sided alternative hypothesis of positive or negative spatial autocorrelation.

2.3.5 Some useful R databases

The package `spdep` contains some datasets that are very useful for additional practice. These datasets will be considered in the rest of this book for examples and practical exercises. In particular we will consider four databases called "`baltimore`", "`boston`", "`columbus`" and "`used.cars`".

For example, to access the data "`used.cars`" type the command:

```
>data(used.cars)
```

After downloading these data, your session contains two new objects. To visualize them type the command:

```
>ls()
```

As a result of this command you will see the two objects: (i) `used.cars` containing the actual data and (ii) `usa48_1960` containing the map information in the form a list of neighbors. To visualize the content of the data file type the command:

```
>str(used.cars)
```

which shows that in the database we now have two variables called `used.cars$tax.charges` and `used.cars$price1960`. Notice that the prefix `used.cars$` is always necessary when you have more than one dataset open in the active session.

The object `usa48_nb` is a list of neighbors from which we can generate the *W* matrix through the command:

```
W<-nb2listw(usa48.nb)
```

as seen previously.

Analogous procedures can be followed to download the other datasets (`baltimore`, `boston` and `columbus`) mentioned above.

Key Terms and Concepts Introduced

- Spatial autocorrelation
- Neighborhood
- Neighborhood criteria: rook's case and queen's case definition
- Neighborhood criteria: maximum distance criterion
- Neighborhood criteria: nearest neighbor criterion
- Weight (or connectivity) matrix
- Standardized weight matrix
- Spatial lag
- Moran's I test of spatial autocorrelation among regression residuals
- Moments of Moran's I test under randomization and under the hypothesis of normality

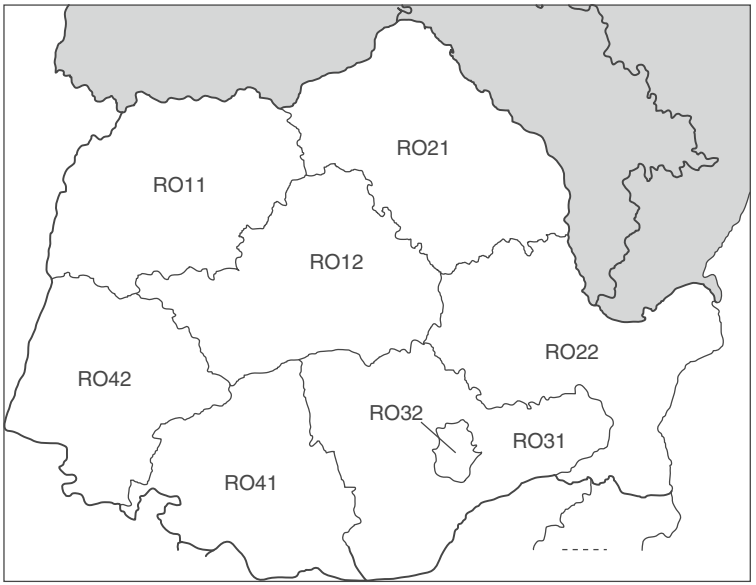
Questions

1. Why can't the Durbin–Watson test (available in all econometric software and used to test the hypothesis of no residual autocorrelation) be employed in the case of regressions estimated on spatial data?
2. What is the meaning of *spatially lagged variable*?
3. What is the meaning of *row-standardization* of a weight matrix? In which case is this operation beneficial?
4. What is the meaning of the *randomization* hypothesis used in deriving the sampling distribution of the Moran I test statistics?
5. In Question 5 of Chapter 1 we introduced the Durbin–Watson test statistics $d = \sum_{t=2}^{T}[e_t - e_{t-1}]^2 \Big/ \sum_{t=1}^{T} e_t^2$ $t=1,\ldots,T$. Rewrite the statistics in matrix notation, making use of an appropriate W matrix which properly describes the proximity between temporal units.

Exercises

Exercise 2.1 The Nomenclature of Units for Territorial Statistics (NUTS) is a geocoded standard for referencing the subdivision of European Union countries for statistical purposes. For each of the Member States, a hierarchy of three NUTS levels is established by Eurostat, level NUTS1

corresponding to the country subdivision. The following figure shows the boundaries of the eight Romanian NUTS2 regions. On the basis of this map derive the corresponding *W* matrix and its row-standardized version. Calculate the percentage of non-zero entries of the *W* matrix ("*sparsity*").



Map of the boundaries of the 8 regions of Romania at the NUTS2 European level. Source: http://epp.eurostat.ec.europa.eu/cache/GISCO/yearbook2007/NUTS2.pdf.

Exercise 2.2 Given the weight matrix derived in Exercise 2.1 and the data shown in the following table, compute the spatially lagged variable of infant mortality rates.

Regions	**Infant Mortality Rates (2011)**
RO11 – Nord-Vest	8.7
RO12 – Centru	10.1
RO21 – Nord-Est	10.1
RO22 – Sud-Est	11.3
RO31 – Sud - Muntenia	10.3
RO32 – Bucuresti - Ilfov	5.7
RO41 – Sud-Vest Oltenia	9.3
RO42 – Vest	8.9

Source: http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/data/database.

Exercise 2.3 (a) Using the procedure illustrated in section 2.3.1 generate a (rook's case based) weight matrix for the following regular square lattice grid of dimension 5-by-5 ($n=25$).

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

(b) Given the data arranged on the previously generated grid, now calculate the spatially lagged variable $L(X)$ of the following variable X.

27	16	–1	23	19
36	21	32	33	26
28	25	3	23	35
14	12	16	14	12
4	15	29	31	–1

Exercise 2.4 Given the 12 UK regions reported in Exercise 1.1, by using the procedure illustrated in section 2.3, create the `*.GAL` file and derive the weight matrix. Consider Northern Ireland to be adjacent to Scotland and Wales.

Exercise 2.5 Given the data shown in Exercise 1.1 and using the results of Exercise 2.4, compute the spatially lagged values of the variable GVA (Gross Value Added) for the 12 UK regions.

Exercise 2.6 On the basis of the results obtained in Exercise 2.5, draw a scatter diagram with the variable *GVA* on the horizontal axis and the lagged variable $L(GVA)$ on the vertical axis. This graph represents the exploratory tool termed "*Moran scatterplot*" in the literature (Anselin, 1995). What kind of insight can you derive from it?

Exercise 2.7 Given the results of Exercise 2.4 re-estimate model 1 of Exercise 1.1 [GVA=f(labor productivity; business birth rate)] and test for the presence of spatial autocorrelation among the residuals. Can we accept the hypothesis of residuals spatial uncorrelation?

Exercise 2.8 Using the procedure illustrated in section 2.3.2, generate a weight matrix for the 20 Italian regions and replicate Examples 2.3 and 2.4.

Exercise 2.9 Visit the website http://gis.cancer.gov/tools/seerstat_bridge/fips_vars/#statefips and download the shapefiles related to the 51 US states. Then, by using the procedure illustrated in section 2.3.3, generate the weight matrix and plot the map of the boundaries of the 51 states.

References to the Chapter

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publishers.

Anselin, L. (1995) 'Local Indicators of Spatial Association', *Geographical Analysis*, 27, 93–115.

Arbia, G. (1989) *Spatial Data Configuration in the Statistical Analysis of Regional Economics and Related Problems*, Dordrecht: Kluwer Academic Press.

Arbia, G. (2006) *Spatial Econometrics: Statistical Foundations and Applications to Regional Economic Growth*, Heidelberg: Springer-Verlag.

Burridge, P. (1980) 'On the Cliff–Ord Test for Spatial Correlation', *Journal of the Royal Statistical Society, B*, 42, 107–8.

Burrough P.A., Mcdonnell, R.A. and Lloyd, C.D. (2014) *Principles of Geographical Information Systems*, 3rd edition, Oxford: Oxford University Press.

Cliff, A.D. and Ord J.K. (1972) *Spatial Autocorrelation*, London: Pion.

Durbin, J. and Watson, G.S. (1950) 'Testing for Serial Correlation in Least Squares Regression', I, *Biometrika* 37, 409–28.

Kelejian, H.H. and Prucha, I. (2001) 'On the Asymptotic Distribution of the Moran I Test Statistic with Applications', *Journal of Econometrics,* 104, 219–57.

Li, H., Calder, C.A. and Cressie, N. (2007) 'Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model', *Geographical Analysis*, 39, 4, 357–75.

Moran, P.A.P. (1950) 'Notes on Continuous Stochastic Phenomena', *Biometrika* 37, 1, 17–23.

Okun, A.M. (1962) *Potential GNP, Its Measurement and Significance*, Cowles Foundation, Yale University.

Phillips, A.W. (1958) 'The Relationship Between Unemployment and the Rate of Change of Money Wages in the United Kingdom 1861–1957', *Economica,* 25, 100, 283–99.

Whittle, P. (1954) 'On Stationary Processes in the Plane', *Biometrika*, 41, 434–49.

3
Spatial Linear Regression Models

3.1 Generalities

This chapter discusses different specifications of linear spatial econometric models that can be considered once the hypothesis of no spatial autocorrelation in the disturbances is violated. A general form to take into account the violation of the ideal conditions for the applicability of OLS is given by the following set of equations:

$$y = \lambda W y + X\beta_{(1)} + WX\beta_{(2)} + u \qquad |\lambda| < 1 \tag{3.1}$$

$$u = \rho W u + \varepsilon \qquad |\rho| < 1 \tag{3.2}$$

with X a matrix of non-stochastic regressors, W a weight matrix exogenously given, $\varepsilon | X \approx i.i.d.N(0, \sigma^2_{\varepsilon\, n} I_n)$ and $\beta_{(1)}$, $\beta_{(2)}$, λ and ρ parameters to be estimated. The restrictions on the parameters λ and ρ hold if W is row-standardized.

The first equation considers the spatially lagged variable of the dependent variable y as one of the regressors and may also contain spatially lagged variables of some or all of the exogenous variables (the term WX). The second equation considers a spatial model for the stochastic disturbances. In principle, there is no need that the three weight matrices in Equations (3.1) and (3.2) are the same, although in practical cases it is difficult to justify a different choice.

Equation (3.1) can also be written as:

$$y = \lambda W y + Z\beta + u \qquad |\lambda| < 1 \tag{3.3}$$

having defined the matrix of all regressors, current and spatially lagged, as $Z = [X, WX]$ and the vector of regression parameters as $\beta = \left[\beta_{(1)}, \beta_{(2)}\right]$.

This model was termed SARAR(1,1) (acronym for Spatial AutoRegressive with additional AutoRegressive error structure) by Kelejian and Prucha (1998) and encompasses several spatial econometric models. In particular we have five remarkable cases:

(i) $\beta = 0$ and either λ or $\rho = 0$, known as the pure spatial autoregressive model
(ii) $\lambda = \rho = 0$, known as the Lagged independent variable model
(iii) $\lambda = 0$, $\rho \neq 0$ known as Spatial Lag Model (SLM)
(iv) $\lambda \neq 0$, $\rho = 0$ known as Spatial Error Model (SEM)
(v) $\lambda \neq 0$, $\rho \neq 0$ the complete model (SARAR)

We will review these five cases in the following sections. Before doing this, however, let us consider a general condition on the model's parameters.

First of all notice that Equation (3.1) can also be written as:

$$\begin{aligned} (I - \lambda W)y &= X\beta_{(1)} + WX\beta_{(2)} + u \\ y &= (I - \lambda W)^{-1}\left[X\beta_{(1)} + WX\beta_{(2)} + u\right] \end{aligned} \tag{3.4}$$

and Equation (3.2) as:

$$u = (I - \rho W)^{-1}\varepsilon \tag{3.5}$$

provided that the two inverse matrices exist. Using the Gerschgorin (1931) theorem Kelejian and Prucha (1998) proved that, when the W matrix is row-standardized, both inverse matrices exist if $|\rho| < 1$ and $|\lambda| < 1$, hence the parameters' restriction reported in Equations (3.1) and (3.2).

3.2 Pure spatial autoregression

When $\beta = 0$ and either $\lambda = 0$ *or* $\rho = 0$, and further assuming $\varepsilon | X \approx i.i.d.N(0, \sigma^2_{\varepsilon\, n} I_n)$ and W non-stochastic, then the model reduces to a simple spatial autoregression that can be estimated via the ML procedure (Whittle, 1954).

In this case we have

$$y = \lambda W y + \varepsilon \qquad |\lambda| < 1 \tag{3.6}$$

when ρ = 0 or

$$y = \rho W y + \varepsilon \qquad |\rho| < 1 \tag{3.7}$$

when $\lambda = 0$ since, in this instance, $y = u$. In this case we can derive the likelihood in the following way. First of all from Equation (3.6) (or (3.7)) we have that:

$$(I - \rho W)y = \varepsilon$$

hence

$$y = (I - \rho W)^{-1}\varepsilon$$

so that

$$E(y) = 0 \tag{3.8}$$

and

$$E(yy^T) = \sigma_\varepsilon^2 (I - \rho W)^{-1}(I - \rho W^T)^{-1} = \sigma_\varepsilon^2 \Omega \tag{3.9}$$

Having assumed normality of the innovations, the likelihood function can therefore be expressed as:

$$L(\rho, \sigma_\varepsilon^2) = const\, \sigma_\varepsilon^2 |\Omega|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} y^T \Omega^{-1} y\right\} \tag{3.10}$$

Substituting the explicit expression for the matrix Ω reported in Equation (3.9), we can write:

$$\begin{aligned} L(\rho, \sigma_\varepsilon^2) = const \left(\sigma_\varepsilon^2\right)^{-\frac{n}{2}} &\left|(I - \rho W)^{-1}(I - \rho W)^{-T}\right|^{-\frac{1}{2}} \\ &\times \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} y^T \left[(I - \rho W)^{-1}(I - \rho W)^{-T}\right]^{-1} y\right\} \end{aligned} \tag{3.11}$$

and, finally, the log-likelihood can be expressed as

$$l(\rho,\sigma_\varepsilon^2) = const - \frac{n}{2}\ln(\sigma_\varepsilon^2) - \frac{1}{2}\ln\left|(I-\rho W)^{-1}(I-\rho W)^{-T}\right| - \frac{1}{2\sigma_\varepsilon^2} y^T\left[(I-\rho W)^{-1}(I-\rho W)^{-T}\right]^{-1} y \tag{3.12}$$

This expression is non-linear in the parameters and requires a numerical maximization.

Example 3.1 Spatial autoregression of price index in the 20 Italian regions

Let us consider again the example of the 20 Italian regions shown on the map in Figure 2.3, and consider the spatial distribution of the variation of the price index shown in Example 2.4. Here we wish to estimate a purely spatial autoregressive model with a constant term, $y = \beta_0 + \lambda Wy + \varepsilon$ (with y = variation of the price index) to test if the variation of the price index in one region affects the variations in the neighboring regions through a mechanism of inflation contagion. As a neighborhood criterion we considered the maximum threshold distance criterion in order to include the two islands that would be otherwise isolated employing a simple contiguity criterion. The results of the estimation of the spatial autoregressive model using the ML criterion (by numerical maximization of the log-likelihood) are shown below:

Parameter		*Standard Error*	*t-test*	*p-value*
β_0	1.841545	0.037795	48.724	2.2e–16

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

λ=–0.56091 LR test = 0.94661 (p-value=0.33058)
Log-likelihood = –2.065929 AIC = 10.132

The table shows the usual t-test for the intercept term and a likelihood ratio test for the spatial parameter λ. The results show that only the constant term is significantly different from zero and thus there is no significant geographical transmission of the price variations at a regional level.

3.3 The classical model with spatially lagged non-stochastic regressors

When $\lambda = \rho = 0$, if we further assume that $\varepsilon | X \approx i.i.d.N(0, \sigma^2_{\varepsilon\, n} I_n)$, that both X and W are non-stochastic and that the matrix of all regressors $Z = [X, WX]$ is full rank, then the model only possibly contains a spatial lag of some or all the independent variables. In this situation no particular estimation problem emerges and the model can be simply estimated using the OLS procedure.

3.4 The Spatial Error Model (SEM)

3.4.1 Introduction

When $\lambda = 0$ and $\rho \neq 0$ the model becomes:

$$y = Z\beta + u \tag{3.13}$$

$$u = \rho W u + \varepsilon \qquad |\rho| < 1 \tag{3.14}$$

with the regressors Z and the weights W non-stochastic. This model is referred to in the literature as the Spatial Error Model (SEM) (Anselin, 1988; Arbia, 2006; LeSage and Pace, 2009). If $\varepsilon | X \approx i.i.d.N(0, \sigma^2_{\varepsilon\, n} I_n)$ then we have that $u = (I - \rho W)^{-1}\varepsilon$ as in Equation (3.5) so that we can write:

$$\begin{aligned} &E(u) = 0 \\ &E(uu^T) = \sigma^2_\varepsilon (I - \rho W)^{-1}(I - \rho W^T)^{-1} = \sigma^2_\varepsilon \Omega \end{aligned} \tag{3.15}$$

a formulation that considers both heteroscedastic and autocorrelated error terms. In these circumstances the *GLS* procedure may be applied only if the value of the parameter ρ is known *a priori*, a circumstance which happens only very rarely in empirical cases. Notice that from Equation (3.14) we have

$$(I - \rho W)u = \varepsilon$$

and model (3.13) (3.14) can thus also be written as:

$$\begin{aligned} &(I - \rho W)y = (I - \rho W)Z\beta + (I - \rho W)u \\ &y = \rho W y + Z\beta - WZ\rho\beta + \varepsilon \end{aligned}$$

$$y = \rho W y + Z\beta - WZ\gamma + \varepsilon \tag{3.16}$$

with $\gamma = \rho\beta$ and one may think of estimating model (3.16) directly. However, two problems emerge. First of all, Equation (3.16) is overparametrized due to the restriction $\gamma = \rho\beta$. Secondly, the term Wy is correlated with the error term, thus producing endogeneity. To convince ourselves of this let us consider that, from Equation (3.16):

$$(I - \rho W)y = Z\beta - WZ\gamma + \varepsilon$$

and so

$$y = (I - \rho W)^{-1}(Z\beta - WZ\gamma) + (I - \rho W)^{-1}\varepsilon \tag{3.17}$$

so that the covariance between the lagged variable Wy and the error term can be expressed as:

$$\begin{aligned} E\left[(Wy)\varepsilon^T\right] &= E\Big[W(I - \rho W)^{-1}(Z\beta - WZ\gamma) \\ &\quad + W(I - \rho W)^{-1}\varepsilon\Big]\varepsilon^T = \\ &= W(I - \rho W)^{-1}(Z\beta - WZ\gamma)E(\varepsilon^T) \\ &\quad + W(I - \rho W)^{-1}E\left[\varepsilon\varepsilon^T\right] = \\ &= \sigma_\varepsilon^2 W(I - \rho W)^{-1} I \neq 0 \end{aligned} \tag{3.18}$$

So the error is endogeneous, in that it is correlated with the spatially lagged variable Wy. As a consequence of the endogeneity of the errors, the OLS procedure loses its optimal properties.

In principle, an instrumental variable procedure could have been adopted to accommodate endogeneity. However, Kelejian and Prucha (1998) proved that such a procedure is not consistent due to the fact that it is not possible to identify instruments for Wy which are linearly independent of the other two regressors, Z and WZ.

As a consequence, unless the parameter ρ is known, there are two viable estimation alternatives:

(i) Maximum Likelihood (*ML*), and
(ii) Feasible GLS (*FGLS*)

These two procedures will be discussed in the next two subsections.

3.4.2 Maximum Likelihood estimator

From Equation (3.13) we derive

$$u = y - Z\beta \tag{3.19}$$

and, since u is normally distributed with a variance-covariance matrix given by Equation (3.15), we can easily obtain the likelihood function given by:

$$L(\rho, \sigma_\varepsilon^2, \beta) = const\ \sigma_\varepsilon^2 |\Omega|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} u^T \Omega^{-1} u\right\} \tag{3.20}$$

Substituting the expression (3.19) into this last equation we obtain:

$$L(\rho, \sigma_\varepsilon^2, \beta) = const \sigma_\varepsilon^2 |\Omega|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} (y - Z\beta)^T \Omega^{-1} (y - Z\beta)\right\} \tag{3.21}$$

and, by substituting the explicit expression for the matrix Ω reported in Equation (3.15), we can write:

$$\begin{aligned} L(\rho, \sigma_\varepsilon^2, \beta) = const \left(\sigma_\varepsilon^2\right)^{-\frac{n}{2}} &\left|(I - \rho W)^{-1}(I - \rho W)^{-T}\right|^{-\frac{1}{2}} \\ &\times \exp\left\{-\frac{1}{2}(y - Z\beta)^T \right. \\ &\left. \times \left[(I - \rho W)^{-1}(I - \rho W)^{-T}\right]^{-1} (y - Z\beta)\right\} \end{aligned} \tag{3.22}$$

Finally, the log-likelihood can be expressed as

$$\begin{aligned} l(\rho, \sigma_\varepsilon^2, \beta) = const &- \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{1}{2} \ln\left|(I - \rho W)^{-1}(I - \rho W)^{-T}\right| \\ &- \frac{1}{2}(y - Z\beta)^T \left[(I - \rho W)^{-1}(I - \rho W)^{-T}\right]^{-1} (y - Z\beta) \end{aligned} \tag{3.23}$$

This expression corresponds to deriving the likelihood function of the regression model

$$y^* = Z^{*T}\beta + \varepsilon$$

with $\varepsilon | X \approx i.i.d.N(0, \sigma^2_{\varepsilon\, n} I_n)$, for the transformed variables:

$$y^*(\rho) = y - \rho W y$$
$$Z^*(\rho) = Z - \rho W Z$$

Lee (2004) formally proves the conditions that ensure that the *ML* estimators are consistent and asymptotically normal in model (3.13) and (3.14).

Equation (3.23) cannot be maximized analytically due to the high degree of non-linearity. It can, however, be maximized numerically in order to produce estimates of the parameters. We must, however, note that the computational procedures employed in the available software are all approximated in that they are based on pseudo-likelihood.

A problem in maximizing the log-likelihood is represented by the term $\ln|I - \rho W|$ in that the determinant has to be evaluated repeatedly for each value of the parameter ρ in a numerical search. If *n* is very large this operation may be demanding. A way out, suggested in the literature, consists of exploiting the so-called *Ord decomposition* (Ord, 1975):

$$\ln|I - \rho W| = \ln\left[\prod_{i=1}^{n}(1 - \rho\phi_i)\right] \tag{3.24}$$

where ϕ_i represents the *i*-th eigenvector of the weight matrix *W*. This decomposition enormously simplifies the computation but, if *n* is very large, it does not completely eliminate the accuracy problems because the spectral decomposition is also approximated in very large matrices, as noted by Kelejian and Prucha (1998). We will go back to these computational issues in more detail in Chapter 5. To tackle the problem of evaluating a log-determinant in very large samples, Kelejian and Prucha (1998) suggested an alternative estimation strategy which will be presented in the next section.

If ρ is known, then the *ML* estimators coincide with the *GLS* estimators presented in section 1.2 obtained by substituting the explicit expression (3.15) of the variance-covariance matrix into Equation (1.47).

3.4.3 Feasible GLS

Let us return to model (3.13) (3.14) reported here for simplicity

$$y = Z\beta + u \tag{3.13}$$

$$u = \rho W u + \varepsilon \quad |\rho| < 1 \tag{3.14}$$

A feasible GLS procedure (FGLS) can be obtained along the following steps (Kelejian and Prucha, 1998):

Step 1: first of all obtain a consistent estimate of β, say $\tilde{\beta}$
Step 2: use this estimate to obtain an estimate of u, say $\hat{u}$
Step 3: use $\hat{u}$ to estimate ρ in Equation (3.14), say $\hat{\rho}$
Step 4: use $\hat{\rho}$ to transform model (3.13) as

$$(I - \hat{\rho} W) y = (I - \hat{\rho} W) Z\beta + \varepsilon$$

Step 5: finally, since the transformed model now contains stochastic disturbances which satisfy the requisites, estimate β via OLS on the transformed data corresponding to the GLS procedure.

These steps will be now discussed in detail.

Step 1: As an estimator of β consider the OLS estimator of Equation (3.13):

$$\tilde{\beta} = \left(Z^T Z\right)^{-1} Z^T y \tag{3.25}$$

Kelejian and Prucha (1998) proved that this estimator is consistent.

Step 2: From Equation (3.13) we derive an estimate of the residuals:

$$\hat{u} = y - Z\tilde{\beta} \tag{3.26}$$

Step 3: To obtain a consistent estimator of the parameter ρ, Keleijan and Prucha (1998) suggested a GMM procedure introducing the following additional assumptions:

(i) $E(\varepsilon^4) < \infty$.
(ii) both matrix W and $(I - \rho W)^{-1}$ are "absolutely summable" in the sense that $\sum_{i=1}^{n} w_{ij} < c$ and $\sum_{j=1}^{n} w_{ij} < c$, with c a constant not depending on n, and similarly for $(I - \rho W)^{-1}$.

(iii) Q_z, Q_1 and Q_2 are non-singular matrices with $Q_z = \lim_{n\to\infty} Z^T Z$; $Q_1 = \lim_{n\to\infty} Z^T \Omega Z$ and $Q_2 = \lim_{n\to\infty} Z^T \Omega^{-1} Z$.

Assuming the previous conditions, let us write model (3.13) (3.14) in terms of scalar quantities. We have:

$$y_i = Z_i\beta + u_i$$

and

$$u_i = \rho \sum_{i=1}^{n} w_{ij} u_j + \varepsilon_i \tag{3.27}$$

Let us now define

$$\bar{u}_i = \sum_{i=1}^{n} w_{ij} u_j; \quad \bar{\bar{u}}_i = \sum_{i=1}^{n} w_{ij} \bar{u}_j \text{ and } \bar{\varepsilon}_i = \sum_{i=1}^{n} w_{ij} \varepsilon_j \tag{3.28}$$

Hence, from Equation (3.27), we have:

$$u_i - \rho \bar{u}_i = u_i - \rho \sum_{i=1}^{n} w_{ij} u_j = \varepsilon_i \tag{3.29}$$

$$\bar{u}_i - \rho \bar{\bar{u}}_i = \bar{\varepsilon}_i \tag{3.30}$$

Three moment conditions are derived in the following way. Squaring and averaging Equation (3.29) and (3.30) we have:

$$\frac{1}{n} \sum_{i=1}^{n} \left(u_i - \rho \bar{u}_i\right)^2 = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 = E\left(\varepsilon_i^2\right) \tag{3.31}$$

$$\frac{1}{n} \sum_{i=1}^{n} \left(\bar{u}_i - \rho \bar{\bar{u}}_i\right)^2 = \sum_{i=1}^{n} \frac{1}{n} \bar{\varepsilon}_i^2 = E\left(\bar{\varepsilon}_i^2\right) \tag{3.32}$$

Furthermore, multiplying Equation (3.29) by Equation (3.30) and averaging we derive a third moment condition:

$$\frac{1}{n} \sum_{i=1}^{n} \left(u_i - \rho \bar{u}_i\right)\left(\bar{u}_i - \rho \bar{\bar{u}}_i\right) = E\left(\varepsilon_i \bar{\varepsilon}_i\right) \tag{3.33}$$

Notice that:

$$E\left(\varepsilon_i^2\right)=\sigma_\varepsilon^2;\quad E\left(\bar{\varepsilon}_i^2\right)=\sigma_\varepsilon^2 tr\frac{W^TW}{n};\ E\left(\varepsilon_i\bar{\varepsilon}_i\right)=0 \tag{3.34}$$

as demonstrated by Kelejian and Prucha (1998). Furthermore substituting into Equations (3.31), (3.32) and (3.33) the empirical counterpart of u, say $\hat{u}$ (which was derived in Equation (3.26)) the three moment conditions, obtained by equating the theoretical moments to their empirical counterparts, become:

$$\begin{cases}\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(\hat{u}_i-\rho\hat{\bar{u}}_i\right)^2=\sigma_\varepsilon^2\\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(\hat{\bar{u}}_i-\rho\hat{\bar{\bar{u}}}_i\right)^2=\sigma_\varepsilon^2 tr\dfrac{W^TW}{n}\\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\left(\hat{u}_i-\rho\hat{\bar{u}}_i\right)\left(\hat{\bar{u}}_i-\rho\hat{\bar{\bar{u}}}_i\right)=0\end{cases} \tag{3.35}$$

with obvious notation, and, after simple algebraic manipulation:

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2+\rho^2\frac{1}{n}\sum_{i=1}^{n}\hat{\bar{u}}_i^2-\frac{2}{n}\rho\sum_{i=1}^{n}\hat{u}_i\hat{\bar{u}}_i=\sigma_\varepsilon^2\\ &\frac{1}{n}\sum_{i=1}^{n}\hat{\bar{u}}_i^2+\frac{1}{n}\rho^2\sum_{i=1}^{n}\hat{\bar{\bar{u}}}_i^2-\frac{2}{n}\rho\sum_{i=1}^{n}\hat{\bar{u}}_i\hat{\bar{\bar{u}}}_i=\sigma_\varepsilon^2 tr\frac{W^TW}{n}\\ &\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\hat{\bar{u}}_i+\frac{\rho^2}{n}\sum_{i=1}^{n}\hat{\bar{u}}_i\hat{\bar{\bar{u}}}_i-\rho\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\bar{u}}_i^2+\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\hat{\bar{\bar{u}}}_i\right)=0\end{aligned} \tag{3.36}$$

or:

$$\begin{aligned}&\rho^2\frac{1}{n}\sum_{i=1}^{n}\hat{\bar{u}}_i^2-\frac{2}{n}\rho\sum_{i=1}^{n}\hat{u}_i\hat{\bar{u}}_i-\sigma_\varepsilon^2=-\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2\\ &\frac{1}{n}\rho^2\sum_{i=1}^{n}\hat{\bar{\bar{u}}}_i^2-\frac{2}{n}\rho\sum_{i=1}^{n}\hat{\bar{u}}_i\hat{\bar{\bar{u}}}_i-\sigma_\varepsilon^2 tr\frac{W^TW}{n}=-\frac{1}{n}\sum_{i=1}^{n}\hat{\bar{u}}_i^2\\ &\frac{\rho^2}{n}\sum_{i=1}^{n}\hat{\bar{u}}_i\hat{\bar{\bar{u}}}_i-\rho\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\bar{u}}_i^2+\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\hat{\bar{\bar{u}}}_i\right)=-\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\hat{\bar{u}}_i\end{aligned} \tag{3.37}$$

The three equations reported in (3.37) can be written more compactly in matrix notation as:

$$A_1\varphi - A_2 = 0 \tag{3.38}$$

having defined:

$$A_1 = \begin{bmatrix} 1/n\sum \hat{\bar{u}}_i^2 & -2/n\sum \hat{u}_i\hat{\bar{u}}_i & -1 \\ 1/n\sum \hat{\bar{\bar{u}}}_i^2 & -2/n\sum \hat{\bar{u}}_i\hat{\bar{\bar{u}}}_i & -1/n\,tr(W^TW) \\ 1/n\sum \hat{\bar{u}}_i\hat{\bar{\bar{u}}}_i & -1/n\left(\sum \hat{\bar{u}}_i^2 + \sum \hat{u}_i\hat{\bar{\bar{u}}}_i\right) & 0 \end{bmatrix} \tag{3.39}$$

$$A_2^T = \left[-\frac{1}{n}\sum \hat{u}_i^2; -\frac{1}{n}\sum \hat{\bar{u}}_i^2 - \frac{1}{n}\sum \hat{u}_i\hat{\bar{u}}_i\right] \tag{3.40}$$

and

$$\varphi^T = \left[\rho^2, \rho, \sigma_\varepsilon^2\right] \tag{3.41}$$

Thus a consistent estimator of the parameters' vector can be obtained by solving Equation (3.38) with respect to φ obtaining:

$$\varphi = A_1^{-1}A_2 = \left(\hat{\rho}^2, \hat{\rho}, \hat{\sigma}_\varepsilon^2\right) \tag{3.42}$$

Step 4: Use the estimator $\hat{\rho}$, thus obtained, to estimate the elements of the correlation matrix Ω reported in Equation (3.15) as:

$$\hat{\Omega} = \left(I - \hat{\rho}W\right)^{-1}\left(I - \hat{\rho}W^T\right)^{-1} \tag{3.43}$$

Step 5: Finally, estimate the regression parameter β via *GLS* by substituting the variance-covariance matrix derived in Equation (3.43) into expression (1.47) thus obtaining:

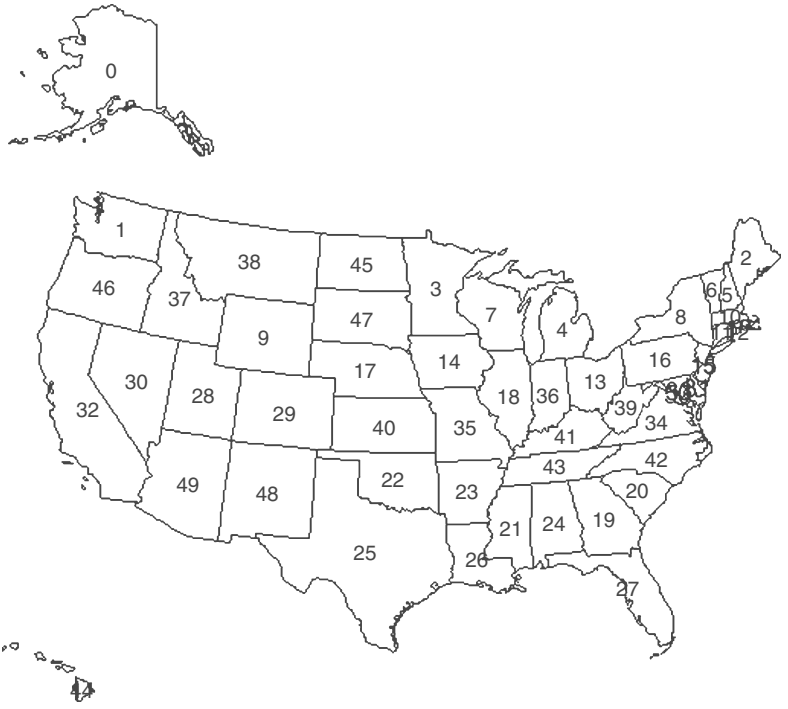
$$\hat{\beta}_{FGLS} = \left(Z^T\hat{\Omega}^{-1}Z\right)^{-1}Z^T\hat{\Omega}^{-1}y \tag{3.44}$$

This operation corresponds to applying the *GLS* method to the transformed data $(I - \hat{\rho}W)y = (I - \hat{\rho}W)Z\beta + \varepsilon$.

Example 3.2 The relationship between used car price and taxes in 48 US states

As an example of the Spatial Error model, let us consider the dataset related to the price of used cars by state in 1960 and of the tax and delivery charges for new cars by state in the period 1955–59. This dataset is very popular in spatial econometrics and can be easily downloaded in the R environment using the procedure illustrated in section 2.3.5.

The map of the 48 states considered is shown in the following figure (Alaska and Hawaii were discarded because they are isolated from the rest. The District of Columbia was also discarded for its exceptional features).



The following table shows the data.

n.	States	tax	price	n.	States	tax	price	n.	States	tax	price
1	*Washington*	129	1461	17	*Nebraska*	159	1547	33	*Maryland*	135	1466
2	*Maine*	218	1601	18	*Illinois*	139	1510	34	*Virginia*	171	1468
3	*Minnesota*	176	1469	19	*Georgia*	96	1572	35	*Missouri*	164	1627
4	*Michigan*	252	1611	20	*S. Carolina*	133	1509	36	*Indiana*	161	1502
5	*New Hampshire*	186	1606	21	*Mississippi*	82	1586	37	*Idaho*	174	1555
6	*Vermont*	154	1491	22	*Oklahoma*	159	1460	38	*Montana*	153	1465
7	*Wisconsin*	92	1536	23	*Arkansas*	136	1468	39	*W. Virginia*	172	1601
8	*New York*	150	1517	24	*Alabama*	196	1631	40	*Kansas*	133	1463
9	*Wyoming*	149	1481	25	*Texas*	97	1584	41	*Kentucky*	178	1511
10	*Massachusetts*	168	1659	26	*Louisiana*	220	1636	42	*N. Carolina*	257	1647
11	*Connecticut*	138	1515	27	*Florida*	96	1539	43	*Tennessee*	112	1559
12	*Rhode Island*	52	1460	28	*Utah*	89	1520	45	*N. Dakota*	93	1495
13	*Ohio*	195	1592	29	*Colorado*	185	1626	46	*Oregon*	265	1592
14	*Iowa*	141	1574	30	*Nevada*	115	1544	47	*S. Dakota*	105	1470
15	*New Jersey*	144	1418	31	*Delaware*	122	1477	48	*New Mexico*	58	1473
16	*Pennsylvania*	165	1509	32	*California*	153	1609	49	*Arizona*	188	1655

Let us start calculating, for the purpose of comparison, the OLS estimates of a simple regression model $y = \beta_0 + \beta_1 x + \varepsilon$, *where y = price in 1960 and x = tax and delivery charges in 1955–59. The results are shown here below together with the main test statistics.*

Parameter		***Standard Error***	***t-test***	***p-value***
β_0	*1435.7506*	*27.5796*	*52.058*	*2e–16****
β_1	*0.6872*	*0.1754*	*3.918*	*0.000294****

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

*F-test = 15.35 (p-value = 0.000294***)*
AIC = 528.3317 BIC = 533.945 JB test =1.8906 (p-value=0.3886)
BP test = 0.0013 (p-value=0.971)

The F-test is highly significant and leads to the acceptance of the model. Furthermore, both parameters are significant at the usual confidence level. Both the JB and the BP tests are not significant, thus leading to the acceptance of the two hypotheses of normality and homescedasticity. The table below summarizes the calculation of the Moran I test statistic for the hypothesis of spatial correlation of the residuals. The W matrix used is based on simple adjacency and is row-standardized.

Moran's I Test

	Observed Value	***Expected Value***	***Variance***	***z-test***	***p-value***
Moran's I	*0.574817771*	*– 0.030300549*	*0.008976437*	*6.38687*	*8.466e–11****

The test shows that there is evidence of a highly significant positive residual spatial correlation. All in all, the model is not satisfactory and, given the evidence of a residual positive and significant spatial correlation, we have clear indications of a Spatial Error model as an alternative framework. First of all, let us estimate the model using the maximum likelihood technique (see section 3.4.2). The results are shown here below:

Parameter		***Standard Error***	***t-test***	***p-value***
β_0	*1528.34521*	*31.96239*	*47.8170*	*2e–16****
β_1	*0.08831*	*0.11923*	*0.7406*	*0.4589*

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

ρ = 0.81899 LR-test = 40.899 (p-value = 1.603e–10) *Wald statistics = 122.32 (p-value = 2.22e–16)*

AIC = 489.43 *BIC = 496.9174* *JB test = 2.0845 (p-value = 0.3527)*

The tests now show that the regression coefficient related to the variable tax is not significant, while the parameter ρ is highly significant evaluated through the likelihood ratio test (see Equation (1.31)) and through the Wald test (see Equation (1.33)). So the spatial dependence among the residuals explains most of the model's variability and the model reduces to a pure autoregression (case of $\beta = 0$ and either ρ or λ also equal to 0. See Section 3.2) where the price in one country is explained just by the price in the neighboring countries.

To finish with, let us estimate the same Spatial Error model using now the Feasible Generalized Least Squares estimators discussed in section 3.4.3. The results are shown in the following table:

Parameter		***Standard Error***	***t-test***	***p-value***
β_0	*1512.98359*	*28.69940*	*52.7183*	*2e–16****
β_1	*0.17802*	*0.15018*	*1.1854*	*0.2359*
ρ	*0.65398*	*0.2184*	*3.1017*	*0.00096206****
σ^2	*1672.5*			

*Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

JB test = 2.85351 (p-value = 0.2423)

The FGLS estimates substantially confirm the conclusions of the ML estimation: the variable tax is not significant and the price of used cars in one country can be explained by a pure spatial autoregressive model. Notice, however, that the estimates of β_0 and ρ are different using the two procedures and, in particular, the value of ρ is larger if estimated through the ML procedure. Notice also that the ML procedure produces standard errors that are different from those produced by the FGLS alternative.

3.5 The Spatial Lag Model (SLM)

3.5.1 Generalities

When $\lambda \neq 0$ and $\rho = 0$ the model becomes:

$$y = \lambda Wy + Z\beta + u \qquad |\lambda| < 1 \tag{3.45}$$

with $u|X \approx i.i.d.N(0, \sigma^2_{u\,n} I_n)$. This model is referred to in the literature as the Spatial Lag Model (SLM) (Anselin, 1988; Arbia, 2006).

In this case, a problem of endogeneity emerges in that the spatially lagged value of y is correlated with the stochastic disturbance. In fact, using the same argument reported in Equation (3.18), we have that $(I - \lambda W)y = Z\beta + u$ and $y = (I - \lambda W)^{-1} Z\beta + (I - \lambda W)^{-1} u$ so that the correlation between the lagged term WY and the error can be expressed as:

$$\begin{aligned} E\left[(Wy)u^T\right] &= E\left[W(I - \lambda W)^{-1} Z\beta + (I - \lambda W)^{-1} u\right]u^T \\ &= W(I - \lambda W)^{-1} Z\beta E(u^T) + W(I - \lambda W)^{-1} E\left[uu^T\right] \\ &= \sigma^2_{\varepsilon} W(I - \lambda W)^{-1} I \neq 0 \end{aligned}$$

so, in the presence of endogeneity, a GLS procedure cannot be employed.

Two alternative estimators have been suggested in the literature:

(i) Maximum Likelihood
(ii) Two-Stage Least Squares (2SLS)

3.5.2 Maximum Likelihood Estimator

First of all, notice that re-writing Equation (3.45) as $(I - \lambda W)y = Z\beta + u$ we have:

$$y = (I - \lambda W)^{-1} Z\beta + (I - \lambda W)^{-1} u$$

so that:

$$E(y) = E\left[(I - \lambda W)^{-1} Z\beta + (I - \lambda W)^{-1} u\right] = (I - \lambda W)^{-1} Z\beta \tag{3.46}$$

and

$$E\left(yy^T\right) = \sigma_\varepsilon^2 (I - \lambda W)^{-1} (I - \lambda W)^{-T} = \sigma_\varepsilon^2 \Omega \tag{3.47}$$

Hence, the likelihood of y can be expressed as:

$$\begin{aligned} L(\sigma^2, \lambda, \beta; y) = const \left|\sigma_\varepsilon^2 \Omega\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left[y - (I - \lambda W)^{-1} Z\beta\right]^T \right. \\ \left. \times \Omega^{-1}\left[y - (I - \lambda W)^{-1} Z\beta\right]\right\} \end{aligned} \tag{3.48}$$

and, therefore, the log-likelihood as:

$$\begin{aligned} l(\sigma^2, \lambda, \beta; y) = const - \frac{1}{2}\ln\left|\sigma_\varepsilon^2 \Omega\right| - \frac{1}{2\sigma_\varepsilon^2}\left[y - (I - \lambda W)^{-1} Z\beta\right]^T \\ \times \Omega^{-1}\left[y - (I - \lambda W)^{-1} Z\beta\right] \end{aligned} \tag{3.49}$$

Using the expression reported in Equation (3.47), the determinant of the matrix $\sigma_\varepsilon^2 \Omega$ can be written as:

$$\left|\sigma_\varepsilon^2 \Omega\right| = \left|\sigma_\varepsilon^2 (I - \lambda W)^{-1} (I - \lambda W)^{-T}\right| = \sigma_\varepsilon^{2n} \left|(I - \lambda W)^{-1} (I - \lambda W)^{-T}\right|$$

and, since $\left|(I - \lambda W)^{-1} (I - \lambda W)^{-T}\right| = \left|(I - \lambda W)^{-1}\right| \left|(I - \lambda W)^{-T}\right|$, it can be also expressed as:

$$\left|\sigma_\varepsilon^2 \Omega\right| = \sigma_\varepsilon^{2n} \left|(I - \lambda W)\right|^{-2} \tag{3.50}$$

Let us now go back to the log-likelihood and substitute Equations (3.47) and (3.50) into Equation (3.49). We obtain:

$$\begin{aligned} l(\sigma^2,\lambda,\beta;y) &= const - \frac{1}{2}\ln\left(\sigma_\varepsilon^{2n}\left|I-\lambda W\right|^{-2}\right) \\ &\quad -\frac{1}{2\sigma_\varepsilon^2}\left[y-(I-\lambda W)^{-1}Z\beta\right]^T\left[(I-\lambda W)^{-1}(I-\lambda W)^{-T}\right]^{-1} \\ &\quad \times\left[y-(I-\lambda W)^{-1}Z\beta\right] \\ &= const - \frac{n}{2}\ln\sigma_\varepsilon^2 + \ln\left|I-\lambda W\right| - \frac{1}{2\sigma_\varepsilon^2}\left[y-(I-\lambda W)^{-1}Z\beta\right]^T \\ &\quad \times(I-\lambda W)^T(I-\lambda W)\left[y-(I-\lambda W)^{-1}Z\beta\right] \end{aligned} \tag{3.51}$$

and, since $(I-\lambda W)\left[y-(I-\lambda W)^{-1}Z\beta\right] = (I-\lambda W)y - Z\beta$, we eventually obtain:

$$\begin{aligned} l(\sigma^2,\lambda,\beta;y) &= const - \frac{n}{2}\ln\sigma_\varepsilon^2 + \ln\left|I-\lambda W\right| \\ &\quad -\frac{1}{2\sigma_\varepsilon^2}\left[(I-\lambda W)y - Z\beta\right]^T\left[(I-\lambda W)y - Z\beta\right] \end{aligned} \tag{3.52}$$

an expression that can be maximized numerically to obtain the estimators for the unknown parameters σ^2, λ and β.

3.5.3 Two-Stage Least Squares estimators

As an alternative to the *ML* estimators, in order to eliminate the endogeneity problem, we can also use a Two-Stage Least Squares strategy. To implement the method we need, first of all, to identify proper instruments that can eliminate the endogeneity problem arising from the spatially lagged term Wy. In other words, we need to identify instruments that are correlated with Wy (relevance) and uncorrelated with the error term (exogeneity).

Consider the fact that, from (3.46) we have:

$$E(y) = E\left[(I-\lambda W)^{-1}Z\beta + (I-\lambda W)^{-1}u\right] = (I-\lambda W)^{-1}Z\beta \tag{3.53}$$

now, since $|\lambda|<1$ we can expand the inverse matrix in (3.53) and write:

$$(I-\lambda W)^{-1} = I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 \ldots$$

Hence

$$\begin{aligned} E(y) &= \left[I + \lambda W + \lambda^2 W^2 + \lambda^3 W^3 \ldots\right] Z\beta \\ &= Z\beta + WZ\lambda\beta + W^2 Z\lambda^2\beta + \ldots \end{aligned} \tag{3.54}$$

so that $E(y)$ can be expressed as a linear function of $Z, WZ, W^2Z, \ldots$. This suggests the use of the first three elements of the expansion (3.54), that is to say Z, WZ, W^2Z, as relevant instruments to eliminate the endogeneity of Wy. Let us refer to this set of instruments as the n-by-$3k$ matrix ${}_nH_{3k} = \left[{}_nZ_k, {}_nW_n\,{}_nZ_k, {}_nW_n^2\,{}_nZ_k\right]$.

Let us now write equation (3.45) as follows:

$$y = M\theta + u \tag{3.55}$$

with the set of regressors ${}_nM_{k+1} = \left[{}_nW_n\,{}_ny_1, {}_nZ_k\right]$ and ${}_{k+1}\theta_1 = \left[\lambda, {}_k\beta_1\right]$ the vector of unknown parameters.

In the *first stage* of the two-stage procedure, the independent variables M are regressed on the instruments H through the instrumental regression:

$$M = H\gamma + \eta \tag{3.56}$$

with η an error term. The parameters in Equation (3.56) are then estimated via OLS producing:

$$\hat{\gamma} = (H^TH)^{-1}H^TM \tag{3.57}$$

by which we derive the estimated value of M, say $\hat{M}$, which is given by:

$$\hat{M} = H\hat{\gamma} = H(H^TH)^{-1}H^TM \tag{3.58}$$

In the *second stage* of the two-stage procedure, we estimate via OLS the relationship between y and the instrumented regressors, that is:

$$y = \hat{M}\theta + u \tag{3.59}$$

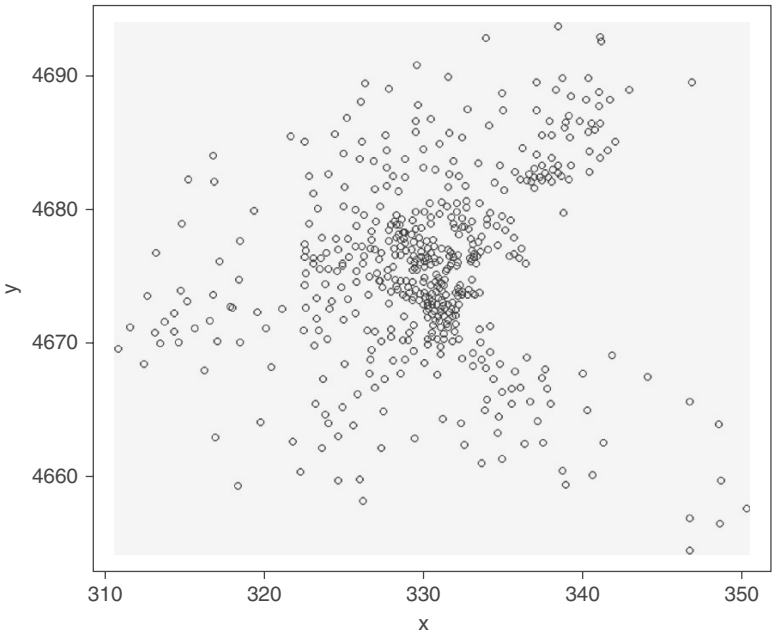
obtaining the two-stage estimators of the parameters θ given by:

$$\hat{\theta}_{2SLS} = \left(\hat{M}^T\hat{M}\right)^{-1}\hat{M}^T y \tag{3.60}$$

Example 3.3 House price determinants in Boston

Let us consider an example of the Spatial Lag model. The data we are using were collected by Harrison and Rubinfield (1978) and integrated by Gilley and Pace (1996) and are very popular in spatial econometrics. They are contained in the dataset Boston and can be downloaded through R *by using the procedure illustrated in section 2.3.5. The data refer to the median house price observed in 506 Boston area census tracts together with a series of variables that can be thought of as being potential determinants of the house value.*

The map of the 506 census tracts, represented through their centroids, is shown in the graph below, while the list of variables, contained in the database, is shown in the following table.



	Variable	*Variable Description*
1	*MEDV*	*Median value of owner-occupied housing expressed in thousands of USD*
2	*CRIM*	*Value of per-capita crimes*
3	*RM*	*Average number of rooms per dwelling*
4	*INDUS*	*Proportion of non-retail business acres per town*
5	*NOX*	*Value of the NOX (nitric oxides) concentration (parts per 10 millions) per town*
6	*AGE*	*Proportion of owner-occupied units built before 1940*
7	*DIS*	*Weighted distance from five Boston employment centers*
8	*RAD*	*Index of accessibility to radial highways per town*
9	*PTRATIO*	*Pupil to teachers ration per town*
10	*B*	*A transformed proportion of blacks*
11	*LSTAT*	*Percentage of lower status population*
12	*TAX*	*Full-value property tax per USD 10,000 per town*

Let us start calculating a simple regression model estimated via OLS. We wish to test if the price of the house can be expressed as a function of the 11 possible factors listed in the previous table. The results of a simple OLS estimation are shown here below, together with the test statistics.

Parameter	*Estimated Value*	*Standard Error*	*t-test*	*p-value*
Intercept	*37.308337*	*5.199690*	*7.175*	*2.66e–12****
CRIM	*–0.103402*	*0.033339*	*–3.102*	*0.002035***
RM	*4.074379*	*0.420639*	*4.074379*	*9.686 < 2e–16****
INDUS	*0.018212*	*0.062015*	*0.294*	*0.769138*
NOX	*–17.829176*	*3.889690*	*–4.584*	*5.79e–06 ****
AGE	*–0.002647*	*0.013353*	*–0.198*	*0.842957*
DIS	*–1.210182*	*0.186123*	*–6.502*	*1.94e–10****
RAD	*0.304603*	*0.066878*	*4.555*	*6.62e–06****
PTRATIO	*–1.131146*	*0.126079*	*–8.972*	*< 2e–16****
B	*0.009853*	*0.002735*	*3.603*	*0.000346 ****
LSTAT	*–0.525072*	*0.051543*	*–10.187*	*< 2e–16****
TAX	*–0.010901*	*0.003710*	*–2.939*	*0.003452***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

*F-test = 121 (p-value = 2.2e–16 ***)*
AIC = 3045.227 BIC = 3100.172 *JB test =936.7417 (p-value =2.2e–16***)*
*BP test = 59.2137 (p-value= 1.297e–08***)*

The F-test is highly significant and leads to the acceptance of the model. Furthermore all variables apart from INDUS and AGE are also significant at the usual confidence level. In particular, the variables RM, RAD and B are significantly positive, while CRIME, NOX, DIS, LSTAT and TAX present a negative sign. Notice that both the JB and the BP tests are significant, thus leading us to reject the hypotheses of normality and homoscedasticity. We employed a row-standardized distance-based weight matrix, considering neighbors as two sites if the distance between their centroids was less than 3.99 units in the previous graph.

The table here below summarizes the calculation of the Moran I test statistic for the hypothesis of spatial correlation of the residuals.

Moran's I test

	Observed value	*Expected value*	*Variance*	*z-test*	*p-value*
Moran's I	*0.0780022170*	*–0.0071438650*	*0.0001598831*	*6.7338*	*8.262e–12****

The test shows that there is evidence of a positive and highly significant spatial correlation in the regression residuals that motivates further analysis. In our case, since we are referring to very small areal units (the census tracts), it is certainly reasonable to speculate that the house value changes smoothly through space, in that expensive neighborhoods will tend to concentrate in certain zones of the city, thus displaying positive spatial correlation. For these reasons we can test if a Spatial Lag model achieves a better fit to our data while removing the residual correlation.

First of all, let us estimate the model using the Maximum Likelihood technique (Section 3.5.2). The results are shown below:

Parameter	*Estimated Value*	*Standard Error*	*t-test*	*p-value*
Intercept	*28.378*	*5.8225*	*4.8739*	*1.094e–06****
CRIM	*–0.097501*	*0.032606*	*–2.9902*	*0.0027877****
RM	*3.8432*	*0.41351*	*9.2941*	*< 2.2e–16****
INDUS	*–0.00071563*	*0.060617*	*–0.0118*	*0.9905805*
NOX	*–13.602*	*4.0537*	*–3.3555*	*0.0007921****
AGE	*0.0016953*	*0.013242*	*0.1280*	*0.8981255*
DIS	*–1.1782*	*0.18339*	*–6.4249*	*1.320e–10****
RAD	*0.29274*	*0.065501*	*4.4693*	*7.848e–06****
PTRATIO	*–0.97610*	*0.13042*	*–7.4845*	*7.172e–14****
B	*0.0098041*	*0.0026659*	*3.6776*	*0.0002354****
LSTAT	*–0.52343*	*0.050249*	*–10.4167*	*< 2.2e–16****
TAX	*–0.010491*	*0.0036223*	*–2.8962*	*0.0037769****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

λ = 0.22019 *LR-test= 12.492* *Wald statistics = 13.16*
*(p-value=0.00040864***)* *(p-value = 0.00028602***)*
AIC = 3034.7 *BIC =3093.906* *JB test = 1144.455*
*(p-value=2.2e–16***)*
*LM test for residuals = 13.341 (p-value = 0.00025969***)*

The t-tests show that, again as in the OLS case, all variables apart from INDUS and AGE are significant, with a sign which is in accordance with the OLS estimation. The regression coefficient related to the variable TAX is not significant, while both the Likelihood Ratio test and the Wald test show that the parameter λ is highly significant.

The Spatial Lag specification, however, while emphasizing the important feature of a spatial dependence of the house prices, has not completely removed the problem of residual spatial autocorrelation. In fact, the LM test for residuals shows that there is still some positive and significant residual correlation.

Let us, finally, estimate the same Spatial Lag model using the Two-Stage Least Squares technique presented in section 3.5.3. The results are displayed here below:

Parameter	***Estimated Value***	***Standard Error***	***t-test***	***p-value***
λ	*0.22047*	*0.068558*	*3.02158*	*0.0013008****
Intercept	*28.367*	*5.8399*	*4.8574*	*1.189e–06****
CRIM	*–0.097493*	*0.032978*	*–2.9563*	*0.0031137****
RM	*3.8429*	*0.04.2163*	*9.1145*	*< 2.2e–16****
INDUS	*–0.00073982*	*0.061532*	*–0.0120*	*0.9904070*
NOX	*–13.597*	*4.0608*	*–3.3484*	*0.0008129****
AGE	*0.0017009*	*0.013257*	*0.1283*	*0.8979109*
DIS	*–1.1782*	*0.18409*	*–6.4001*	*1.553e–10****
RAD	*0.29273*	*0.066155*	*4.4248*	*9.651e–06****
PTRATIO	*–0.97590*	*0.13355*	*–7.3073*	*2.727e–13****
B	*0.0098041*	*0.0027008*	*3.6300*	*0.0002834****
LSTAT	*–0.52343*	*0.050909*	*–10.2818*	*< 2.2e–16****
TAX	*–0.010490*	*0.0036662*	*–2.8614*	*0.0042179****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

*JB test = 1144.613 (p-value =2.2e–16***)*

The Two-Stage Least Squares substantially confirm the conclusions of the ML estimation in terms of both the sign and the significance of the variables.

In particular, it confirms that the proportion of non-retail business acres per town (INDUS) and the presence of old buildings in the area (AGE) have no significant effects on the median house price. Finally, the significance of the parameter λ is confirmed and its estimated value is very similar to the one estimated via Maximum Likelihood. Notice, however, that in contrast to the ML method, the 2SLS technique does not require the hypothesis of normality to be satisfied so that in this case (given the evidence of non-normality provided by the Jarque–Bera test) it provides more reliable estimates. Both estimation methods, however, lead to models that are not entirely satisfactory due to the presence of persistent positive significant correlation among the residuals. If in a spatial model the inclusion of the spatial lag is not enough to remove the non-sphericalness among the residuals it might be necessary to include some extra spatial components. This provides a scope to the theory which we will develop in the next section.

3.6 The general SARAR(1,1) Model

3.6.1 Generalities

To start with, let us consider the case where, in Equations (3.1) and (3.2), we set β = 0. We have:

$$y = \lambda W y + u \quad |\lambda| < 1 \tag{3.61}$$

$$u = \rho W u + \varepsilon \quad |\rho| < 1 \tag{3.62}$$

we thus have:

$$(I - \lambda W) y = u \qquad \mathrm{y} = (I - \lambda W)^{-1} u \tag{3.63}$$

and

$$(I - \rho W) u = \varepsilon \qquad \mathrm{u} = (I - \rho W)^{-1} \varepsilon \tag{3.64}$$

Combining (3.63) and (3.64) we have:

$$\mathrm{y} = (I - \lambda W)^{-1} (I - \rho W)^{-1} \varepsilon \tag{3.65}$$

and, as a consequence

$$\begin{aligned} E(yy^T) &= E\left[(I-\lambda W)^{-1}(I-\rho W)^{-1}\varepsilon\varepsilon^T(I-\lambda W)^{-T}(I-\rho W)^{-T}\right] \\ &= \sigma_\varepsilon^2(I-\lambda W)^{-1}(I-\rho W)^{-1}(I-\lambda W)^{-T}(I-\rho W)^{-T} \\ &= \sigma_\varepsilon^2\Omega \end{aligned} \tag{3.66}$$

so that the inverse of Ω is now:

$$\begin{aligned} \Omega^{-1} &= (I-\lambda W^T)(I-\rho W^T)(I-\rho W)(I-\lambda W) \\ &= \left[I-(\lambda+\rho)W^T+\lambda\rho W^TW^T\right] \\ &\quad\times\left[I-(\lambda+\rho)W^T+\lambda\rho W^TW^T\right]^T \end{aligned} \tag{3.67}$$

where the two parameters λ and ρ are present in the form of a sum and of a product and so they cannot be identified univocally. This fact has been considered in the literature to suggest that a complete model of the kind reported in Equations (3.61) and (3.62) is not feasible in practice. However, Kelejian and Prucha (1998) proved that this only happens when $\beta=0$ and it is not the case conversely when $\beta\neq 0$, which is what usually happens in the generality of cases of interests in spatial econometrics. In this case we can define a more general spatial model which encompasses the Spatial Lag and the Spatial Error models previously discussed in sections 3.4 and 3.5. This model, as already said, was termed a SARAR(1,1) model by Kelejian and Prucha (1998), but is also referred to in the literature as the *General Spatial Model* by Anselin (1988) or as an *SAC* model by LeSage and Kelly (2009).

If we consider the general SARAR model we thus have

$$y = Z\beta + \lambda Wy + u \qquad |\lambda| < 1 \tag{3.68}$$

$$u = \rho Wu + \varepsilon \qquad |\rho| < 1 \tag{3.69}$$

with $\varepsilon\,|X \approx i.i.d.N(0,\sigma_{\varepsilon\,n}^2 I_n)$.

Model (3.68) and (3.69) presents two major estimation problems. First, similar to the case of the *SLM* examined in the previous section, we have a problem of endogeneity associated to the presence of the lagged term *Wy*. Secondly, due to the presence of the autoregression in the stochastic disturbance in Equation (3.69), we cannot employ a GLS

strategy unless the parameter ρ is known. In this case we can exploit the following estimation alternatives:

(i) Maximum Likelihood
(ii) A spatial version of the Two-Stage Least Squares (GS2SLS)
(iii) Lee's Instrumental Variable estimators (LIV)

Maximum likelihood is feasible, but presents the limitation that there is currently no formal proof that the estimators possess the usual optimal large sample properties. The GS2SLS estimators are not fully efficient. The LIV estimator achieves more efficient estimators than GS2SLS, even if the gain in efficiency is only limited. The three alternative estimators will be now discussed in turn.

3.6.2 Maximum Likelihood Estimator

Let us consider again the full model contained in Equations (3.1) and (3.2)

$$y = Z\beta + \lambda W y + u \quad |\lambda| < 1 \tag{3.70}$$

$$u = \rho W u + \varepsilon \quad |\rho| < 1 \tag{3.71}$$

From Equation (3.70) we have

$$\mathrm{E}(\mathrm{y}) = (I - \lambda W)^{-1} Z\beta \tag{3.72}$$

and also

$$\begin{aligned} E(yy^T) &= E\left[(I-\lambda W)^{-1}(I-\rho W)^{-1}\varepsilon\varepsilon^T (I-\lambda W)^{-T}(I-\rho W)^{-T}\right] \\ &= \sigma_\varepsilon^2 (I-\lambda W)^{-1}(I-\rho W)^{-1}(I-\lambda W)^{-T}(I-\rho W)^{-T} = \sigma_\varepsilon^2 \Omega \end{aligned} \tag{3.73}$$

hence, maintaining the hypothesis of normality on the disturbances, we have:

$$y \approx N\left[(I-\lambda W)^{-1} X\beta; \sigma_\varepsilon^2 \Omega\right] \tag{3.74}$$

Now, remembering the simplification of the determinant $\left|\sigma_\varepsilon^2\Omega\right|$ reported in Equation (3.50) the likelihood is easily derived as:

$$\begin{aligned} L(\sigma^2,\rho,\lambda,\beta;y) = coust\left(\sigma_\varepsilon^2\right)^{-n/2}\left|I-\lambda W\right|\left|I-\rho W\right| \\ \times\exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left[y-(I-\rho W)^{-1}Z\beta\right]^T\right. \\ \left.\times\Omega^{-1}\left[y-(I-\rho W)^{-1}Z\beta\right]\right\} \end{aligned}$$

and the log-likelihood as:

$$\begin{aligned} l(\sigma^2,\lambda,\rho,\beta;y) = coust-\frac{n}{2}\ln\sigma_\varepsilon^2+\ln\left|I-\lambda W\right|+\ln\left|I-\rho W\right| \\ -\frac{1}{2\sigma_\varepsilon^2}\left[y-(I-\lambda W)^{-1}Z\beta\right]^T(I-\lambda W)^T \\ \times(I-\rho W)^T(I-\rho W)(I-\lambda W) \\ \times\left[y-(I-\lambda W)^{-1}Z\beta\right] \end{aligned} \tag{3.75}$$

and, since $(I-\lambda W)\left[y-(I-\lambda W)^{-1}Z\beta\right] = (I-\lambda W)y-Z\beta$, we eventually obtain:

$$\begin{aligned} l(\sigma^2,\lambda,\rho,\beta;y) = coust-\frac{n}{2}\ln\sigma_\varepsilon^2+\ln\left|I-\lambda W\right|+\ln\left|I-\rho W\right| \\ -\frac{1}{2\sigma_\varepsilon^2}\left[(I-\rho W)(y-Z\beta-\lambda Wy)\right]^T \\ \times\left[(I-\rho W)(y-Z\beta-\lambda Wy)\right] \end{aligned} \tag{3.76}$$

This, as usual, can only be maximized numerically to derive the estimators of the unknown parameters. The previous expression can be written in a different way by considering the following transformations known in the literature as the spatial *Cochrane–Orcutt transform*:

$$y^* = (I-\rho W)y$$

and

$$Z^* = (I-\rho W)Z$$

Expressed in this way the log-likelihood becomes:

$$l(\sigma^2,\lambda,\rho,\beta;y) = coust - \frac{n}{2}\ln\sigma_\varepsilon^2 + \ln|I-\lambda W| + \ln|I-\rho W| - \frac{1}{2\sigma_\varepsilon^2}\left[y^* - Z^*\beta - \lambda W y^*\right]^T \times \left[y^* - Z^*\beta - \lambda W y^*\right] \tag{3.77}$$

As mentioned already, currently there is no formal proof that the above *ML* estimator possesses the usual optimal large sample properties of an *ML* estimator. For this reason (and also to overcome the computational problems arising from the calculation of the log-determinant in large samples) the literature has suggested a spatial version of the Two-Stage Least Squares which will be discussed in the next section.

3.6.3 The Generalized Spatial Two-Stage Least Squares (GS2SLS)

The Generalized Spatial Two-Stage Least Squares (GS2SLS) was introduced by Kelejian and Prucha (1998) and accounts for both the problem of endogeneity of Wy and the problem of spatial correlation among the stochastic disturbances. It is an extension of the 2SLS methodology already illustrated in section 3.5.3 for Spatial Lag models, but it is combined with the GMM estimator presented in section 3.4.3 to account for the spatial correlation structure in the disturbances.

The GS2SLS procedure can be obtained using the following steps:

Step 1: first of all obtain a consistent estimate of the parameters β and λ, say $\tilde{\beta}$ and $\tilde{\lambda}$

Step 2: use these estimates to obtain an estimate of u in Equation (3.70), say $\hat{u}$

Step 3: use $\hat{u}$ to estimate ρ in Equation (3.71), say $\hat{\rho}$

Step 4: use $\hat{\rho}$ to transform model (3.70) as

$$(I-\hat{\rho}W)y = (I-\hat{\rho}W)Z\beta + \varepsilon$$

Step 5: finally, estimate the parameters of such a transformed model using 2SLS with the transformed variables $Z^* = (I-\hat{\rho}W)Z$; $WZ^* = W(I-\hat{\rho}W)Z$ and $W^2Z^* = W^2(I-\hat{\rho}W)Z$ as instruments.

These steps will be now discussed in detail.

Step 1: Estimate the parameters of (3.70) consistently accounting for the problem of endogeneity via the 2SLS estimator using Z and WZ as instruments. The motivation for this choice follows the same argument already used in section 3.4.3. Let us refer to the estimates thus obtained with the symbols $\tilde{\beta}$ and $\tilde{\lambda}$.

Step 2: From Equation (3.68) derive

$$\hat{u} = y - Z\tilde{\beta} - \tilde{\lambda}WY \tag{3.78}$$

and consequently we define:

$$\hat{\bar{u}} = W\hat{u} \tag{3.79}$$

$$\hat{\bar{\bar{u}}} = W^2\hat{u} \tag{3.80}$$

Step 3: Use the terms in Equations (3.78) to (3.80) to obtain a consistent estimator of ρ via the generalized Method of Moments procedure introduced in section 3.5.3 and in particular using the moments conditions contained in Equation (3.38). Let us call $\hat{\rho}$ such a consistent estimator of ρ.

Step 4: Use the estimate $\hat{\rho}$ obtained in Step 3 to transform the original model as follows:

$$(I - \hat{\rho}W)y = (I - \hat{\rho}W)(Z\beta - \lambda Wy) + \varepsilon \tag{3.81}$$

Step 5: Finally, estimate β and λ in Equation (3.81) using a 2SLS procedure with $H = [X, WX, W^2X]$ as instruments. In this way we obtain:

$$\tilde{\delta}_{GS2SLS} = \left[\hat{Q}^{*T}Q\right]\hat{Q}^{*T}y^{*} \tag{3.82}$$

where $\delta \equiv [\beta, \lambda]$, $Q = [Z, Wy]$, $Q^{*} = (I - \tilde{\rho}W)Q$ and $\hat{Q}^{*} = H(H^TH)^{-1}H^TQ^{*}$. Kelejian and Prucha (1998) showed that, under the model's assumptions, the GS2SLS estimators are consistent with an asymptotic variance equal to:

$$\sigma_{\varepsilon}^2\left[\hat{Q}^{*T}\hat{Q}^{*}\right]^{-1} \tag{3.83}$$

3.6.4 The fully efficient Lee estimators

Even if the GS2SLS estimators are consistent, it has been proved that they are not asymptotically fully efficient. To eliminate this problem, an asymptotically efficient alternative estimator was suggested by Lee (2003) known in the literature as Best Feasible *GS2SLS* or *BFGS2SLS* for short.

In the suggested procedure the optimal instrument matrix is defined as follows:

$$\bar{Q}^{\star} = (I - \tilde{\rho}W)\left[Z, W(I - \tilde{\lambda}W)^{-1}Z\tilde{\beta}\right] \tag{3.84}$$

with the symbols already used in section 3.6.3. The *BFGS2SLS* estimator is defined as:

$$\hat{\delta}_{BFGS2SLS} = \left[\bar{Q}^{\star T}Q^{\star}\right]^{-1}\bar{Q}^{\star T}y^{\star} \tag{3.85}$$

Such an estimator achieves the theoretical lower bound for the variance in large samples. The computation of the instrument in Equation (3.84), however, involves an operation that can be numerically challenging in very large samples. For this reason, even if Lee (2003) himself derives a numerical algorithm, he also suggests an alternative estimator, much simpler in terms of computation. Kelejian et al. (2004) show with a simulation study that in small samples both the BFGS2SLS and its simplified version do not differ substantially in terms of efficiency from the GS2SLS.

Example 3.4 House price determinants in Boston (continued)

In Example 3.3 we estimated a Spatial Lag model seeking to explain the spatial variability of house prices among 506 census tracts in Boston. The estimation phase was not entirely satisfactory in that, both using the ML and the 2SLS techniques, there was still some positive significant spatial correlation among the regression residuals. Let us re-estimate again the same model by specifying it as a SARAR(1,1).

First of all, as usual, let us start by estimating the model with the maximum likelihood technique (section 3.6.2). The results are shown in this table:

Parameter	*Estimated Value*	*Standard Error*	*t-test*	*p-value*
λ	0.072404	0.093996	0.77029	0.44113
ρ	0.52612	0.1141	4.611	4.0068e-06***
Intercept	38.2472606	6.0539170	6.3178	2.654e-10***
CRIM	–0.1164539	0.0325364	–3.5792	0.0003447***
RM	3.8363958	0.4074820	9.4149	< 2.2e-16***
INDUS	–0.0069422	0.0617867	–0.1124	0.9105396
NOX	–19.4589672	4.1417000	–4.6983	2.623e-06***
AGE	–0.0177129	0.0140407	–1.2615	0.2071146
DIS	–1.4637506	0.2609413	–5.6095	2.029e-08***
RAD	0.3216872	0.0726696	4.4267	9.568e-06***
PTRATIO	–1.0251807	0.1379809	–7.4299	1.088e-13***
B	0.0098786	0.0026440	3.7362	0.0001868***
LSTAT	–0.5162812	0.0496707	–10.3941	< 2.2e-16***
TAX	–0.0112292	0.0038528	–2.9145	0.0035622***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

ρ = 0.52612 LR-test= 26.375 (p-value= 1.8734e–06***) JB test = 1085.642 (p-value =2.2e–16***)

AIC = 3022.9 BIC =3086.249

The results are comparable to those of the models estimated in Example 3.3 in terms of the significance and of the sign of the variables. The parameter ρ related to the residual spatial autocorrelation is positive and significantly different from zero, while the parameter λ is also positive, but not significantly different from zero. Let us further consider the results related to the same model, but estimated using the Generalized Spatial Two-Stage Least Squares technique presented in section 3.6.3. These results are shown in the following table:

Parameter	*Estimated Value*	*Standard Error*	*t-test*	*p-value*
λ	0.1685872	0.0824706	2.0442	0.0409328**
Intercept	33.8711396	5.9812413	5.6629	1.488e–08***
CRIM	–0.1096767	0.0329242	–3.3312	0.0008648***
RM	3.8187552	0.4155824	9.1889	< 2.2e–16***
INDUS	–0.0064382	0.0624589	–0.1031	0.9179000
NOX	–17.1885687	4.1399033	–4.1519	3.297e–05***
AGE	–0.0112438	0.0138795	–0.8101	0.4178827
DIS	–1.3975787	0.2282525	–6.1229	9.186e–10***
RAD	0.3156587	0.0713019	4.4271	9.552e–06***
PTRATIO	–1.0002552	0.1384726	–7.2235	5.067e–13***
B	0.0098127	0.0026849	3.6548	0.0002574***
LSTAT	–0.5195838	0.0505312	–10.2824	< 2.2e–16***
TAX	–0.0110596	0.0038402	–2.8800	0.0039772***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

ρ = 0.35868

*JB test = 1129.996 (p-value =2.2e–16***)*

The estimates based on the Two-Stage Least Squares substantially confirm the conclusions of the ML estimation in terms of both the sign and the significance of the variables. In particular, it confirms that the proportion of non-retail business acres per town (INDUS) and the presence of old buildings in the area (AGE) have no significant effects on the median house price. Finally, the significance of the parameter ρ is confirmed. The presence of significant non-normality of the residuals indicates that we should use the GSTSLS method which does not require such a hypothesis. Using this alternative estimator, the parameter λ, which was not significant using ML estimator, is also significantly different from zero.

3.7 Testing spatial autocorrelation among the residuals with an explicit alternative hypothesis

In section 2.3.4 we discussed a testing procedure for the hypothesis of no spatial correlation among the *OLS* regression residuals based on Moran's I statistics (Moran, 1950). A pitfall of this test statistic is that no alternative hypothesis is explicitly considered to contrast the null of uncorrelation. In this chapter we have introduced some alternative formulations to the classical linear regression model based on various ways of taking into account the spatial dependence that is likely to be observed when dealing with spatial samples. These models can be considered as explicit alternative hypotheses to the case of uncorrelation in a testing procedure. Therefore the problem can now be approached in a more comprehensive way.

When we can explicitly express the alternative hypothesis either in the form of a Spatial Lag or of a Spatial Error, and a Maximum Likelihood strategy has been followed in the estimation phase, a Lagrange Multiplier test strategy can be followed (see section 1.1). Furthermore, for all the spatial models treated in this chapter, a *modified* version of Moran's I statistics can be considered. These two alternatives will be now reviewed in sections 3.7.1 and 3.7.2 respectively.

3.7.1 Testing spatial autocorrelation among the residuals using SEM or SLM as alternatives

To begin, let us consider the general form of a Lagrange Multiplier test (see Equation (1.35)):

$$LM = s(\theta_0)^T I(\theta_0)^{-1} s(\theta_0) \tag{3.86}$$

where θ is a vector of parameters, $s(\theta_0) = \frac{\partial L(\theta)}{\partial \theta}$ is the score function and $I(\theta_0) = E\left[\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta^T}\right]$ is the Fisher's Information matrix associated with the likelihood function $L(\theta)$ under the null of no spatial correlation.

When the alternative hypothesis is specified as a Spatial Error Model the log-likelihood function assumes the expression derived in Equation (3.23). As a consequence, in this case, Equation (3.86) assumes the explicit expression:

$$LM_{SEM} = \frac{n^2}{tr\left(W^T W + WW\right)}\left[\frac{\hat{\varepsilon}^T W \hat{\varepsilon}}{\hat{\varepsilon}^T \hat{\varepsilon}}\right]^2 \tag{3.87}$$

which is simply the square of the Moran I test as demonstrated by Burridge (1980). So, using Moran's *I* or the LM test will lead to the same inferential conclusions.

Conversely, in case the alternative hypothesis is specified in the form of the Spatial Lag Model, the log-likelihood is specified in Equation (3.52) and so Equation (3.86) becomes:

$$LM_{LAG} = \frac{n^2}{Q}\left[\frac{\hat{\varepsilon}^T W y}{\hat{\varepsilon}^T \hat{\varepsilon}}\right]^2 \tag{3.88}$$

with $Q = \left(WX\hat{\beta}\right)^T (I - M_x)\frac{WX\hat{\beta}}{\hat{\sigma}_\varepsilon^2} + T$, $M_x = X\left(X^T X\right)X^T$, $T = tr(W^T W + WW)$ and with $\hat{\beta}$ and $\hat{\sigma}_\varepsilon^2$ denoting the maximum likelihood estimators of the corresponding parameters of Equation (3.45). Sometimes a further alternative hypothesis is considered with an error structure that follows a Spatial Autoregressive and Moving Average structure. However, a detailed review of this alternative is not presented here. Both LM_{SEM} and LM_{LAG} are asymptotically distributed, under the null, as a χ^2 with 1 degree of freedom. However the two test statistics are not independent on one another so that one can only test the alternative hypothesis that the errors follow a SEM model assuming that there no spatial lag component and vice versa. For this reason

Anselin et al. (1996) proposed a robust version of both tests which can be expressed as:

$$RLM_{SEM} = \frac{1}{T(1-TQ)}\left[\frac{n\hat{\varepsilon}^T W\hat{\varepsilon}}{\hat{\varepsilon}^T\hat{\varepsilon}} - TQ^{-1}\frac{n\hat{\varepsilon}^T W y}{\hat{\varepsilon}^T\hat{\varepsilon}}\right]^2 \tag{3.89}$$

for the alternative hypothesis of a Spatial Error Model and respectively:

$$RLM_{LAG} = \frac{1}{Q-T}\left[\frac{n\hat{\varepsilon}^T W\hat{\varepsilon}}{\hat{\varepsilon}^T\hat{\varepsilon}} - \frac{n\hat{\varepsilon}^T W y}{\hat{\varepsilon}^T\hat{\varepsilon}}\right]^2 \tag{3.90}$$

using the Spatial Lag model as an alternative.

3.7.2 Testing spatial autocorrelation among the residuals using a spatial model as an alternative: the modified Moran I test

In section 2.2 we considered the statistic introduced by Moran (1950) and studied by Cliff and Ord (1972) to test the hypothesis of no spatial correlation among regression residuals. The statistic is reported here again for the convenience of the reader:

$$I = \frac{n\hat{\varepsilon}^T W\hat{\varepsilon}}{\hat{\varepsilon}^T\hat{\varepsilon}\left[\sum_i\sum_j w_{ij}\right]} \tag{3.91}$$

$\hat{\varepsilon}$ being the model's residuals. More recently, Kelejian and Prucha (2001) have criticized this measure, arguing that the normalizing factor used by Cliff and Ord (1972) to derive its expected value and the variance under the null of no spatial correlation is not theoretically justified. In fact, the denominator of (3.91) represents the estimator of the standard deviation of the quadratic form appearing in the numerator and this can be proved to be inconsistent. For this reason, they proposed a different normalizing factor that removes this inconsistency and achieves the aim of normalizing the variance to unity. The alternative Moran test, in general, assumes the following expression (see Kelejian and Prucha, 2001):

$$\bar{I} = \frac{\hat{\varepsilon}^T W\hat{\varepsilon}}{\tilde{\sigma}^2} \tag{3.92}$$

with $\tilde{\sigma}^2$ a normalizing factor that depends on the particular model chosen as an alternative hypothesis. In particular, if the alternative

hypothesis is constituted by a Spatial Error model, the normalizing factor assumes the expression:

$$\tilde{\sigma}^2 = \frac{\hat{\varepsilon}^T\hat{\varepsilon}\left\{tr\left[\left(W^T + W\right)W\right]\right\}^{-1/2}}{n} \tag{3.93}$$

with the term $tr(A)$ indicating the trace of matrix A, which is the sum of its main diagonal elements. As a consequence the test statistics can be defined as:

$$\bar{I} = \frac{n\hat{\varepsilon}^T W\hat{\varepsilon}}{\hat{\varepsilon}^T\hat{\varepsilon}\left\{tr\left[\left(W^T + W\right)W\right]\right\}^{-1/2}} \tag{3.94}$$

The two expressions reported in Equations (3.91) and (3.94) coincide if the weight matrix has dichotomous entries (0 and 1) in which case $w_{ij} = w_{ij}^2$ and, therefore, $\sum_i\sum_j w_{ij} = \left\{tr\left[\left(W^T + W\right)W\right]\right\}^{-1/2}$.

In contrast, in the case of a SARAR(1,1) model, the normalizing factor can be derived in the following way. Consider again the two equations of a SARAR model:

$$y = \lambda Wy + X\beta_{(1)} + WX\beta_{(2)} + u \;\; = \mathrm{Q}\delta + \mathrm{u} \qquad |\lambda| < 1 \tag{3.95}$$

$$u = \rho Wu + \varepsilon \qquad |\rho| < 1 \tag{3.96}$$

where we now set $Q = [Z, Wy]$; $Z = [X, WX]$ as usual, $\delta^T = \left[\beta^T, \rho^T\right]$ and $\beta^T = \left[\beta^T{}_{(1)}, \beta^T{}_{(2)}\right]$ and let us assume again the validity of the assumptions behind the *GMM* procedure described in section 3.4.3. Consider further the Generalized Spatial Two-Stage Least Square Estimators derived in Equation (3.82)

$$\tilde{\delta}_{GS2SLS} = \left[\hat{Q}^{*T}Q\right]\hat{Q}^{*T}y^* \tag{3.97}$$

with $\hat{Q}^* = H(H^TH)^{-1}H^TQ^*$ and H the matrix of instruments $H = \left[X, WX, W^2X\right]$, and let us further indicate with $\hat{\varepsilon} = y - Q\tilde{\delta}$ the

GS2SLS residuals. In order to test the null hypothesis of no spatial correlation among the residuals (that is ρ = 0) against the alternative hypothesis that ρ ≠ 0 Kelejian and Prucha (2001) derived the following normalizing factor:

$$\tilde{\sigma}^2 = n^{-2}\left(\hat{\varepsilon}^T\hat{\varepsilon}\right)^2\left\{tr\left[\left(W^T + W\right)W\right] + \left(n^{-1}\hat{\varepsilon}^T\hat{\varepsilon}\right)\hat{c}^T\hat{c}\right\}^{1/2} \tag{3.98}$$

which, substituted into Equation (3.92), leads to the following modification of the Moran I test statistics:

$$\bar{I} = \frac{\left(\hat{\varepsilon}^T W\hat{\varepsilon}\right)}{n^{-2}\left(\hat{\varepsilon}^T\hat{\varepsilon}\right)^2\left\{tr\left[\left(W^T + W\right)W\right] + \left(n^{-1}\hat{\varepsilon}^T\hat{\varepsilon}\right)\hat{c}^T\hat{c}\right\}^{1/2}} \tag{3.99}$$

where, in addition to the previously introduced symbolism, we define $\hat{c} = -H\hat{P}\hat{\alpha}$, $\hat{P} = \left(\frac{H^TH}{n}\right)^{-1}\frac{H^T\hat{Q}^*}{n}\left(\frac{\hat{Q}^{*T}\hat{Q}^*}{n}\right)^{-1}$ and $\hat{\alpha} = \frac{\hat{Q}^{*T}(W + W^T)\hat{\varepsilon}}{n}$. In their contribution, Kelejian and Prucha (2001) prove that the modified Moran test $\bar{I}$ converges in distribution to a standardized normal distribution even when the a priori assumption of the normality of the errors is not satisfied. Even if in large samples $\bar{I}$ is $N(0,1)$, in small samples its expected value and variance may be different. Their formal expressions are derived in the quoted paper by Kelejian and Prucha (2001).

Example 3.5 Phillips curve (continued)

Let us consider again the Phillips curve estimated for the 20 Italian regions already discussed in Example 2.4. The OLS estimation leads to the model $\Delta unempl = -9.827 + 8.746\Delta prices$ *and the Moran test (0.3212607) on the regression residuals (using a contiguity-based W matrix) was judged not significantly different from zero at 94%. As we know, Moran's I test does not have an explicit alternative hypothesis. Let us now test the hypothesis of no spatial correlation in the residuals using the Spatial Lag and the Spatial Error as explicit alternatives to the hypothesis of no residual spatial*

dependence. The results obtained are shown below together with their robust versions.

Test Statistic	*Test Value*	*p-value*
LM_{ERR}	*2.8633*	*0.0906235*
LM_{LAG}	*12.7722*	*0.0003518****
RLM_{ERR}	*1.7170*	*0.1900785*
RLM_{LAG}	*11.6260*	*9.0006504****

The table clearly reveals that, if compared with the Spatial Error model (as expected by the equivalence demonstrated by Burridge, 1980), the LM test confirms the finding of Moran's I test in detecting non-significant residual spatial correlation. However, if compared with a Spatial Lag model the hypothesis of error spatial independence cannot be accepted. The robust versions of the two tests substantially confirms these findings.

3.8 Interpretation of the parameters in spatial econometric models

In a standard linear regression model the regression parameters have an easy interpretation in that they represent the partial derivative of the dependent variable y with respect to the independent variables:

$$\beta_i = \frac{\partial}{\partial X_i} y_i \tag{3.100}$$

they can therefore be straightforwardly interpreted as the variation induced on variable y of a unitary increase in the single independent variable X_i.

However, in the spatial econometric models presented in this chapter the interpretation of the parameters is less immediate and requires some clarification. In fact, a variation of variable X observed in location i does not only have an effect on the value of variable y in the same location, but also on variable y observed in other locations. Consider, for instance, Okun's Law model presented in Example 2.3. The model predicts that an increase in the level of GDP produces a decrease in the level of unemployment. Consider also, for simplicity, a Spatial Lag model. The model can be expressed as follows:

$$\Delta unempl_i = \beta_0 + \beta_1 \Delta GDP_i + \lambda \sum_{j=1}^{n} w_{ij} \Delta unempl_j \tag{3.101}$$

In this case, an increase in the *GDP* in region *i* produces as an immediate effect, a decrease in the level of unemployment in that region. However, given the spatial autoregressive mechanism considered in this framework, the variation of the level of unemployment in region *i*, also produces an effect on the level of unemployment in other neighboring regions so that all impacts have to be evaluated simultaneously. This topic has been treated in the literature by Kelejian et al. (2006) and LeSage and Pace (2009) amongst others. The formal solution to the problem consists of evaluating the partial derivative (3.100) in each specific model. Consider, for example, the case of the Spatial Lag model:

$$y = \lambda Wy + X\beta + u \qquad |\lambda| < 1 \tag{3.102}$$

which can also be written in the reduced form as:

$$y = (I - \lambda W)^{-1} X\beta + (I - \lambda W)^{-1} u \tag{3.103}$$

so that:

$$E(y) = (I - \lambda W)^{-1} X\beta \tag{3.104}$$

The impact of each variable X on y can then be described through the partial derivatives $\frac{\partial E(y)}{\partial X}$ which can be arranged in the following matrix:

$$\frac{\partial E(y)}{\partial X} = S = \begin{bmatrix} \frac{\partial E(y_1)}{\partial X_1} & \cdots & \frac{\partial E(y_i)}{\partial X_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial E(y_n)}{\partial X_1} & \cdots & \frac{\partial E(y_n)}{\partial X_n} \end{bmatrix} \tag{3.105}$$

whose single entry is defined as:

$$s_{ij} = \frac{\partial E(y_i)}{\partial X_j} \tag{3.106}$$

On this basis, LeSage and Pace (2009) suggested the following impact measures that can be calculated for each independent variable X_i included in the model:

1. A global measure, termed *Average Direct Impact*. This measure refers to the average total impact of a change in X_i on y_i for each observation, which is simply the average of all diagonal entries in matrix S:

$$ADI = n^{-1}tr(S) = n^{-1}\sum_{i=1}^{n}\frac{\partial E(y_i)}{\partial X_i} \tag{3.107}$$

2. A measure related to the impact produced on one single observation by all other observations, termed *Average Total Impact To* an observation. For each observation this is calculated as the sum of the i-th row of matrix S:

$$ATIT_j = n^{-1}\sum_{i=1}^{n}s_{ij} = n^{-1}\sum_{i=1}^{n}\frac{\partial E(y_i)}{\partial X_j} \tag{3.108}$$

3. A measure related to the impact produced by one single observation on all other observations, termed *Average Total Impact From* an observation. For each observation this is calculated as the sum of the j-th column of matrix S:

$$ATIF_i = n^{-1}\sum_{j=1}^{n}s_{ij} = n^{-1}\sum_{j=1}^{n}\frac{\partial E(y_i)}{\partial X_j} \tag{3.109}$$

4. A global measure of the average impact obtained from the two preceding measures:

$$ATI = n^{-1}i^{T}Si = n^{-1}\sum_{j=1}^{n}ATIT_i = n^{-1}\sum_{j=1}^{n}ATIF_j \tag{3.110}$$

   which is simply the average of all entries of matrix S.
5. A measure of the average indirect impact obtained as the difference between ATI and ADI:

$$AII = ATI - ADI \tag{3.111}$$

   which is simply the average of all off-diagonal entries of matrix S.

Measures of relative impact for a general *SARAR* model can be similarly obtained by substituting the appropriate reduced form expressions in Equation (3.103).

Example 3.6 Okun's law in Italy (continued)

As an illustration of the impact measures described in the last section, let us consider again Okun's Law in the 20 Italian regions described in Example 2.3 in Chapter 2. In the example we estimated the model with a simple OLS estimator neglecting spatial effects, but we observed a significant residual correlation. Now, on the basis of the same W matrix used to calculate Moran's I in Example 2.3 (that is, a row-standardized W matrix based on mere adjacency between regions corrected to include the islands), let us re-estimate the model as a Spatial Lag and let us compute the impact measures described in section 3.8. The results are shown in the following table together with those based on OLS for the sake of comparison.

	OLS	*Spatial Lag Model*
Intercept	*10.971****	*3.12275****
GDP	*–3.326****	*–1.13532****
λ	–	*0.7476****
ADI	–	*–1.542448*
AII	–	*–2.95571*
ATI	–	*–4.498159*

In the OLS framework the interpretation is straightforward: if we observe an increase of 1 unit in the GDP in one region we expect a decrease of –3.326 units in the unemployment of the same region. Within the Spatial Lag framework it would be wrong to interpret the regression slope in the same way. The "direct" effect of increasing the GDP of 1 unit is NOT a decrease of 1.13532 units of unemployment, but rather a decrease of 1.542448 (see ADI in the Table). Furthermore, given the presence of "indirect" effects produced by the spillover of the increase of GDP from one region to all other regions (which is connatural with the Spatial Lag mechanism), an increase of 1 unit of GDP in one region produces a "total" impact that, on the average, is a reduction of 4.498159 units of unemployment (ATI). Thus the impact is greater if we consider a Spatial Lag model in place of the traditional a-spatial model due to the mechanism of geographical transmission which is incorporated in the former.

3.9 `R Codes`: estimating linear spatial models

All the `R` commands related to the procedures described in this chapter are contained in the library `spdep`. All the commands are very similar to the standard commands for regressions described in section 1.4 and are very straightforward to apply. In all cases we assume that we

want to estimate the basic model $y = \beta_0 + \beta_1 x + \beta_2 z + \varepsilon$ (with the relevant addition of a spatial lag or a spatial error, or both) using a weight matrix $\boldsymbol{W}$ which is generated by one of the procedures described in section 2. 3. We also assume that the observations of the variables y, x and z are stored in a file called `filename`. If the data are stored in the active `R` session, all the options including this specification can be omitted. Once a model (say `model`) is estimated, a summary of the estimation output can be obtained, as usual, by typing:

```
> summary(model)
```

To start, if we want to estimate the parameters of a **purely autoregressive** model, the command

```
> model0<-spautolm(x~1 , data=filename, listw=W)
```

provides the ML solution to the estimation problem.

If we want estimate a **Spatial Error** Model via the Maximum Likelihood (**ML**) criterion (section 3.4.2) use the command

```
> model1<-errorsarlm(formula= y~x + z, data=filename,
listw=W)
```

Conversely, to estimate the Spatial Error model by the Feasible Generalized Least Squares **(GLS)** procedure (section 3.4.3) we use, instead, the command:

```
> model2<-GMerrorsar (formula= y~x + z, data=filename,
listw=W)
```

When dealing with a **Spatial Lag** Model, if we wish to employ the **ML** method (section 3.5.2), use the command:

```
> model3<-lagsarlm(formula= y~x+z, data=filename, listw=W)
```

if we add the option

```
> model3<-lagsarlm(formula= y~x+z, data=filename, listw=W,
type="mixed")
```

we consider the case of a Spatial Lag Model with the additional spatial lag of all independent variables.

In contrast, if we want to estimate the Spatial Lag model with the Two-Stage Least Squares (**2SLS**) estimator (section 3.5.3), use the command:

```
> model4<-stsls (formula= y~x+z, data=filename, listw=W)
```

Finally, to estimate the parameters of a **SARAR** model by **ML** (section 3.6.2) use the command:

```
> model5<-sacsarlm(formula= y~x+z, data=filename, listw=W)
```

and to estimate the same model using the Generalized Spatial Two-Stage Least Squares (**GS2SLS)** method (section 3.6.3), type the command:

```
> model6 <- gstsls(Y~x+z, listw = w)
```

For the calculation of the autocorrelation test for residuals of a regression specifying the alternative hypothesis, use the command:

```
>lm.LMtests(Model1, listw=W, test="all")
```

through which we calculate the *LM* test using the Spatial Error or the Spatial Lag model as alternatives (LM_{SEM} and LM_{LAG} in section 3.7), their robust versions (RLM_{SEM} and RLM_{LAG}) or the modified Moran I test (I_n).

Notice that if we have estimated a Spatial Lag model and we want to test if there is still a residual spatial autocorrelation, an LM test procedure appears by default in the output of the `lagsarlm` procedure with no additional commands.

Finally, for the computation of the **impact measures** described in section 3.8, the `spdep` library contains a command with reference to the Spatial Lag model. If such a model has been estimated (say `model3`) and we have available a matrix `W` used in the estimation process, we can obtain the impact measures by typing the command:

```
> impact <- impacts(model3, listw = W)
```

Key Terms and Concepts Introduced

- Spatial autocorrelation
- Pure spatial autoregression
- Spatial Error model
- Spatial Lag model
- Spatial AutoRegressive with AutoRegressive error model (SARAR)
- Lagged independent variables model
- Maximum Likelihood solution for Spatial Error model estimation
- Ord's decomposition
- Feasible Generalized Least Squares solution for Spatial Error model estimation
- Generalized Method of Moments estimation
- Maximum Likelihood solution for Spatial Lag model estimation
- Two-Stage Least Squares solution for Spatial Lag model estimation
- Maximum Likelihood solution for SARAR model estimation
- Generalized Spatial Two-Stage Least Squares solution for SARAR model estimation
- Lee Instrumental Variable estimators
- Cochrane–Orcutt transformation
- Best Feasible Generalized Spatial Two-Stage Least Squares estimators
- Rao's score test (Lagrange Multiplier) for regression residuals with spatially lagged variables
- Robust Lagrange Multiplier test
- Modified Moran I test
- Impact measures

Questions

1. What is a major limitation of using the Moran's I statistics in testing the hypothesis of no spatial autocorrelation among the regression residuals? How can this limitation be addressed? What alternative tests are available?
2. What are the relative advantages and weaknesses of employing a Maximum Likelihood strategy and a Generalized Least Squares strategy in estimating a SARAR model?
3. In what sense is the procedure in section 3.4.3 described as "feasible"?
4. Why is it necessary within the Feasible GLS to consider a Method of Moments procedure?
5. How can we define a Cochrane–Orcutt transformation in space? How can this transformation help in the process of model estimation?

6. What is the aim of Ord's decomposition?
7. Why can we use Moran's I instead of a Lagrange Multiplier test for spatial dependence obtaining the same inferential conclusions?
8. Why do we need to consider the robust versions of the Lagrange Multiplier tests of spatial dependence?

Exercises

Exercise 3.1 The following map shows the boundaries of the 27 countries that were the members of the European Union in July 2012.

Map of the 27 EU member States in July 2012. (Courtesy of Carrie Dolan).

Using the procedure illustrated in Section 2.3.2, prepare a `.GAL` file and build up the row-standardized W matrix with the criterion of adjacency (join the island to the closest country, that is, Ireland to UK, UK to France, Malta to Italy and Cyprus to Greece. Also join Finland to Estonia and Denmark to Sweden). The table below shows the data related to the per capita growth of the Gross Domestic Product in the period 2010–11 and the percentage of GDP devoted to education in 2009.

COUNTRY CODE	COUNTRY	% Education Expenses 2009	Growth 2010–2011	COUNTRY CODE	COUNTRY	% Education Expenses 2009	Growth 2010–2011
BE	Belgium	42	1.046931408	AT	Austria	23.5	1.057823
BG	Bulgaria	27.9	1.058252427	PL	Poland	32.8	0.992958
CZ	Czech Republic	17.5	1.041237113	PT	Portugal	21.1	1.037234
DK	Denmark	40.7	1.072413793	RO	Romania	16.8	1.054054
DE	Germany	29.4	1.074074074	SI	Slovenia	31.6	1.118227
EE	Estonia	35.9	1.170068027	SK	Slovakia	17.6	1.05848
IE	Ireland	48.9	1.09	FI	Finland	45.9	1.107807
ES	Spain	39.4	1.070247934	SE	Sweden	43.9	1.099291
FR	France	43.2	1.04296875	UK	United Kingdom	41.5	1.084615
IT	Italy	19	1.069672131	EL	Greece	26.5	1.045249
CY	Cyprus	44.7	1.059574468	LU	Luxembourg	46.6	1.098333
LT	Lithuania	40.6	1.191176471	LV	Latvia	30.1	1.149606
HU	Hungary	23.9	1.045751634	MT	Malta	21	1.020202
NL	Netherlands	40.5	1.083870968				

Test the hypothesis that growth is fostered by education. Since we can conjecture that the expenses for education present some spatial pattern, in principle we can also include the lagged values of the variable education among the regressors. Using a Maximum Likelihood strategy, start estimating a SARAR model and then, on the basis of the results obtained, choose the best model between the Spatial Lag, the Spatial Error and the complete SARAR.

Exercise 3.2 Show that a Spatial Error model $y_i = \beta X_i + u_i;$ $u_i = \rho \dfrac{\sum_{j=1}^{n} w_{ij} u_j}{\sum_{j=1}^{n} w_{ij}} + \varepsilon_i;\ |\rho| < 1;\ \ \varepsilon_i$ n.*i.d* $N(0,\sigma_\varepsilon^2)$ can be expressed as a Spatial Lag model with an additional spatially lagged independent variable. Explain why, even if the error term ε is "well-behaved", the OLS criterion does not provide reliable estimators for the unknown parameters.

Exercise 3.3 Show that in the Spatial Error model, if the spatial correlation parameter ρ and the error variance σ_ε^2 are both known, the *ML* estimators of β coincide with the *GLS* estimators. Derive the explicit expression of the GLS solution and of their error variances.

Exercise 3.4 Show that a Spatial Lag model

$$y_i = \lambda \frac{\sum_{j=1}^{n} w_{ij} y_j}{\sum_{j=1}^{n} w_{ij}} + \beta X_i + u_i;\ |\lambda| < 1; \qquad u_i | X \approx i.i.d.N(0, \sigma_{u\,n}^2 I_n)$$

can be expressed as a pure autoregression of the independent variable y with a (non-zero mean) error term which incorporates the (non-stochastic) independent variable.

Exercise 3.5 Consider the SARAR model:

$$y_i = \lambda \frac{\sum_{j=1}^{n} w_{ij} y_j}{\sum_{j=1}^{n} w_{ij}} + u_i;\ |\lambda| < 1; \qquad u_i = \rho \frac{\sum_{j=1}^{n} w_{ij} u_j}{\sum_{j=1}^{n} w_{ij}} + \varepsilon_i;\ |\rho| < 1$$

$\varepsilon | X \approx i.i.d.N(0, \sigma_{\varepsilon\,n}^2 I_n)$ and consider the case of $\lambda = -\rho$. Show that in this case the model reduces to a particular case of a Spatial Lag model, discuss the parameter space and comment on the form of the W matrix.

Then derive the variance-covariance matrix of y and the associated likelihood function.

Exercise 3.6 Consider again the dataset discussed in Examples 3.3 and 3.4 related to the determinant of the median house price observed in 506 Boston area census tracts. To download the dataset type:

```
>data(boston)
```

Estimate the model used in Example 3.3 and 3.4, but specifying the spatial component in the Spatial Error form. Use both the Maximum Likelihood and the Feasible GLS procedure. Compare the results obtained with those presented in Examples 3.3 and 3.4.

References to the Chapter

Anselin, L. (1988) *Spatial Econometric: Methods and Models*, Dordrecht: Kluwer Academic Publishers.

Anselin, L., Bera, A., Florax, R. and Yoon, M. (1996) 'Simple Diagnostic Tests for Spatial Dependence', *Regional Science and Urban Economics,* 26, 77–104.

Arbia, G. (2006) *Spatial Econometrics: Statistical Foundations and Applications to Regional Economic Growth*, Heidelberg: Springer-Verlag.

Burridge, P. (1980) 'On the Cliff–Ord Test for Spatial Autocorrelation', *Journal of the Royal Statistical Society, Series B,* 42, 107–8.

Cliff, A.D. and Ord, J.K. (1972) *Spatial Autocorrelation*, London: Pion.

Gerschgorin, S. (1931) 'Über die Abgrenzung der Eigenwerte einer Matrixi', *Izv. Akad. Nauk. USSR Otd. Fiz.-Mat.*, 7, 749–54.

Gilley, O.W. and Pace, R.K. (1996) 'On the Harrison and Rubinfeld Data', *Journal of Environmental Economics and Management*, 31, 403–5.

Harrison, D. and Rubinfeld, D.L. (1978) 'Hedonic Housing Prices and the Demand for Clean Air', *Journal of Environmental Economics and Management*, 5, 81–102.

Kelejian, H.H. and Prucha, I. (1997) 'Estimation of the Spatial Autoregressive Parameter by Two Stage Least Squares Procedure: A Serious Problem', *International Regional Science Review,* 20, 103–11.

Kelejian, H.H. and Prucha, I. (1998) 'A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances', *Journal of Real Estate Finance and Economics*, 17, 99–121.

Kelejian, H.H. and Prucha, I. (2001) 'On the Asymptotic Distribution of the Moran I Test Statistics with Applications', *Journal of Econometrics,* 104, 219–57.

Kelejian, H.H., Prucha, I. and Yuzefovich, E. (2004) 'Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbance: Large and Small Sample Results', in J. LeSage and K. Pace (eds), *Advances in Econometrics: Spatial and Spatiotemporal Econometrics,* New York: Elsevier, pp. 163–98.

Kelejian, H.H., Tavlas, G.S. and Hondronyiannis, G. (2006) 'A Spatial Modeling Approach to Contagion Among Emerging Economies', *Open Economies Review,* 17, 4–5, 423–42.

Lee, L.F. (2003) 'Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances', *Econometric Reviews,* 22, 307–35.

Lee, L.F. (2004) 'Asymptotic Distribution of Maximum Likelihood Estimators for Spatial Autoregressive Models', *Econometrica,* 72, 1899–925.

LeSage, J. and Pace, R.K. (2009) *Introduction to Spatial Econometrics*, Boca Raton, FL: Chapman & Hall/CRC Press.

Moran, P.A.P. (1950) 'Notes on Continuous Stochastic Phenomena', *Biometrika*, 37, 1, 17–23.

Ord, J.K. (1975) 'Estimation Methods for Models of Spatial Interaction', *Journal of the American Statistical Association*, 70, 120–6.

Whittle, P. (1954) 'On Stationary Processes in the Plane', *Biometrika*, 41, 434–49.

4
Further Topics in Spatial Econometrics

This chapter discusses some advanced special topics in spatial econometrics that have recently been introduced in the literature. The primary purpose is to make the reader knowledgeable on a set of techniques that represent the evolution of the methods presented in Chapter 3 and that constitute an essential part of the skill set currently required by spatial econometricians. These methods have the potential to make a tremendous impact in the analysis of real world problems in many scientific fields. In particular, section 4.1 discusses the case of non-constant innovation variances (*heteroscedastic models*), section 4.2 refers to the case where the dependent variable assumes a *discrete* (in particular, a binary) form, section 4.3 contains some of the modeling strategies in the field of diachronic spatial econometric models estimated on *panel data* and, finally, section 4.4 discusses regression models that are non-stationary in the geographical space. Following the introductory nature of the current presentation, we will discuss the various methods with less analytical detail compared to the rest of the book, while referring the interested reader to the current literature for more detail.

4.1 Heteroscedastic innovations

4.1.1 Generalities

Up to this point in the text, all of the models presented have considered the innovations to be $\varepsilon | X \approx i.i.d.N(0, \sigma^2_{\varepsilon} {}_n I_n)$ thus assuming the hypothesis of constant variance (homoscedasticity). However, as already discussed in section 1.2, for many reasons this assumption can be considered rather unrealistic when dealing with spatial data. In fact, in contrast to other typologies of data (for example, time series data), spatial data are observed within units that, in general, are different in terms of

their size and shape. This characteristic, which is peculiar to all spatial data observed on irregular lattices, may result in heteroscedasticity of the innovations with larger variances in larger areas. Furthermore, even if we can assume homescedasticity at a given level of spatial aggregation (for instance, the county level), such a characteristic is generally lost if we aggregate the data at a higher level of aggregation (for example, the state level) due to the different dimension of the various geographical areas (for a proof see, for example, Arbia, 1989).

Part of the most recent literature has concentrated on this aspect by introducing heteroscedastic spatial econometric models of the class SARAR (Kelejian and Prucha, 2007, 2010). We will discuss some of these advances in the present section. In particular, we will distinguish two main estimation approaches. The first is an approach where the heteroscedasticity of the innovations is taken into account by parametrically estimating their variance-covariance matrix (see section 4.1.2). The second is a non-parametric approach in which an estimation procedure is developed that is robust to possible misspecifications of the innovation process and the structure of their variance-covariance matrix is inferred non-parametrically from the observed data (section 4.1.3).

4.1.2 The SARAR model with heteroscedastic disturbances

Let us consider again the SARAR(1,1) model introduced in section 3.6 and shown here below for convenience

$$y = Z\beta + \lambda Wy + u \qquad |\lambda| < 1 \tag{4.1}$$

$$u = \rho Wu + \varepsilon \qquad |\rho| < 1 \tag{4.2}$$

but now, in place of the usual assumption $\varepsilon | X \approx i.i.d.N(0, \sigma_{\varepsilon}^2{}_n I_n)$, let us assume instead $E(\varepsilon_i^2) = \sigma_i^2$. It should be evident that, by using the same argument considered in section 3.5, a problem of endogeneity emerges in that the term Wy is correlated with the innovations. This motivates the use of a two-stage strategy. Furthermore a problem of inefficiency emerges as a result of the heteroscedastic innovations so that a weighted procedure needs to be considered.

The procedure that we will illustrate was introduced by Kelejian and Prucha (2010) and uses an approach which is similar to the Spatial Two-Stage Least Squares introduced in section 3.6 for the general SARAR(1,1) model with homescedastic innovations, with some important differences. In what follows we will try to provide the intuitions behind the procedure, without discussing it in all its mathematical details, while referring the interested reader to the referenced paper to gain a deeper

understanding in the subject. In particular there are two intuitions. The first is that, when using a feasible GLS approach, in the third step of the procedure illustrated in Section 3.4.3, a GMM estimation was obtained imposing three moments conditions. The first of these conditions was on the variance, $\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{u}_i - \rho\hat{\tilde{u}}_i\right)^2 = \sigma_\varepsilon^2, \text{ see Equation (3.35)}\right)$. However, in the present context we no longer have a single parameter σ_ε^2 due to the relaxation of the homescedasticity hypothesis and due to the fact that heteroscedasticity is not expressed in a specified form. So, in this new case, we will have to consider the error variances σ_i^2 as nuisance parameters and, instead of imposing three moments conditions to estimate the whole parameter vector $\varphi^T = \left[\rho^2, \rho, \sigma_\varepsilon^2\right]$ (see Equation (3.41)), we will only impose two moments conditions and we will estimate the reduced parameter vector $\varphi^T = \left[\rho^2, \rho\right]$. So the estimation in this phase will concentrate on the only parameter of interest ρ. The second intuition is that, dropping the hypothesis of homescedasticity (by analogy with the discussion reported in section 1.2) the *GMM* estimation of the model's parameter ρ, will still provide consistent estimators, but with a variance that will no longer be minimal. As a consequence, in a similar way to the correction of the *OLS* employed in section 1.2, we will also need to correct the *GMM* estimation with a weighted procedure.

The phase of estimation is rather complex and moves sequentially through different steps, making use of moment conditions and instruments. A modified *GS2SLS* procedure can, in this case, be obtained following these sequential steps:

Step 1: first of all, obtain a consistent estimate of the parameters β and λ, say $\tilde{\beta}$ and $\tilde{\lambda}$

Step 2: use the estimates obtained in the first step in order to obtain an initial estimate of u in Equation (4.1), say $\hat{u}$

Step 3: use $\hat{u}$ to obtain an initial estimate the autoregressive parameter ρ in Equation (4.2), say $\tilde{\rho}$, imposing moments conditions. This estimation is consistent, but inefficient due to the lack of constancy of the variances,

Step 4: an efficient estimation of the parameter ρ is obtained by using a weighted GMM procedure; call $\hat{\rho}$ such an efficient estimate.

Step 5: use the efficient estimate $\hat{\rho}$ to transform model (4.1) as

$$(I - \hat{\rho}W)y = (I - \hat{\rho}W)Z\beta + \varepsilon$$

Step 6: estimate the parameters of such a transformed model using a 2SLS strategy using the transformed variables $Z^* = (I - \hat{\rho}W)Z$; $WZ^* = W(I - \hat{\rho}W)Z$ and $W^2Z^* = W^2(I - \hat{\rho}W)Z$ as instruments. Call this a Feasible Generalized Spatial Two Stage Least Squares estimator (FGS2SLS).

Step 7: Define the new GS2SLS residuals.

Step 8: Finally obtain an efficient GMM estimate of ρ based on the residuals of the FGS2SLS procedure discussed in Step 6.

These steps will now be discussed in detail.

Step 1: Estimate the parameters of (4.1), consistently accounting for the problem of endogeneity via the 2SLS estimator using Z and WZ as instruments. The motivation for this choice is the same used in the previous sections. Let us call the estimates thus obtained $\tilde{\beta}$ and $\tilde{\lambda}$.

Step 2: From Equation (4.1) derive

$$\tilde{u} = y - Z\tilde{\beta} - \tilde{\lambda}WY$$

Step 3: Use the terms in Equation (4.2) to obtain an initial estimator of ρ via an unweighted *GMM* procedure analogous to that considered in section 3.6.3, but based on just two moments conditions since no condition can be now derived from the parameter σ^2. In other words, since heteroscedasticity is not expressed in a prespecified form, the *GMM* procedures can only focus on the estimation of the unknown parameter ρ. Let us call $\breve{\rho}$ the initial estimator of ρ thus obtained. Kelejian and Prucha (2010) suggest that this estimator can be interpreted as an unweighted nonlinear least squares estimator (Greene, 2011). This estimator, in our hypotheses, is consistent, but inefficient due to the heterogeneity of the disturbances' variances.

Step 4: Starting from the result obtained in Step 3, an efficient *GMM* estimator of ρ can be defined as a weighted version of the nonlinear least square estimator discussed above with weights provided by the elements of a matrix (which we will refer to with the symbol Ψ) which represents an estimator of the variance-covariance matrix of the limiting distribution of the normalized sample moments. Given the crucial role played by the matrix Ψ in the estimation procedure, let us first of all define it formally as:

$$_2\Psi_2 = n^{-1}{}_2H_n^T{}_n\Sigma_n{}_nH_2 \tag{4.3}$$

with n the sample size, ${}_nH_2 = [Z, WZ]$ an n-by-2 matrix of instruments and $\Sigma = E[uu^T]$ the variance-covariance matrix of the disturbances of generic elements $\sigma_{ij} \in \Sigma$. Let us further refer to the generic element of the matrix Ψ as ψ_{rs}, $r, s = 1, 2$. Notice that the matrix of instruments H now has dimension n-by-2 because we only impose two moments conditions.

In their work Kelejian and Prucha (2010) propose the following parametric estimator for the generic element of ψ_{rs}

$$\tilde{\psi}_{rs} = (2n)^{-1} tr\left[\left(A_r + A_r^T\right)\tilde{\Sigma}\left(A_s + A_s^T\right)\tilde{\Sigma}\right] + n^{-1}\tilde{a}_r^T\tilde{\Sigma}\tilde{a}_s \quad r, s = 1, 2 \tag{4.4}$$

Let us clarify the meaning of the many elements contained in the previous expression. First of all, let us define $i_{\circ j}$ as the j-th element of the identity matrix and $w_{\circ j}$ as the j-th element of the weight matrix W. Secondly, let us define the transformed residuals as $\tilde{\varepsilon} = (I - \tilde{\rho}W)\tilde{u}$ and consequently the estimator of the diagonal elements of the variance-covariance matrix of the disturbances as $\tilde{\Sigma} = diag_{i=1}^{n}\left(\tilde{\varepsilon}_i^2\right)$.

Given such definitions we can now define the remaining elements of Equation (4.4) as follows:

$$A_1 = W^T W - \sum_{j=1}^{n} i_{\circ j} w_{\circ j}^T w_{\circ j} i_{\circ j}^T$$

$$A_2 = W$$

$$\tilde{a}_r = (I - \tilde{\rho}W)^{-1} H \tilde{P} \tilde{\alpha}_r$$

with $\tilde{\alpha}_r = -n^{-1}\left[Q^T(I - \tilde{\rho}W)(A_r + A_r^T)\tilde{\Sigma}(I - \tilde{\rho}W)\tilde{u}\right]$, $Q = [Z, W\gamma]$ and $\tilde{P} = (n^{-1}H^T H)^{-1}(n^{-1}H^T Q)\left[\left(n^{-1}Q^T H\right)(n^{-1}H^T H)^{-1}(n^{-1}H^T Q)\right]$.

The outcome of the weighted *GMM* procedure described above is a new estimator of the parameter ρ. Such an estimator has been proved by Kelejian and Prucha (2010) to be consistent and efficient. Let us call this new estimator $\hat{\rho}$.

Step 5: use the estimate $\hat{\rho}$ obtained in Step 4 to transform the original model in the usual way by premultiplying each side of Equation (4.1) by the term $(I - \hat{\rho}W)$:

$$(I - \hat{\rho}W)\gamma = (I - \hat{\rho}W)(Z\beta - \lambda W\gamma) + \varepsilon \tag{4.5}$$

Step 6: estimate β and λ in Equation (4.1) using a Generalized Spatial 2SLS procedure with H as instruments, thus obtaining:

$$\hat{\delta}_{GS2SLS} = \left[\hat{Q}^{*T}Q\right]\hat{Q}^{*T}y^{*} \tag{4.6}$$

where, as before, $Q = [Z, Wy]$, $Q^{*} = (I - \hat{\rho}W)Q$ and $\hat{Q}^{*} = H(H^{T}H)^{-1}H^{T}Q^{*}$

Step 7: Define the new GS2SLS residuals as:

$$\hat{u} = y - Q\hat{\delta}_{GS2SLS}$$

Step 8: Produce a final, efficient, GMM estimator of ρ based on the residuals of the GS2SLS procedure introduced in Step 7. Such an efficient estimator is obtained as a weighted *GMM* procedure that uses the elements of the matrix Ψ as weights.

4.1.3 Spatial HAC estimators

As seen, one crucial element in the procedure illustrated in the previous section is represented by the variance-covariance matrix of the limiting distribution of the normalized sample moments (the term Ψ). In one paper, on which we based the discussion of the previous section 4.1.2, Kelejian and Prucha (2010) suggested a parametric estimator for the elements of such a matrix. In another paper the same authors (Kelejian and Prucha, 2007) suggest estimating the matrix Ψ non-parametrically by using a heteroscedasticity and autocorrelation consistent (*HAC*) estimator. *HAC* estimators are rather popular in econometrics and have been studied for years especially in the time series literature after the seminal contributions of Grenander and Rosenblatt (1957) and Newey and West (1987). In the spatial context, the earlier attempts of estimating non-parametrically a variance-covariance matrix from a spatially dependent sample can be dated back to the contribution of Conley (1999), with respect to a continuous space, and of Driscoll and Kraay (1998) and Pinkse et al. (2002) for discrete spatial units. In the following discussion we will briefly summarize this method of estimation that can be used in Steps 4 and 8 of the procedure described in section 4.1.2.

Rather than estimating parametrically the elements of the matrix Ψ, as shown in Section 4.1.2 Kelejian and Prucha (2007) suggest that, under a series of working assumptions, consistent and efficient estimators of the elements of the matrix Ψ can be obtained non-parametrically. First of all consider that, as already shown in Equation (4.3), the true

variance-covariance matrix is defined as $\Psi = n^{-1}H^T\Sigma H$, so that the algebraic expression of its generic element ψ_{rs}, $r,s = 1,2$ is given by:

$$\psi_{rs} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ir}h_{is}\sigma_{ij} \tag{4.7}$$

with h_{ir}, h_{is} the *i-th* element of the vector of (respectively) the *r-th* and the *s-th* instrument.

By analogy, the spatial *HAC* estimator of the generic element of the Ψ matrix can be expressed as:

$$\hat{\psi}_{rs} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} h_{ir}h_{is}\hat{u}_i\hat{u}_j K(d_{ij}^{*}/d) \tag{4.8}$$

In the previous expression, $\hat{u}$ is an estimator for the disturbances u, $K(\circ)$ denotes a kernel function, d_{ij} represents the distance (however it is measured) between the *i-th* and the *j-th* spatial observation, $d_{ij}^{*} = d_{ij} + \upsilon_{ij}$ denotes the distance including a possible measurement error υ_{ij}, independent of the disturbances ε and such that $0 < |\upsilon_{ij}| < \infty$. Finally the term d represents a properly selected standardizing distance such that for $n \to \infty$, $d \uparrow \infty$.

As for the kernel function $K(\circ)$, which incorporates various ways of smoothing the function, various specifications are possible. Some of the most common specifications are shown here below:

1. Uniform

$$K(v) = \begin{cases} 0.5 & for \quad |v| \le 1 \\ 0 & otherwise \end{cases}$$

2. Bartlett (triangular)

$$K(v) = \begin{cases} 1-|v| & for \quad |v| \le 1 \\ 0 & otherwise \end{cases}$$

3. Epanechnikov (quadratic)

$$K(v) = \begin{cases} \frac{3}{4}(1-v^2) & for \quad |v| \le 1 \\ 0 & otherwise \end{cases}$$

4. Quadratic (bi-weight)

$$K(v) = \begin{cases} \frac{15}{16}(1-v^2)^2 & for \quad |v| \le 1 \\ 0 & otherwise \end{cases}$$

5. Quadratic (tri-weight)

$$K(v)=\begin{cases}\dfrac{35}{32}\left(1-v^2\right)^3 & \textit{for} \quad |v|\le 1\\ 0 & \textit{otherwise}\end{cases}$$

6. Gaussian

$$K(v)=\frac{1}{(2\pi)^{1/2}}\exp\left(\frac{1}{2}v^2\right)$$

7. Tukey–Hanning

$$K(v)=\begin{cases}\dfrac{1}{2}\left[1+\cos(\pi v)\right] & \textit{for} \quad |v|\le 1\\ 0 & \textit{otherwise}\end{cases}$$

8. Cosine

$$K(v)=\begin{cases}\dfrac{\pi}{4}\left[\cos\left(\dfrac{\pi}{2}v\right)\right] & \textit{for} \quad |v|\le 1\\ 0 & \textit{otherwise}\end{cases}$$

9. Partzen

$$K(v)=\begin{cases}1-6v^2+6|v|^3 & \textit{for} \quad |v|\le \dfrac{q}{2}\\ 2(1-|v|)^3 & \textit{for} \quad \dfrac{q}{2}|v|\le q\end{cases}$$

10. Quadratic spectral

$$K(v)=\frac{25}{12\pi^2v^2}\left(\frac{\sin(6\pi v)}{6\pi v}-\frac{\cos(6\pi v)}{5}\right)$$

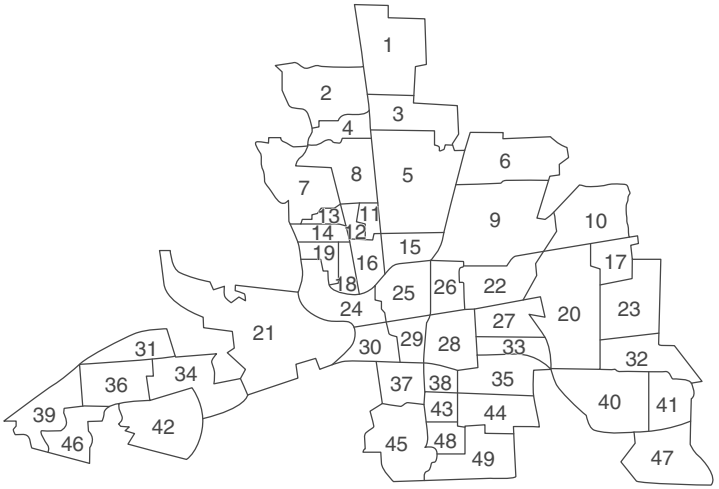
In the quoted paper Kelejian and Prucha (2007) proved the consistency of the estimators (4.8) and derived their asymptotic distributions.

Example 4.1 An urban study: The determinants of crime in Columbus (Ohio)

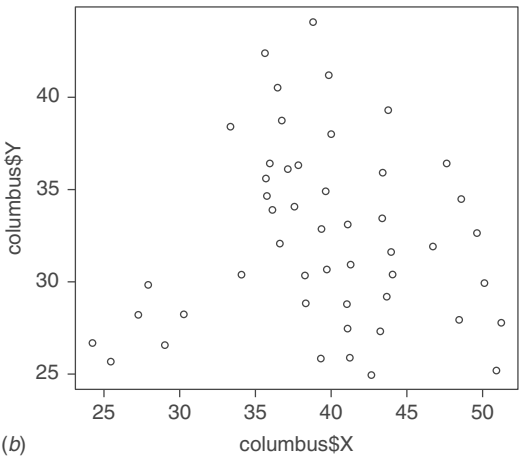
A very popular data set in spatial econometrics refers to an urban study on the determinants of crime in Columbus (Ohio). This data set was collected and studied for the first time by Anselin (1988) and then used several times

in the literature to test new procedures. Both the data and the maps necessary for a geographical analysis are available in `R`*. To retrieve them type* `data(columbus)` *and* `data(coldis)` *in an* `R` *session The following figures show the map of the 49 planning neighborhoods corresponding to census tracts of Columbus (Ohio) together with their centroids.*

(a) Map of the 49 neighborhoods of Columbus (Ohio) (b) map of the centroids of the 49 neighborhoods in Columbus (Ohio)



(*a*)



(*b*)

In addition, the following table shows the data related to a measure of crime (the dependent variable "crime", say y, defined as the total of residential burglaries and vehicle thefts per thousand households), and of the two independent variables "income" (say variable X) and "house value" (say variable Z) observed in the 49 neighborhoods.

NEIGHBORHOODS	CRIME	HOUSE VALUE	INCOME
1	15.726	80.467	19.531
2	18.802	44.567	21.232
3	30.626	26.350	15.956
4	32.387	33.200	4.477
5	50.731	23.225	11.252
6	26.066	28.750	16.029
7	0.178	75.000	8.438
8	38.425	37.125	11.337
9	30.515	52.600	17.586
10	34.000	96.400	13.598
11	62.275	19.700	7.467
12	56.705	19.900	10.048
13	46.716	41.700	9.549
14	57.066	42.900	9.963
15	48.585	18.000	9.873
16	54.838	18.800	7.625
17	36.868	41.750	9.798
18	43.962	60.000	13.185
19	54.521	30.600	11.618
20	0.224	81.267	31.070
21	40.074	19.975	10.655
22	33.705	30.450	11.709
23	20.048	47.733	21.155
24	38.297	53.200	14.236
25	61.299	17.900	8.461
26	40.969	20.300	8.085
27	52.794	34.100	10.822
28	56.919	22.850	7.856
29	60.750	32.500	8.681
30	68.892	22.500	13.906
31	17.677	31.800	16.940
32	19.145	40.300	18.942
33	41.968	23.600	9.918
34	23.974	28.450	14.948

(*continued*)

Continued

NEIGHBORHOODS	CRIME	HOUSE VALUE	INCOME
35	39.175	27.000	12.814
36	14.305	36.300	18.739
37	42.445	43.300	17.017
38	53.710	22.700	11.107
39	19.100	39.600	18.477
40	16.241	61.950	29.833
41	18.905	42.100	22.207
42	16.491	44.333	25.873
43	36.663	25.700	13.380
44	25.962	33.500	16.961
45	29.028	27.733	14.135
46	16.530	76.100	18.324
47	27.822	42.500	18.950
48	26.645	26.800	11.813
49	22.541	35.800	18.796

Let us start by estimating, for the sake of comparison, the homescedastic SARAR model defined by $y = \beta_0 + \beta_1 X + \beta_2 Z + \lambda Wy + u;\ u = \rho Wu + \varepsilon$ *with a constant error variance* $E(\varepsilon_i) = \sigma^2 \forall i$ *and with W a row-standardized weight matrix based on threshold distance of 3.5. We estimate the model, via the Generalized Spatial Two-Stage Least Squares procedure illustrated in Section 3.6.3. The results are shown here below.*

Parameter		***Standard Error***	***t-test***	***p-value***
β_0	*40.285663*	*8.912493*	*4.5201*	*6.18e–06****
β_1	*–0.860819*	*0.324179*	*–2.6554*	*0.0079218***
β_2	*–0.286200*	*0.085272*	*–3.3563*	*0.0007899****
λ	*0.492052*	*0.167476*	*2.9381*	*0.0033028***
ρ	*0.27048*			
σ^2	*85.65*			

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

All coefficients are significantly different from zero including the spatial effect λ *showing that the number of crimes in one census tract is significantly related to that of the neighboring census tracts.*

However, as is evident from the map, the various neighborhoods are very different in terms of their area and so the hypothesis of constant error variance is rather implausible. Indeed, we can expect the variance to be different from one area to another with a larger variability in larger areas. For this reason we want to estimate the same SARAR model as before $y = \beta_0 + \beta_1 X + \beta_2 Z + \lambda W y + u$; $u = \rho W u + \varepsilon$, *but now considering a space-varying error variance* $E(\varepsilon_i) = \sigma_i^2$. *Let us first consider the parametric estimation obtained via the modified Generalized Spatial Two-Stage Least Squares procedure illustrated in section 4.1.2. The results of the estimation procedure are shown in the following table.*

Parameter		***Standard Error***	***t-test***	***p-value***
β_0	*40.22132*	*6.41924*	*6.2657*	*3.710e–10****
β_1	*–0.85734*	*0.43197*	*–1.9847*	*0.04718**
β_2	*–0.28657*	*0.14316*	*–2.0017*	*0.04531*
λ	*0.49257*	*0.11578*	*4.2543*	*2.097e–05 ****
ρ	*0.27881*	*0.20754*	*1.3434*	*0.17914*

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

Wald test for ρ and λ both zero

	Observed Value	***p-value***
Wald test	*15.697*	*7.4348e–05*

The estimation results reveal that there is still a significant relationship of the variable "crime" with the variable "income", but, now that we considered a heteroscedasticity effect, the relationship with the variable "house value" is no longer significant (parameter β_2*). The contagion effect in crime* ($\lambda \neq 0$) *is confirmed to be significant in this new setting while no significant spatial correlation is found in the residuals* ($\rho = 0$). *The Wald test rejects the hypothesis that both parameters are equal to 0. Notice that in the heteroscedastic SARAR model we do not estimate the single variance parameter* σ^2 *anymore as it was clarified in section 4.1.2.*

Finally, let us estimate again the model using the non-parametric estimation of the variance-covariance matrix discussed in the spatial heteroscedastic autocorrelation consistent (HAC) version of the SARAR model (see section 4.1.3). In particular, in order to estimate non-parametrically the variance-covariance matrix reported in Equation (4.8) we considered a Triangular kernel function.

Parameter		*Standard error*	*t-test*	*p-value*
β_0	*41.79910*	*7.14254*	*5.8521*	*4.853e–09****
β_1	*–0.94449*	*0.46596*	*–2.0270*	*0.042665**
β_2	*–0.27756*	*0.15375*	*–1.8053*	*0.071031..*
λ	*0.47802*	*0.12523*	*3.8172*	*0.000135****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

The results are substantially similar to those obtained with the parametric estimator although with larger standard errors. Notice that in this new setting no estimate of the error correlation parameter appears because now the entire error correlation matrix (and not the single parameter ρ) is non-parametrically estimated.

4.2 Spatial models for binary response variables

4.2.1 Introduction

In all the previous sections of the book (chapter 3 and section 4.1), we have considered the case of a spatial regression related to continuous dependent variables; however, in many instances the dependent variable of a regression model may assume only a discrete number of possible outcomes. For example, we might be interested in the presence or absence of a certain technology in one region as a function of a number of independent variables. Similar situations emerge in explaining consumer choice between different shopping centers, in modeling count data, in studying spatial interaction, or in modeling electoral behavior or patient choices in health economics, in explaining criminal behavior and in many other situations. All these cases fall within the category of what is known in the literature as *discrete choice modeling* (see Greene, 2011). When the dependent variable is discrete rather than continuous, the models we have discussed so far cannot be employed and they need to be adjusted to follow their statistical nature. As the nature of this book is introductory, we will limit ourselves only to the case of binary choices and, within this context, we will only consider some of the possible model specifications. Notwithstanding the enormous interest of these models under an applied perspective, they have received comparatively less attention in the literature with respect to continuous dependent variable models discussed in Chapter 3 and section 4.1. This is partly due to the higher analytical complexity of the models, the need for more sophisticated estimation techniques and the increased

computational obstacles as we consider medium to large sample sizes. The interested reader is referred to the book by LeSage and Pace (2009) for a Bayesian point of view on the subject and to papers such as Beron and Vijverberg (2004), Fleming (2004) and Smirnov (2010) for reviews. The various specifications of spatial discrete choice models follow the general strategy used in the literature to deal with a-spatial discrete models, but adapted to account for spatial dependence employing the two fundamental paradigms of a Spatial Lag or of a Spatial Error discussed at length in chapter 3 that, by now, should be familiar to the reader. In the next section we will start presenting some a-spatial binary choice models that will form the basis for the spatial specifications that will be discussed later in this chapter.

4.2.2 The a-spatial logit and probit models

Let us start considering the following linear regression model:

$$y^{\bullet} = X\beta + \varepsilon \tag{4.9}$$

where $y^{\bullet}$ is a continuous variable to which we associate a binary variable defined as $y = I(y^{\bullet} > 0)$ with $I(.)$ the indicator function such that $I(a > 0) = \begin{cases} 1 & \text{if } a > 0 \\ 0 & \textit{otherwise} \end{cases}$. Equation (4.9) substantially represents the basic regression model presented in section 1.1, but expressed now in terms of a variable, $y^{\bullet}$ which cannot be observed (and, for this reason, is called *latent* variable), while the only observable variable y can assume the value 1 (e.g. presence) or 0 (e.g. absence). In Equation (4.9) the unobserved variable $y^{\bullet}$ may be thought of as a utility: when the utility is positive, the economic event materializes (e.g. a purchase in consumer choice). In this new setting Equation (4.9) is called a *latent regression,* whereas the term $X\beta$ is called the *index function* (Greene, 2011). From the above assumptions and from Equation (4.9) we have:

$$\begin{aligned} P(y_i = 1|X) &= P(y_i^{\bullet} > 0|X) = P(X_i\beta > \varepsilon|X) \\ &= P(\varepsilon < X_i\beta|X) = F(X_i\beta) = \int_{-\infty}^{\mu_i} f(\mu)d\mu \end{aligned} \tag{4.10}$$

similarly:

$$\begin{aligned} P(y_i = 0|X) &= P(y_i^{\bullet} < 0|X) = P(X_i\beta < \varepsilon|X) \\ &= P(\varepsilon > X_i\beta|X) = 1 - F(X_i\beta) = 1 - \int_{-\infty}^{\mu_i} f(\mu)d\mu \end{aligned} \tag{4.11}$$

In the previous expressions F is the disturbances' cumulative probability distribution function, $f = f(X\beta)$, the associated density function such that $\left(f = \frac{\partial F}{\partial X_i \beta}\right)$ and $\mu = X\beta$ the systematic component of the model. Different models can be defined by specifying different cumulative distribution functions in the previous expressions. In particular, two specifications are very popular in the econometrics literature, leading to two different binary choice models: (i) the *standardized logistic distribution* with zero expected value and variance $\pi^2/3$ and (ii) the *standardized normal distribution*. Since the value of the dichotomous variable y depends only on the sign of $y^{\bullet}$ and not its absolute value, it is not affected by the amount of the error variance so that it is not a limitation to standardize it to 1. If we assume the hypothesis of a logistic distribution for the disturbances, (with cumulative probability distribution function, say, Λ) Equations (4.10) and (4.11) define the so-called *Logit model*. Conversely, if we specify the innovations as being Normally distributed (with cumulative distribution function, say, Φ) Equations (4.10) and (4.11) define the *Probit model*. Although other models have been suggested, the Logit and the Probit models are by far the most commonly used frameworks in econometrics (Greene, 2011).

A popular estimation method for models (4.10) and (4.11) is based on the Maximum Likelihood strategy which will be now briefly outlined. First of all, consider that, since the dependent variable is binary, we can build up a likelihood function considering the probability of drawing a random sample of size n from a Bernoulli distribution, leading to:

$$L(\beta) = P(Y_1 = y_1, \ldots, Y_n = y_n | X) = P(Y_1 = y_1 | X) \ldots P(Y_n = y_n | X) \quad (4.12)$$

due to independence and, given definitions (4.10) and (4.11):

$$L(\beta) = \prod_{y_i=1} F(X_i\beta) \prod_{y_i=0} \left[1 - F(X_i\beta)\right] \quad (4.13)$$

where $\prod_{y_i=1}$ represents the product for all values of y such that $y_i = 1$ and similarly for $\prod_{y_i=0}$. Equation (4.13) can be re-written as:

$$L(\beta|X) = \prod_{i=1}^{n} F(X_i\beta)^{yi} \prod_{y_i=0} [1 - F(X_i\beta)]^{y_i - 1} \quad (4.14)$$

and consequently the log-likelihood as:

$$l(\beta) = \ln\left[L(\beta)\right] = \sum_{i=1}^{n} y_i \left\{\ln F\left(X_i\beta\right) + (1-y_i)\ln\left[1-F\left(X_i\beta\right)\right]\right\} \quad (4.15)$$

Apart from some practically irrelevant cases, Equation (4.15) is non-linear and cannot be maximized analytically so a numerical solution has to be identified. To obtain the maximum likelihood solution let us consider the score function:

$$\frac{\partial}{\partial \beta} l(\beta) = \sum_{i=1}^{n} \left[\frac{y_i f_i}{F_i} + (1-y_i)\frac{-f_i}{1-F_i}\right] x_i = 0 \quad (4.16)$$

In particular, if we assume the errors to be distributed according to the *logistic* distribution, for the score functions we have (Greene, 2011):

$$\frac{\partial}{\partial \beta} l(\beta) = \sum_{i=1}^{n} \left(y_i - \Lambda_i\right) X_i = 0 \quad (4.17)$$

with $\Lambda_i = \Lambda\left(X_i\beta\right)$. In this case the second derivate is equal to:

$$\frac{\partial^2}{\partial \beta^2} l(\beta) = \sum_{i=1}^{n} \Lambda_i \left(1-\Lambda_i\right) X_i X_i^T \quad (4.18)$$

and the expression can be used in confidence interval estimation and hypothesis testing.

In contrast, under the hypothesis of *normally* distributed innovations, we have that the score function is equal to:

$$\frac{\partial}{\partial \beta} l(\beta) = \sum_{y_i=0} \frac{-\phi_i}{1-\Phi_i} X_i + \sum_{y_i=1} \frac{\phi_i}{\Phi_i} X_i = 0 \quad (4.19)$$

with $\phi_i = \phi\left(X_i\beta\right)$, the standard normal density function such that $\phi = \dfrac{\partial \Phi(t)}{\partial t}$ while, in this case, the second derivate is:

$$\frac{\partial^2}{\partial \beta^2} l(\beta) = \sum_{i=1}^{n} -\kappa_i\left(\kappa_i + X_i\beta\right) X_i X_i^T \quad (4.20)$$

with $\kappa_i = \frac{(2y_i - 1)\phi\left[(2y_i - 1)X_i^T\beta\right]}{\Phi\left[(2y_i - 1)X_i^T\beta\right]}$. It should be noted that, both in the case of the Logit and the Probit models, the interpretation of the parameters is not as straightforward as in the case of the (a-spatial) linear regression model. In this case, in fact, the marginal effect of a unitary increase of the independent variables on the binary dependent variable is not simply the regression coefficient β, and assumes, instead, the value:

$$\frac{\partial E(y|X)}{\partial X} = f(X\beta)\beta \tag{4.21}$$

So, for the Logit model the marginal effect is equal to:

$$\frac{\partial E(y|X)}{\partial X} = \Lambda(X\beta)\left[1 - \Lambda(X\beta)\right]\beta \tag{4.22}$$

while for the Probit model it is:

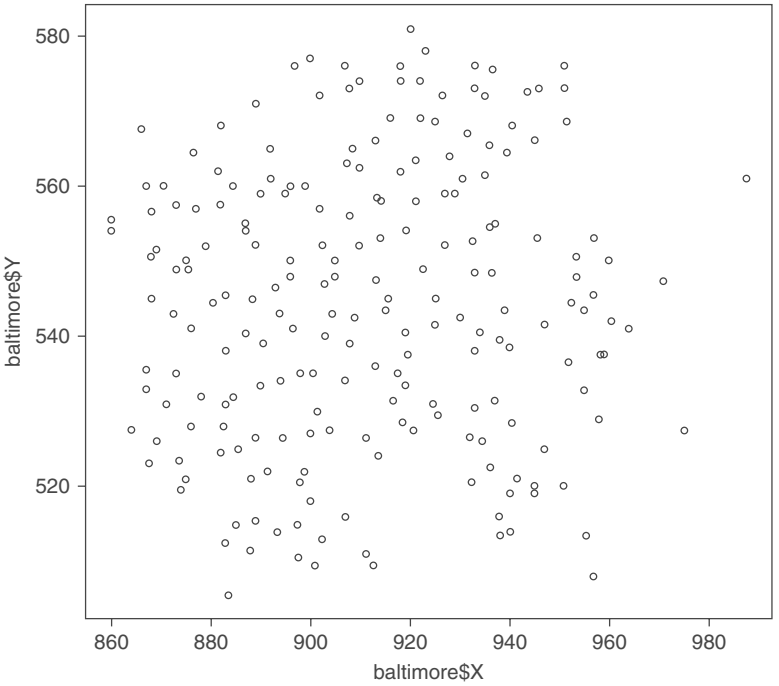
$$\frac{\partial E(y|X)}{\partial X} = \phi(X\beta)\beta \tag{4.23}$$

In order to judge the significance of the parameter estimates, most of the tools discussed in a linear regression framework are still available. In particular, in testing the significance of the single parameter estimates we can use the *t-test* (employing the standard errors derived from the Information matrix (4.18) or (4.20)) or the z-scores approximation, exploiting the property of asymptotic normality of the *ML* estimators. Furthermore, all likelihood-based tests (such as Likelihood ratio test, Wald test and Lagrange Multiplier) discussed in section 1.1 are also available. Finally, in order to measure the degree of fit of the model to the observed data, while the R^2 criterion is obviously not available (being based on the variance decomposition of the variable y which is now binary), we can still use the *AIC* and *BIC* criteria (see Equations (1.27) and (1.28)).

Example 4.2 Explaining luxury house prices in Baltimore

In order to illustrate the models for discrete choices described in this section, let us examine a set of data (prepared by Anselin and originally made available

by Dubin, 1992) on house price and a series of variables related to their value observed in 211 locations in Baltimore. The data can be downloaded from R using the command `data(baltimore)` *as illustrated in Section 2.3.5. The dataset also contains the X, Y coordinates of the houses. The map of their location is shown in the figure below.*



The dataset contains the following variables: Price of the house (PRICE), number of rooms (NROOM), number of bathrooms (NBATH), age of construction (AGE), square feet (SQFT) and other qualitative variables related to the presence or absence of important characteristics (like, e. g., fireplace, patio, air conditioning, garage and so on). For our purposes, we transformed the variable price into a binary variable classifying a house as "expensive" if its price is greater than $40,000. Our model is built up with the purpose of testing if the variables NROOM, NBATH, AGE and SQFT have a significant influence in the classification of a house as expensive. In other words, we want to test

whether the utility deriving from buying an expensive house can be explained by these characteristics. We will consider both the Logit and the Probit specifications of the model. The results of the ML estimations are shown below.

	Logit	***Probit***
Intercept	*–3.78782 (0.000201***)*	*–2.162195 (0.000136***)*
NROOM	*0.52624 (0.023392*)*	*0.252867 (0.048217*)*
NBATH	*0.56735 (0.106308)*	*0.365626 (0.072704)*
AGE	*–0.05183 (1.97e–05***)*	*–0.021129 (0.000775***)*
SQFT	*0.10725 (0.002539**)*	*0.057961 (0.003452**)*
AIC	*226.63*	*232.45*

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

The two models obviously lead to different estimates of the parameters, however the inferential conclusions are very similar: the AGE and the dimension of the house (SQFT) are the most significant factors in explaining the classification of an expensive house, followed by the number of rooms (NROOM) whereas the number of bathrooms (NBATH) appears to be less relevant. Although different in the absolute values, the sign of the parameters are always the same in the two model specifications and are in the expected direction. The AIC values are also comparable with only a slight relative advantage in the Probit model.

4.2.3 The Spatial Probit model

4.2.3.1 Generalities

When we observe spatial data, the binary models discussed in the previous section need to be adapted to account for spatial dependence. In the spatial econometrics literature the Probit specification is certainly more popular than the Logit version, due also to a severe criticism made by Anselin (2002), who noticed that in the spatial Logit version the error term is analytically intractable. On the other hand, Smirnov (2010) noticed that the spatial Probit presents the limitation that it cannot be easily extended to more than two alternatives (multivariate Probit, see Greene, 2011). A model that is particularly popular in the spatial econometrics literature is the Spatial Lag Probit model which can be expressed through the equation:

$$y^{\bullet} = \lambda W y^{\bullet} + Z\beta + u \tag{4.24}$$

with $y = I(y^{\bullet} > 0)$, $u|X \approx i.i.d.N(0, {}_{n}I_{n})$, W the usual weight matrix (however it is defined), $y^{\bullet}$ an n-by-1 vector of the unobservable continuous variable, λ the spatial autoregressive coefficient, y the observed

binary variable and Z the matrix of regressors, both current and spatially lagged, ($Z = [X, WX]$). As already noticed in the previous section, the error variance σ_u^2 is normalized to 1 to avoid identification problems with no loss in generalities. In the case described, by analogy with the case of the Spatial Lag model for continuous dependent variables described in section 3.5, a problem of endogeneity emerges in that the spatially lagged value of $y^{\bullet}$ is correlated with the stochastic disturbances in Equation (4.24). However, this is not the only estimation problem. In fact, when estimated on spatially dependent observations, the standard *ML* estimators are also inconsistent due to the heteroscedasticity induced by spatial dependence (Case, 1992; Pinkse and Slade, 1998). Furthermore, we also observe inefficiency as a consequence of the fact that we neglect the information contained in the off-diagonal terms of the non-spherical variance-covariance matrix (Fleming, 2004). Model (4.24) is expressed in the so called *structural* form (Fleming, 2004). Under the usual assumption that all the diagonal elements of W are zero and that $\lambda < 1$, it can also be expressed in a *reduced form* as follows:

$$y^{\bullet} = (I - \lambda W)^{-1}(Z\beta + u) = (I - \lambda W)^{-1} Z\beta + (I - \lambda W)^{-1} u = Z^{*}\beta + \varepsilon \quad (4.25)$$

with $\varepsilon = (I - \lambda W)^{-1} u$ and $\varepsilon \approx MVN(0, \Omega)$ and $Z^{*} = (I - \lambda W)^{-1} Z$. Since ε is expressed as a spatial autoregressive process, we have that, from the results presented in section 3.5.2.

$$E(\varepsilon\varepsilon^{T}) = \Omega = \left[(I - \lambda W)^{T} (I - \lambda W) \right]^{-1} \quad (4.26)$$

In the a-spatial Probit model we described the probability $P(y_i = 1)$ through the integral $P(y_i = 1) = F(Z_i\beta) = \int_{-\infty}^{\mu_i} f(\mu)\, d\mu$ and $\mu = (I - \lambda W) y^{\bullet} - Z_i\beta$ and similarly for $P(y_i = 0)$ (see Equations (4.10) and (4.11)) and these expressions define the likelihood of the single observation. In a similar fashion, in a spatial Probit model, we can define the likelihood of the single observation as:

$$\begin{aligned} P(y_i = 1) &= P(y_i^{\bullet} > 0 \,|\, Z_i^{*}, w_{ij} y_i^{\bullet}) \\ &= P(z_i^{*}\beta + \varepsilon_i > 0 \,|\, Z_i^{*}, w_{ij}, y_i^{\bullet}) \\ &= P(-\varepsilon_i \le Z_i^{*}\beta \,|\, Z_i^{*}, w_{ij}, y_i^{\bullet}) \cong \Phi(Z_i^{*}\beta) \end{aligned} \quad (4.27)$$

A similar expression can be derived for $P(y_i = 0)$. The approximation in the previous expression is due to the presence of heteroscedasticity

which needs to be properly taken into account. In fact, generalizing the result obtained for the a-spatial Probit model (see Equation (4.10)), in this case we have:

$$P(y_i = 1 \left| z_i^*, w_{ij} y_i^\bullet \right.) = \Phi\left(\frac{z_i^* \beta}{\Omega_{ii}}\right) \tag{4.28}$$

where Ω_{ii} is the i-th diagonal element of the variance-covariance matrix (4.26). So, even if the error terms u are homescedastic, the transformed error terms ε are heteroscedastic.

In the a-spatial Probit model, exploiting independence, the marginal probabilities are combined multiplicatively in the likelihood function (4.13). However, in the spatial context this simplification is not possible and, in order to build up the likelihood, we have to evaluate simultaneously the joint probabilities at each location:

$$P(y_1 = 1, y_2 = 1, \ldots, y_n = 1); \quad y_i = \{0;1\} \tag{4.29}$$

In principle this probability should be evaluated through the n-dimensional integral:

$$P(y_1 = 1, y_2 = 1, \ldots, y_n = 1;) = \int_{-\infty}^{\mu_1} \int_{-\infty}^{\mu_2} \ldots \int_{-\infty}^{\mu_n} \phi(\mu) d\mu \tag{4.30}$$

with $\phi(\mu)$ now representing the density function of the n-dimensional Multivariate Normal distribution expressed by:

$$\phi(\mu) = (2\pi)^{-\frac{n}{2}} \left|\Omega\right| \exp\left[-\frac{1}{2}\left(\mu^T \Omega \mu\right)\right] \tag{4.31}$$

However, there is no analytical solution for a univariate normal probability distribution function and the problem becomes even more severe in the multivariate case. Since the multiple integral contained in Equation (4.30) cannot be evaluated analytically the only solution is that of obtaining numerical approximations.

As an additional problem, the presence of spatial dependence also introduces problems in the interpretation of the marginal effects which in this case for the *i*-th observation is:

$$\frac{\partial E(y \left| X, w_{ij} y_j^\bullet \right.)}{\partial X_i} = \phi\left(\frac{Z_i^* \beta_i}{\sigma_{\varepsilon,ii}}\right) \frac{\beta_i^*}{\Omega_{ii}} \tag{4.31}$$

where, with the now obvious notation, we have set $\beta_i^* = \left[(I - \lambda W)^{-1}\right]_{ii} \beta_i$.

As stated, there are two major problems in defining an optimal estimation strategy: a problem of endogeneity and a problem of non-sphericity (both in the form of autocorrelation and of heteroscedasticity) of the variance-covariance matrix. In the literature, various likelihood-based and moments-based alternatives were proposed, starting from the early contributions of McMillen (1992), Pinkse and Slade (1998) and Klier and McMillen (2008) amongst others (see Fleming, 2004 for a review). In particular, here we will discuss the following estimation strategies:

(i) Maximum likelihood (ML),
(ii) Generalized Method of Moments (GMM), and
(iii) A Linearized version of the GMM (LGMM)

A maximum likelihood estimator cannot be found analytically while the numerical procedures may computationally be very demanding. The GMM approach reduces, but does not eliminate completely the computational problems especially when dealing with very large datasets. Its linearized version represents a trade-off between accuracy and computational efficiency.

4.2.3.2 Maximum Likelihood estimation

In the standard a-spatial Probit model, the maximum likelihood estimators can be derived by maximizing Equation (4.15), a procedure that requires numerical optimization due to the high degree of non-linearity in the parameters. However, if we consider spatial data, by wrongly assuming independent errors, the likelihood function maximization produces estimators that are still consistent, but no longer efficient due to the fact that they neglect the information contained in the off-diagonal terms of the variance-covariance matrix. From Equation (4.15), considering the hypothesis of normal disturbance and the Spatial Lag specification expressed in Equation (4.24), the spatial Probit log-likelihood function for a sample of dimension n can be expressed as:

$$l(\beta,\lambda\,|\,Z,W\mathbf{y}^{\bullet}) = \sum_{i=1}^{n} y_i \ln\Phi\left(\frac{\lambda\sum_{i=1}^{n} w_{ij}y_j^{\bullet} + Z_i^T\beta}{\Omega_{ii}}\right) + \sum_{i=1}^{n}(1-y_i)\ln\Phi\left(1-\frac{\lambda\sum_{i=1}^{n} w_{ij}y_j^{\bullet} + Z_i^T\beta}{\Omega_{ii}}\right) \tag{4.32}$$

The term $y^{\bullet}$ is not observable and so, in order to derive an operational likelihood function, we have to exploit the reduced form of the model. Using (4.25), the log-likelihood function can be re-written as:

$$l(\beta,\lambda\,|Z,Wy^{\bullet}) = \sum_{i=1}^{n} y_i \ln\Phi\left(\frac{Z_i^{\star T}\beta}{\Omega_{ii}}\right) + \sum_{i=1}^{n}(1-y_i)\ln\Phi\left(1-\frac{Z_i^{\star T}\beta}{\Omega_{ii}}\right) \quad (4.33)$$

Expression (4.33) has a complicated structure that creates computational problems in the phase of the numerical maximization. To maximize such an expression, McMillen (1992) suggested the use of the *EM* algorithm. The *EM* algorithm (first introduced by Dempster et al., 1977) is an iterative procedure that develops in two steps: an E-step (expectation) and an M-step (maximization). The E-step consists of calculating the expectation of the likelihood using a starting value for the unknown parameters. The M-step consists of maximizing the expected likelihood found in the E-step with respect to the unknown parameters. The two steps are iterated until the parameters converge to a stable solution, which can be proved to coincide with the *ML* estimators of the original likelihood. McMillen (1992) suggested generalizing the *EM* algorithm to the spatial case, by replacing the binary variable with the expectation of the underlying latent variable and to maximize the expectation of the log-likelihood as if the artificial variable was the actual latent variable. In particular, in the case of the Spatial Lag Probit model (4.25), the expected value can be derived as:

$$E(y_i^{\bullet}\,|y_i = 1) = Z_i\beta + E(\varepsilon_i\,|\varepsilon_i > -Z_i^{\star}\beta) + \sigma_i \frac{\phi\left(Z_i^{\star}\beta \big/ \Omega_{ii}\right)}{\Phi\left(Z_i^{\star}\beta \big/ \Omega_{ii}\right)} \quad (4.34)$$

The expected value $y_i^{\bullet}$ thus obtained can be used in an *M*-step where the following log-likelihood is maximized:

$$l = \cos t - \frac{1}{2}\ln|\Omega| - \frac{1}{2}\left(\mu^T\mu\right) \quad (4.35)$$

In this last expression, $\Omega = (I-\lambda W)^{-1}(I-\lambda W)^{-T}$, $\mu = (I-\lambda W)\hat{y}^{\bullet} - Z_i\beta$, and $\hat{y}^{\bullet}$ is the vector of the predicted values of the latent variable derived in the *E*-step.

The use of the *EM* approach presents two major problems. First of all, the variance-covariance matrix Ω which is present in Equation (4.35) is

unknown and needs to be properly estimated. In this respect, McMillen (1992) suggested interpreting the Probit model as a non-linear weighted least square model conditional on the spatial parameter (Amemiya, 1985), but unfortunately this approach leads to biased estimators (Fleming, 2004). Secondly, the estimation process can be very slow because the approach requires the calculation of the determinant of the matrix Ω at each iteration of the *M*-step until the process converges. Not only can this operation be very heavy computationally, but the calculation of the determinant is based on approximations that can become inaccurate when n is large and W dense that is, it contains a large number of non-zero entries.

4.2.3.3 Generalized Method of Moments estimation

An alternative estimation method for a spatial version of a binary choice model was introduced by Pinkse and Slade (1998) who developed a *GMM* technique for a Probit model expressed in the form of a Spatial Error with heteroscedasticity.

Consider the following model expressed in a latent variable form

$$y^{\bullet} = Z\beta + u \tag{4.36}$$

$$u = \rho Wu + \varepsilon \tag{4.37}$$

with $\varepsilon \,|Z \approx i.i.d\ N(0,\sigma_{\varepsilon}^2 I)$. Similar to model (4.24), Equations (4.36) and (4.37) can be written in a reduced form as:

$$y^{\bullet} = Z\beta + (I - \rho W)^{-1}\varepsilon \tag{4.38}$$

where $u = MVN(0,\Omega)$ (see section 3.4) and $\Omega = (I - \rho W)^{-1}(I - \rho W^T)^{-1}$ (see Equation 3.15). Using Equation (4.15) the Spatial Error Probit model leads to the following log-likelihood function:

$$l(\beta,\rho) = \ln\left[L(\beta,\rho)\right] = \sum_{i=1}^{n} y_i \left\{\ln F\left(X_i\beta\right) + (1-y_i)\ln\left[1 - F\left(X_i\beta\right)\right]\right\} \tag{4.39}$$

which, in the case of the normal distribution (Probit model) and, using expression (4.28) can be written as:

$$l(\beta,\rho\,|Z,Wy^{\bullet}) = \sum_{i=1}^{n} y_i \ln\Phi\left(\frac{z_i^T\beta}{\Omega_{ii}}\right) + \sum_{i=1}^{n}(1-y_i)\ln\Phi\left(1 - \frac{z_i^T\beta}{\Omega_{ii}}\right) \tag{4.40}$$

Since the traditional notion of disturbances in this new context has no meaning, let us now introduce the notion of the *generalized errors* which, in the context of a Probit model, can be defined as:

$$\tilde{u}_i(\theta) = \left\{ y_i - \Phi\left(\frac{z_i^T \beta}{\sigma_i} \right) \right\} \frac{\phi\left(\frac{z_i^T \beta}{\sigma_i} \right)}{\Phi\left(\frac{z_i^T \beta}{\sigma_i} \right) \left\{ 1 - \Phi\left(\frac{z_i^T \beta}{\sigma_i} \right) \right\}} \tag{4.41}$$

Let us further consider a set of *k* instruments whose observations are arranged in a *n*-by-*k* matrix *H*. The instruments are exogenous by definition, thus suggesting the moments condition:

$$E\left(H^T \tilde{u} \right) = 0 \tag{4.42}$$

or, for the single *i*-th condition

$$E\left[h_i \frac{\left\{ y_i - \Phi\left(\frac{z_i^T \beta}{\sigma_i} \right) \right\} \phi\left(\frac{z_i^T \beta}{\sigma_i} \right)}{\Phi\left(\frac{z_i^T \beta}{\sigma_i} \right) \left\{ 1 - \Phi\left(\frac{z_i^T \beta}{\sigma_i} \right) \right\}} \right] = 0 \tag{4.43}$$

where h_i indicates the *i*-th row of a matrix of instruments *H*. Finally, using the sample analogue to Equation (4.43), we have the following set of conditions:

$$m(\beta, \rho) = \frac{1}{n} \sum_{i=1}^{n} h_i \frac{\left\{ y_i - \Phi\left(\frac{z_i^T \beta}{\sigma_i} \right) \right\} \phi\left(\frac{z_i^T \beta}{\sigma_i} \right)}{\Phi\left(\frac{z_i^T \beta}{\sigma_i} \right) \left\{ 1 - \Phi\left(\frac{z_i^T \beta}{\sigma_i} \right) \right\}} \tag{4.44}$$

In a *GMM* procedure the number of the moments conditions is larger than the number of parameters to be estimated, so that the estimators are the solutions to the minimization problem:

$$m(\beta, \rho)^T M^{-1} m(\beta, \rho) = \min \tag{4.45}$$

M being a positive definite matrix which defines the weights assigned to each sample moments $m(\beta, \rho)$. Pinkse and Slade (1998) prove the consistency and asymptotic normality of the GMM procedure and derived

the variance-covariance matrix of the estimators developed within a Newey and West (1987) framework.

The GMM estimator presents us two major advantages with respect to the Maximum Likelihood procedure illustrated in section 5.2.3.2. First of all, it does not rely on the assumption of normality of the disturbances. Secondly, it does not require calculating the determinant and the inverse of matrices of dimension *n*-by-*n*. However, expression (4.45) cannot be calculated in an analytical way and so also the GMM estimator can only be identified by numerically solving the minimization problem. This requires the evaluation of the variance-covariance matrix Ω repeatedly for each candidate value of the parameter ρ in a numerical search and this operation can be computationally challenging due to the complex form of Ω.

4.2.3.4 A linearized version of the GMM

To overcome the computational difficulties connected with both the *ML* and the *GMM* approaches, more recently Klier and McMillen (2008) proposed a linearized version of Pinkse and Slade's *GMM* approach which avoids the problem of inverting *n-by-n* matrices. The paper applies the methodology to the estimation of a spatial Logit model, instead of the Probit model considered by Pinkse and Slade (1998). The authors start from the consideration that a spatial model is always an approximation since the true structure of dependence of the disturbances is generally unknown. So, they suggest making the approximation explicit and to linearize the non-linear model by expanding it around a reasonable starting point. This procedure will be now described.

Consider, to start, the Spatial Lag Logit model:

$$y^{\bullet} = \lambda W y^{\bullet} + Z\beta + u \tag{4.46}$$

with $u|X \approx i.i.d.\ Logistic(0, \sigma^2_{u\,n} I_n)$ or expressed in the reduced form:

$$y^{\bullet} = (I - \lambda W)^{-1}(Z\beta + u) = (I - \lambda W)^{-1} Z\beta + (I - \lambda W)^{-1} u = Z^{*}\beta + \varepsilon \tag{4.47}$$

as in (4.25). In (4.47) we have that the transformed errors are no longer homoscedastic so that $Var(\varepsilon_i) = \sigma_i^2$ or in matrix notation $Var(\varepsilon) = \hat{\Sigma}^T\hat{\Sigma}$ with $\hat{\Sigma}$ a diagonal matrix such that:

$$\hat{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & & & 0 \\ & & & & \\ & & & & \\ 0 & 0 & & & \sigma_n \end{bmatrix} \tag{4.48}$$

Let us now define the probability of success as:

$$P_i = P(y_1 = 1) = \frac{\exp(Z_i^{**}\beta)}{1+\exp(Z_i^{**}\beta)} \tag{4.49}$$

due to the hypothesis that the errors are distributed according to the logistic law and with Z^{**} representing the variable Z transformed as in Equation (4.25) (that is $Z^{*} = (I - \lambda W)^{-1}Z$), but also normalized to the heteroscedastic variances as $Z^{**} = Z^{*}\hat{\Sigma}^{-1}$.

In this new setting, the generalized Logit residuals can be simply defined as:

$$\tilde{u}_i = y_i - P_i \tag{4.50}$$

Klier and McMillen (2008) suggest a GMM procedure which develops along the following steps:

Step 1: Assume an initial value for the vector of parameters, say $\delta = (\beta, \lambda)$ and call these initial values $\delta_0 = (\beta_0, \lambda_0))$

Step 2: Use these initial values in Equation (4.49) and calculate through them the generalized residuals given by Equation (4.50). Call these generalized residuals $\tilde{u}_i^0 = y_i - P_{i0}$

Step 3: Calculate the gradient terms defined as:

$$G_{\delta_0} = \frac{\partial P_{i0}}{\partial \delta_0}$$

Step 4: Regress the values thus obtained of G_{δ_0} on a set of instruments, say H, defined as

$$H = (I - \lambda W)^{-1}WZ^{**}$$

Let us call $\hat{G}_{\delta_0}$ the estimates of the gradients thus obtained.

Step 5: Construct the new estimates of the parameters using the updating expression

$$\delta_1 = \delta_0 + (\hat{G}_0^T\hat{G}_0)^{-1}\hat{G}_0^T\tilde{u}^0$$

In their contribution Klier and McMillen (2008) derive the explicit expressions for the gradients of β and λ, given by:

$$G_{\beta_i} = P_i(1-P_i)Z_i^{**} \tag{4.51}$$

and

$$G_{\lambda_i} = P_i(1-P_i)\left[H_i\beta - \frac{Z_i^{**}\beta}{\sigma_i^2}\Xi_{ii}\right] \tag{4.52}$$

with Ξ_{ii} the i-th diagonal element of matrix $\Xi = (I-\lambda W)^{-1}W(I-\lambda W)^{-1}$ $(I-\lambda W)^{-1}$.

The *GMM* procedure described above is still computationally heavy due to the fact that each step of the iteration requires the inversion of the matrix $(I-\lambda W)^{-1}$. For this reason, in the same paper Klier and McMillen (2008) take a step further and propose linearizing the model around the starting point $\lambda_0 = 0$. In this starting point, in fact, no matrix needs to be inverted because $(I-\lambda W)^{-1} = I$. Having made such assumptions, the gradient terms simplify substantially. In fact, expanding the generalized error around the initial estimates $\delta_0 = (\beta_0, \lambda_0)$ and stopping at the first linear term of the expansion, we have:

$$\tilde{u}_i \cong \tilde{u}_i^0 + G(\delta - \delta_0) \tag{4.53}$$

Now define the transformed generalized errors:

$$v_i = \tilde{u}_i^0 + G\delta_0 - G\delta \tag{4.54}$$

and let $M = \left(H^T H\right)^{-1}$. The objective function to minimize becomes:

$$v^T H\left(H^T H\right)H^T v \tag{4.55}$$

In such a way, no matrix inversion is required and the procedure reduces to an a-spatial Logit estimation followed by a Two-Stage Least Squares procedure.

If the true structure of the model is captured by model (4.46) and (4.47), Klier and McMillen (2008) show that the linearization provides accurate estimates as long as the parameter λ is small. Indeed, when $\lambda < 0.5$ there is no bias, although when $\lambda > 0.5$ the estimates are upward biased. In general, the estimators obtained through the linearized model provide a good approximation to the unknown

parameters although a certain loss of efficiency is the price to pay for the increased computational performance.

4.2.3.5 Further computational solutions

The need for computationally efficient procedures is still felt to be an important issue in binary choice modeling, even with the current powerful computer machines (see also chapter 5). In order to reduce the computational burden, LeSage and Pace (2009) suggest the use of Bayesian techniques to simulate the probabilities in a Monte Carlo Markov Chain (*MCMC*) context. The authors report that the use of a Gibbs sampler to estimate a spatial Probit model with only 2 independent variables is a computationally intensive operation. In a simulation experiment LeSage and Pace (2009) reported that with a sample size of $n = 400$ and 1,200 draws of the *MCMC* sampler (each repeated m times in the m step of the Gibbs procedure), the calculations required 20 minutes for $m = 10$ and 5 minutes for $m = 2$. The time increase is proportional to n. Therefore if, for instance, the sample size goes up to $n = 10{,}000$ the time required increases to 8 hours and 49 minutes. Furthermore, they also report that the computing time is less than proportional to m so, even reducing the m step to, say, $m = 1$ (at the expenses of accuracy) the time is still 1 hour and 21 minutes, which is not negligible. Thus the approach is limited to small samples because it does not eliminate the problem of inverting the n-by-n W matrices. Beron and Vijverberg (2004) proposed a further alternative based on the *GHK* simulator to evaluate the n-dimension integral, but without succeeding to substantially reducing the computational burden. Wang et al. (2013) suggest the use of a partial bivariate likelihood, an approach that will be discussed to a deeper extent in section 5.4.

4.2.3.6 Further spatial discrete choice models

Apart from the simple binary discrete choice models considered in this section, in the econometric literature we find various other specifications of discrete choice models including bivariate and multivariate Probit and Logit, ordered Probit and Logit, truncation, censoring, sample selection, and models for count data and duration (see Greene, 2011). Some of these topics have been dealt with in the spatial context (for example, by Wang and Kockelman, 2009), but the field is still largely unexplored.

Example 4.3 Luxury house prices in Baltimore (continued)

Let us now go back to the dataset considered in Example 4.2 and let us test if there are spatial effects in the classification of "expensive" houses in

Baltimore. In the table here below we show the results of the estimation of a spatial Probit model using the three methodologies presented in the previous section (Maximum Likelihood, GMM and linearized GMM). In all three cases we used a row-standardized W matrix based on minimum threshold distance and defining as neighbors points falling within a distance < 22. To facilitate the comparison we also show again the results of the estimation of the a-spatial Probit model already considered in Example 4.2. Notice the spatial models now include the extra parameter λ that captures the effects due to spatial dependence in the latent variable.

	A-spatial Probit	***Spatial Probit (ML)***	***Spatial Probit (GMM)***	***Spatial Probit (linearized GMM)***
Intercept	*–2.162195 (0.000136***)*	*–1.02618685 (0.0930688)*	*–0.99759055 (0.04417*)*	*–1.73088 (0.00163***)*
NROOM	*0.252867 (0.048217*)*	*0.21958292 (0.2186963)*	*0.18258271 (0.1712854)*	*0.22432 (0.04927*)*
NBATH	*0.365626 (0.072704)*	*–0.06945085 (0.7766100)*	*0.01589953 (0.9363988)*	*0.21695 (0.27898)*
AGE	*–0.021129 (0.000775***)*	*–0.02364809 (0.000***)*	*–0.02040384 (0.01461387*)*	*–0.01437 (0.03206*)*
SQFT	*0.057961 (0.003452**)*	*0.04463416 (0.08450417.0)*	*0.04038605 (0.02747669*)*	*0.04102 (0.04998*)*
λ		*0.81640265 (0.000***)*	*0.78396376 (0.0000014***)*	*1.04295 (0.00002***)*

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

The first observation is that, no matter what estimation method is selected, the spatial parameter λ is always highly significant. From the comparison between the four models, a dominant role of the variable AGE also emerges as an explicative factor, whereas the role of variables like NROOM and SQFT becomes less relevant with respect to the a-spatial version of the model. This effect is due to the fact that much of the phenomenon is already explained by the spatial variations of the latent variable.

4.3 Spatial panel data models (Written by Giovanni Millo)

4.3.1 Generalities

Spatial panels are a special case of *panel* where data are observed on two dimensions: across spatial units and over time. Panel data models have

become widespread with the availability of databases containing multiple observations on individual units, for example continuously updated countries' and administrative records, periodical national surveys, and repeated measurements of various phenomena in different moments in time. Let us begin with the a-spatial panel data model which can be expressed through the following equation:

$$y_{it} = \alpha + \beta^T X_{it} + \varepsilon_{it} \tag{4.56}$$

In Equation (4.56) the index $i = 1, \ldots n$ refers to individuals (groups, countries, regions and so on), the index $t = 1, \ldots, T$ is the time index, X_{it} is a non-stochastic vector of observations of the independent variables in individual i and time t, ε_{it} the innovation term such that $\varepsilon \,|X \approx i.i.d.N(0, \sigma_{\varepsilon}^2{}_n I_n)$ and α and β are parameters to be estimated. With respect to the previous specifications of the analogous cross-sectional models (see chapter 1), in this context we prefer to indicate the constant α with a different symbol for reasons that will become apparent later. Panel datasets can have a large number of cross-sectional units observed over a few points in time (*short panels*, typical of microdata), a limited number of relatively long time series (*long panels*, or *pooled time series*, typical of financial or macroeconomic data), or even a balanced behavior between the two dimensions. The typical *spatial panel* leans towards short time series with large spatial dimensions as it usually consists of repeated observations over a sizable cross-section of spatially referenced data, such as countries of the world, regions within one country, or geographical areas. In all these cases, the double dimensionality of panel data allows for richer modeling possibilities than one single cross-section or time series. In particular, panel data models are used to control for *unobserved heterogeneity* related to individual-specific, time-invariant characteristics which are difficult or even impossible to observe, but might lead to biased or inefficient estimates of the parameters of interest if omitted. This topic will be introduced in the next section.

4.3.2 Unobserved heterogeneity and individual effects

To model individual heterogeneity, one often assumes that the error term has two separate components, one of which is specific to the individual and does not change over time. This is called the *unobserved effects* model. Formally we can express this model as follows:

$$y_{it} = \alpha_i + \beta^T x_{it} + u_i = \alpha_i + \beta^T x_{it} + (\mu_i + \varepsilon_{it}) \tag{4.57}$$

In the previous expression the overall (or *composite*) error u_{it} is broken down into two terms. The first term (μ_i) represents *the individual error component* typical of location *i,* while the second term (ε_{it}) represents the *idiosyncratic error component* and it is usually assumed to be well-behaved and independent from both the independent variables and the individual error component. The appropriate estimation method for the model (4.57) depends fundamentally on the properties of the two error components. The individual component may be either correlated or independent from the independent variables. If it is correlated, the ordinary least squares (*OLS*) estimator for β would be inconsistent, so it is customary to treat the μ_i as a further set of n parameters to be estimated, as if in the general model we would have n different intercepts, say $\alpha_{it} = \alpha_i$, constant with respect to t. In this sense, the error component μ_i will be assumed to include the intercept term α_i. This model is referred to in the literature as the *fixed effects* (or *within* or *least squares* dummy variables) model and it is usually estimated using *OLS* on transformed data, a procedure that guarantees consistent estimates for β (see Baltagi, 2008). Conversely, if the individual-specific component μ_i is assumed to be uncorrelated with the regressors, we have a situation which is usually referred to in the literature as *random effects*. In this case, the overall error, u_{it}, is also uncorrelated with the independent variables so that the *OLS* criterion leads to consistent estimators. However, the common error component over individuals induces a correlation across the composite error terms, which is responsible for a loss in the efficiency of the *OLS* estimator, so that we have to resort to some form of feasible generalized least squares (*GLS*) estimators based on the estimation of the variance of the two error components. Finally, if the individual component is missing altogether, a pooled *OLS* estimation is the most efficient criterion for β. This set of assumptions is usually labeled as *pooling model*, although this term more formally refers to the properties of the errors and to the associated appropriate estimation method rather than to the model itself. The panel literature has recently considered panel regression models with a spatially lagged dependent variable or spatially autocorrelated disturbances, both in the context of fixed and random effects specifications (Lee and Yu, 2011). We will devote the next sections of this chapter to discussing some of the most commonly used specifications and related estimation techniques.

Example 4.4 Munnell's model of public capital productivity

The Munnell (1990) model of public capital productivity expresses the level of Gross Domestic Product (GDP) in each state of the US (excluding the

"islands" Alaska and Hawaii and the district of Columbia, for a total of 48 states; see Example 3.2 for the map) observed over the years between 1970 and 1986, through a production function where the explanatory variables considered are the endowment of public capital (roads, water facilities and other utilities, coded as "pcap"), private capital (coded as "pc") and employment ("emp"). The unemployment rate ("unemp") is added in order to proxy for the effects of the business cycle. The model can be expressed formally as follows:

$$\log(GDP)_{it} = \alpha + \beta_1 \log(pcap_{it}) + \beta_2 \log(pc_{it}) + \beta_3 \log(emp_{it}) + \beta_4 unemp_{it} + \varepsilon_{it}$$

In this example we consider a subset of the original data including only the observations between 1970 and 1974. The dataset used are not shown here for succinctness. The original model proposed by Munnell (1990) was estimated using OLS, which amounts to pooling all observations together under the assumption of no individual components. The results are shown in the table here below:

Parameter		***Standard Error***	***t-test***	***p-value***
Intercept	*1.2054162*	*0.1151322*	*10.4698*	*< 2.2e–16****
log(pcap)	*0.2140447*	*0.0355133*	*6.0272*	*6.405e–09****
log(pc)	*0.3421465*	*0.0212932*	*16.0684*	*< 2.2e–16****
log(emp)	*0.5113731*	*0.0261593*	*19.5484*	*< 2.2e–16****
unemp	*0.0118651*	*0.0047056*	*2.5215*	*0.01235**

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

where the intercept and all variables considered in the model are significant. However, as Baltagi and Pinnoi (1995) observed, OLS are inefficient if we suspect the errors to display a random component, and random effects estimators can provide better estimates. A random effects model estimated on the same data provides the following results:

Parameter		***Standard Error***	***t-test***	***p-value***
Intercept	*1.6794841*	*0.1961392*	*8.5627*	*< 2.2e–16****
log(pcap)	*0.0946853*	*0.0525858*	*1.8006*	*0.0717678*
log(pc)	*0.3411077*	*0.0424213*	*8.0409*	*8.915e–16****
log(emp)	*0.6272028*	*0.0377440*	*16.6173*	*< 2.2e–16****
unemp	*–0.0100228*	*0.0026808*	*–3.7388*	*0.0001849****
variance component	*17.5966*	*4.2466*	*4.1437*	*3.418e–05****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

As it turns out, the significance of public capital is spurious, and it disappears if individual error components are considered (see also Example 3 in Baltagi, 2008).

4.3.3 Spatial Panel Models with random effects

In a random effects specification, the unobserved individual effects are assumed to be uncorrelated with the other explanatory variables in the model, and can therefore be safely treated as components of the error term (see, for example, Wooldridge, 2002). Within this context we can consider two alternative specifications: the Spatial Error model with random effects (*SEM-RE*) and the so-called *KKP* model from the initials of the authors who originally proposed the method (see Kapoor et al., 2007). We will treat the two specifications here below.

4.3.3.1 Random Effect Spatial Error Model (SEM-RE)

Let us start by assuming that the individual error component is $\mu_i \sim$ i.i.d.N(0, σ_μ^2) and that the idiosyncratic error term ε obeys a Spatial Error formulation such that, in each moment of time (t=1,…, T), we can write:

$$\varepsilon_{it} = \rho W \varepsilon_{it} + \eta_i \tag{4.58}$$

as in section 3.4 and consequently $\varepsilon_{it} = (I - \rho W)^{-1}\eta_i$. Let us now introduce the symbol ⊗ (known in matrix algebra as the *Kronecker product*) such that if $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$, then $A \otimes C = \begin{bmatrix} a_{11}C & a_{12}C \\ a_{21}C & a_{22}C \end{bmatrix}$ $= \begin{bmatrix} a_{11}c_{11} & a_{11}c_{12} & a_{11}c_{11} & a_{21}c_{12} \\ a_{11}c_{21} & a_{11}c_{22} & a_{21}c_{21} & a_{21}c_{22} \\ a_{12}c_{11} & a_{12}c_{12} & a_{22}c_{11} & a_{22}c_{12} \\ a_{12}c_{21} & a_{12}c_{22} & a_{22}c_{21} & a_{22}c_{22} \end{bmatrix}$. Furthermore, in order to simplify the notation, from now on let us also define the matrix $B = (I_n - \rho W)$, with I_n an n-by-n identity matrix, W the usual spatial weight matrix and ρ the spatial error dependence parameter. So, for the whole panel we can rewrite the idiosyncratic error in a compact way as:

$$\varepsilon = \left(I_T \otimes B^{-1}\right)\eta \tag{4.59}$$

with η an nT-by-1 vector of the disturbances such that $\eta \approx n.i.d.N(0, \sigma_\eta^2)$. As a consequence, the composite error term $u_i = (\mu_i + \varepsilon_{it})$ can be written in a compact way as:

$$u = \left(i_T \otimes I_n\right)\mu + \left(I_T \otimes B^{-1}\right)\eta \tag{4.60}$$

with i_T a vector of ones of dimension T and I_T a T-by-T identity matrix. Now, let us define $J_T = i_T i_T^T$ as a $T \times T$ matrix of ones. After some straightforward algebra, the variance-covariance matrix of the composite error can be written as:

$$_{nT}\Omega_{nT,u} = \sigma_\mu^2 \left(J_T \otimes I_n\right) + \sigma_\eta^2 \left(I_T \otimes B_n^T B_n\right)^{-1} \tag{4.61}$$

which allows us to derive the likelihood function for estimation and hypothesis testing purposes (see section 5.3.5.1 below).

4.3.3.2 The KKP specification

An alternative specification for the disturbances was considered in Kapoor et al. (2007) and is known by the acronym *KKP*. In the quoted paper the authors assume that a spatial correlation structure could be applied to both the individual effects and the other error components. Although the two data-generating processes look similar, they do imply different spatial spillover mechanisms governed by a different structure of the implied variance-covariance matrix. In this case, the composite disturbance term, $u = \left(i_T \otimes I_n\right)\mu + \varepsilon$, is assumed to follows a first order spatial autoregressive process of the form:

$$u = \rho \left(I_T \otimes W\right) u + \eta \tag{4.62}$$

with all the symbols already defined. It follows that the variance-covariance matrix of the composite error u can now be expressed as:

$$_{nT}\Omega_{nT,u} = \left(I_T \otimes B^{-1}\right)\Omega_\varepsilon \left(I_T \otimes B^{-T}\right) \tag{4.63}$$

where $\Omega_\varepsilon = \sigma_\mu^2 J_T + \sigma_\eta^2 I_T \otimes I_n$ is the typical variance-covariance matrix of a one-way error component model. As Baltagi et al. (2013) observe, the economic meaning of the *SEM-RE* and the one implied by the *KKP* model are very different. In the first model only the time-varying components diffuse spatially, while in the second also the spatial spillovers display a permanent component.

4.3.4 Spatial panel models with fixed effects

As observed in section 4.3.2, if individual effects are uncorrelated with the independent variables, they can be considered as a component of the error term and treated in a generalized least squares fashion. Conversely, if we relax this hypothesis then this strategy leads to inconsistency. In this case we have to resort to the so-called *fixed effects*

methods: the individual effects will have to be estimated or, more frequently, just eliminated by first differencing the data, a procedure often called *time-demeaning* (Wooldridge, 2002). From a statistical sampling viewpoint, the random effects hypothesis is considered consistent with sampling individuals from a potentially infinite population, which has led Elhorst (2009) to dismiss its practical utility in spatial econometric contexts, where sampling typically takes place over a fixed set of countries or regions. Nevertheless, the modern approach to the issue (tracing back to Mundlak, 1978 and summarized by, among others, Wooldridge, 2002), centers on the statistical properties of the individual effects, which, despite traditional terminology, are always considered as random variates, the crucial distinction becoming whether one can assume them to be uncorrelated with the regressors or not. Hence, the distinction between fixed and random effects should better be regarded as an empirical issue. The Hausman (1978) test is the standard device for assessing the hypothesis of no correlation, and hence of using random effects methods. In a spatial setting, Lee and Yu (2012) give an extensive treatment of this topic.

4.3.5 Estimation

Spatial panel models, expressed in terms of either random or fixed effects, can be estimated by Maximum Likelihood (*ML*) or by Generalized Method of Moments (GMM) procedures. In general terms, *ML* is most efficient provided all distributional assumptions are met; it is also by far the most computationally intensive method, *GMM* being much easier to calculate. With respect to *ML, GMM* estimators also relax the normality assumption (an exception is discussed below) and, therefore, provide more robust estimates. All model specifications discussed in sections 4.3.3 and 4.3.4 can be estimated by either method. An exception is made for the specification reported in Equation (4.60), where the individual effects are independent across individuals. In fact, the only random effects specification for which *GMM* estimators are available is represented by the *KKP* model. In the following Section 4.3.5.1 we take a *ML* perspective, while we will devote Section 5.3.5.2 to the *GMM* alternative.

4.3.5.1 Maximum Likelihood estimation

The standard procedures developed for *ML* estimation of Spatial Lag and Spatial Error panels are due to Elhorst (2003). His approach is based on a combination of the time-demeaning technique familiar from

standard panel data (Wooldridge, 2002), and with Anselin's Maximum Likelihood framework. Data are transformed through an operational called *time-demeaning* which consists of subtracting the temporal mean from each observation in order to eliminate the individual spatial effects. After time-demeaning the standard Spatial Lag or Spatial Error estimators can be applied to the transformed data so that the first-order conditions simplify to those of *OLS*, with an additional spatial filter on y in the Spatial Lag case. We will illustrate separately the two cases of fixed effects and random effects here below.

Fixed effect model

From a computational point of view, according to the framework introduced by Elhorst (2003), fixed effects estimation of spatial panel models is accomplished as a pooled estimation on time-demeaned data. Let us start by considering the panel version of the **Spatial Lag model** described in section 3.5. In this case, following Elhorst (2003), in order to estimate the parameters we need to correct the likelihood of the pooled model by adding a spatial filtering on y using the filter $I_T \otimes A$, with $A = (I_n - \lambda W)$ and λ being, as usual, the spatial lag parameter. We also need to consider the explicit expression for the determinant term of the spatial filter matrix $|I_T \otimes A|$ which, in this case, is equal to $|A|$ to the power of T. The validity of Elhorst's procedures relies on a property that guarantees that $\Sigma(I_n \otimes A)y = (I_n \otimes A)\Sigma y$, for each matrix Σ, so that demeaning the spatially lagged data is equivalent to spatially lagging the demeaned data (see Mutl and Pfaffermayr, 2011, and Kapoor et al., 2007).

An efficient two-step iterative estimation procedure can be obtained as follows. First of all, let us consider the vector of the *demeaned* values for X and Y defined as:

$$\tilde{y}_{it} = y_{it} - \overline{y}_i \tag{4.64}$$

with $\overline{y}_i = \frac{\sum_{t=1}^{T} y_{it}}{T}$ the time mean at location i and similarly for X. This operation has the effect of removing the constant term from the regression. Secondly, consider the residuals derived from the demeaned model with a further spatial filter on y, defined as:

$$\tilde{\eta} = (I_T \otimes A)\tilde{y} - \tilde{X}\beta \tag{4.65}$$

Thirdly, we can derive the likelihood, concentrated with respect to β and σ_η^2:

$$l = const - \frac{nT}{2}\ln(\sigma_{\tilde{e}}^2) + T\ln|A| - \frac{nT}{2}\ln(\tilde{\eta}^T\tilde{\eta}) \tag{4.66}$$

that can be maximized with respect to λ. The value of λ thus obtained is used in a Generalized Least Squares step, imposing the following first order conditions:

$$\hat{\beta} = \left(\tilde{X}^T\tilde{X}\right)^{-1}\tilde{X}(I_T \otimes A)\tilde{y} \tag{4.67}$$

and

$$\sigma_\eta^2 = \frac{\tilde{\eta}^T\tilde{\eta}}{nT} \tag{4.68}$$

In this way, we obtain a new expression for the errors to be used in Equation (4.65). The procedure is then iterated until convergence.

Elhorst's procedure can be easily adapted to the **Spatial Error Model** specification. Again, an efficient two-step procedure can be based on concentrating the likelihood with respect to β and σ_η^2 as:

$$l = const - \frac{nT}{2}\ln(\sigma_\eta^2) + T\ln|B| - \frac{nT}{2}\ln(\tilde{\eta}^T\tilde{\eta}) \tag{4.69}$$

where the residuals from the demeaned model are now filtered through:

$$\tilde{\eta} = (I_T \otimes B)\tilde{y} - \tilde{X}\beta \tag{4.70}$$

with $B = (I - \rho W)$. Expression (4.69) can then be maximized with respect to ρ. Again as before, the value of ρ obtained from the maximization of Equation (4.69) is used in a Generalized Least Squares step, imposing the following first-order conditions:

$$\hat{\beta} = \left(\tilde{X}^T\tilde{X}\right)^{-1}\tilde{X}\tilde{y} \tag{4.71}$$

and

$$\sigma_{\eta}^{2} = \frac{\tilde{\eta}^{T}\tilde{\eta}}{nT} \tag{4.72}$$

obtaining new expressions for the errors to be used in Equation (4.70), and the procedure can be iterated until convergence.

Although not explicitly stated by the author, Elhorst's methodology is also easily extended to the SARAR specification (for an application see Millo and Pasini, 2010).

While still representing the standard in applied practice and in the available software, Elhorst's procedure has been criticized by Anselin et al. (2008) because the operation of time-demeaning alters the properties of the joint distribution of errors, introducing serial dependence. See Lee and Yu (2010b) for a discussion of the issue, and Millo and Piras (2012) for an evaluation of its practical significance through Monte Carlo simulation. To solve the problem, Lee and Yu (2010a) suggest a different orthonormal transformation of the data.

Random effects

The standard algorithms for the estimation of the Spatial Lag and Spatial Error versions of spatial panels with random effects are due to Elhorst (2003). His approach is based on the partial time-demeaning technique (Wooldridge, 2002) with Anselin's Maximum Likelihood framework: once data are quasi-time-demeaned in order to eliminate the random effects structure, then standard Spatial Lag or Spatial Error models estimators can be applied to the transformed data so that the first-order conditions simplify to those of OLS, plus a spatial filter on y in the Spatial Lag case (see Millo, 2013). Here we will instead consider untransformed data and we will specify the random effects of either type as a feature of the error covariance. The general likelihood function for the **Random Effect Spatial Lag** panel model combined with any error covariance structure Σ represents the panel data version of Equation (3.49):

$$\begin{aligned} l = const &- \frac{nT}{2}\ln(\sigma_{\eta}^{2}) + \frac{1}{2}\ln|\Sigma| + T\ln|A| \\ &- \frac{1}{2\sigma_{\eta}^{2}} \frac{\left[(I_T \otimes A)y - X\beta\right]^{T} \Sigma^{-1} \left[(I_T \otimes A)y - X\beta\right]}{nT} \end{aligned} \tag{4.73}$$

with Σ the composite error variance-covariance matrix. An iterative procedure, *à la* Oberhofer and Kmenta (1974), can be employed to

obtain the maximum likelihood estimates of the unknown parameters. Starting from an initial value for the spatial lag parameter λ and the error covariance parameters, we obtain estimates for β and σ_η^2 from the first-order conditions:

$$\hat{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}(I_T \otimes A)y \tag{4.74}$$

and

$$\sigma_\eta^2 = \frac{\left[(I_T \otimes A)y - X\beta\right]^T \Sigma^{-1}\left[(I_T \otimes A)y - X\beta\right]}{nT} \tag{4.75}$$

The likelihood reported in Equation (4.73) can be concentrated and maximized with respect to the parameters contained in A and Σ. The estimated values are then used to update the expression for Σ^{-1}. These steps are repeated until convergence. In other words, for any specific Σ, the estimation can be operationalized by a two-step iterative procedure that alternates between GLS (for β and σ_η^2) and concentrated likelihood (for the remaining parameters) until convergence. This general scheme can be applied to the random effects case. For example, the Spatial Lag model with random effects (SLM-RE) can be written as a combination of a spatial filtering on the dependent variable y and a random effects structure for the disturbances. More formally, we have:

$$(I_T \otimes A)y = X\beta + u \tag{4.76}$$

and

$$u = (i_T \otimes \mu) + \eta \tag{4.77}$$

where the variance-covariance matrix, say Σ_{SLM-RE}, is defined as $\Sigma_{SLM-RE} = \varphi(J_T \otimes I_n) + I_{nT}$ and the extra parameter φ is defined as:

$$\varphi = \frac{\sigma_\mu^2}{\sigma_\eta^2} \tag{4.78}$$

and represents the ratio between the variance of the individual effect and the variance of the idiosyncratic error.

As mentioned in section 4.3.3 above, the **Random Effect Spatial Error** model gives rise to two possible specifications, depending on the

interaction between the spatial autoregressive effect and the individual error components.

In the first specification (*SEM-RE*), only the idiosyncratic error is spatially correlated and the model can be expressed through the following three equations:

$$y = X\beta + u \tag{4.79}$$

$$u = (i_T \otimes \mu) + \varepsilon \tag{4.80}$$

and

$$\varepsilon = \rho(i_T \otimes \mu)\varepsilon + \eta \tag{4.81}$$

with the scaled errors' covariance expressed by:

$$\Sigma_{SEM-RE} = J_T \otimes \left(T\varphi I_n + (B^T B)^{-1}\right) + E_T \otimes \left(B^T B\right)^{-1} \tag{4.82}$$

with $\overline{J}_T = \dfrac{J_T}{T}$ and $E_T = I_T - J_T$.

In the second specification (*KKP*) the same spatial model applies to both the individual and the idiosyncratic error component, and can be expressed through the set of equations:

$$y = X\beta + u \tag{4.83}$$

$$u = (i_T \otimes \mu) + \varepsilon \tag{4.84}$$

and

$$u = \rho(I_T \otimes W)u + \varepsilon \tag{4.85}$$

where the scaled errors' covariance to be substituted into the likelihood is:

$$\Sigma_{KKP} = \left(\varphi J_T + I_T\right) \otimes \left(B^T B\right)^{-1} \tag{4.86}$$

4.3.5.2 GMM estimation

GMM estimators of spatial panel models are based on the strategy of alternating spatial Cochrane–Orcutt transformations to filter out the spatial dependence with the standard *GLS* (for *RE*) or time-demeaning transformations familiar to the panel data literature. The spatial

Cochrane–Orcutt transformations are in turn based on consistent estimates of the spatial parameters. As in Kapoor et al. (2007) and Millo and Piras (2012), for the sake of simplicity we will only consider here a Spatial Error model, while referring the interested reader to Mutl and Pfaffermayr (2011) and Piras (2011) for the extension to the full model.

Random effects

Kapoor et al. (2007) extended the generalized moments estimator suggested in Kelejian and Prucha (1999) for the spatial parameter of a cross-sectional model (see section 3.4.3) to the panel case. They outline the estimation procedure for the spatial autoregressive parameter of the error process (ρ) and the two variance components of the disturbance process (defined by the two terms $\sigma_1^2 = \sigma_\mu^2 + \sigma_\eta^2$ and σ_η^2). Three alternative *GMM* estimators are defined on the basis of the following moments conditions:

$$E\begin{bmatrix} \frac{1}{n(T-1)}\varepsilon^T Q_0 \varepsilon \\ \frac{1}{n(T-1)}\bar{\varepsilon}^T Q_0 \bar{\varepsilon} \\ \frac{1}{n(T-1)}\bar{\varepsilon}^T Q_0 \varepsilon \\ \frac{1}{n(T-1)}\varepsilon^T \varepsilon \\ \frac{1}{n(T-1)}\bar{\varepsilon}^T \bar{\varepsilon} \\ \frac{1}{n(T-1)}\bar{\varepsilon}^T \varepsilon \end{bmatrix} = \begin{bmatrix} \sigma_\eta^2 \\ \sigma_\eta^2 \frac{1}{n} tr(W^T W) \\ 0 \\ \sigma_1^2 \\ \sigma_1^2 \frac{1}{n} tr(W^T W) \\ 0 \end{bmatrix} \tag{4.87}$$

where, $Q_0 = \frac{I_T - J_T}{T}$ and I_N is the (time-)demeaning matrix, so that $Q_0 y = \tilde{y}$ (see previous section). Furthermore, similar to Section 3.4.3, we define $\varepsilon = u - \rho \bar{u}$, $\bar{\varepsilon} = \bar{u} - \rho \bar{\bar{u}}$, $\bar{u} = (I_T \otimes W_n)u$ and $\bar{\bar{u}} = (I_T \otimes W_n)\bar{u}$.

The estimator implied by Equation (4.87) is based on the fact that in a random effects model without a spatial lag of the dependent variable, the OLS estimator of β is consistent. The OLS residuals can therefore be employed in the GMM procedure.

The first set of GMM estimators is based only on the first three equations and assigns equal weights to each of them. The second set of GMM estimators uses all of the moments conditions and an optimal

weighting scheme: the inverse of the variance-covariance matrix of the sample moments at the true parameter values under the assumption of normality (derived by Kapoor et al., 2007, who also discuss its appropriateness in the case of non-normality). Finally, the third set of *GMM* estimators uses all moments conditions, but a simplified weighting scheme, and can prove useful in cases of computational difficulties in calculating the elements of the asymptotic variance-covariance matrix of the sample moments.

Using one of the above estimators, we can obtain estimates of the spatial parameter and of the variance components. The first estimator can be used to perform a spatial Cochrane–Orcutt type transformation, while, in order to eliminate individual effects, the data are further transformed by pre-multiplying them by the term $I_{nT} - \left(\frac{1-\sigma_v}{\sigma_1}\right)Q_0$ as in the standard panel data literature. The feasible *GLS* estimator reduces in this case to an *OLS*, calculated on the "doubly" transformed model. Finally, small sample inference can be based on the following expression for the parameter's variance-covariance matrix:

$$\Gamma = (X^{*T}\Omega_{\varepsilon}^{-1}X^{*})^{-1} \tag{4.88}$$

where the variables X^* are the result of a spatial Cochrane–Orcutt type transformation of the original model (see Chapter 3), and both X^* and $\Omega_{\varepsilon}^{-1}$ depend on the estimated values of ρ, σ_{μ}^2 and σ_1^2.

Fixed effects

A modification of the above procedure has been recently suggested for the case when the random effects assumption of uncorrelation between the individual effects and the regressors cannot reasonably be maintained. Mutl and Pfaffermayr (2011) note that, under the fixed effects assumption, *OLS* estimation of the regression equation is no longer consistent. They suggest replacing *OLS* residuals with spatial two-stage least squares *within residuals*. In the Spatial Error case, a simple within-estimator will produce consistent estimates of the model parameters. In the quoted paper, Mutl and Pfaffermayr (2011) reformulate the first three moments conditions of Kapoor et al. (2007) in terms of the within residuals and then estimate the spatial parameter ρ using the *GMM* procedure described in Kapoor et al. (2007) based only on these first three moments conditions. The model parameters are then obtained by *OLS* after a further spatial Cochrane–Orcutt-type transformation of the within transformed variables.

4.3.6 Further modeling frameworks in spatial panel data modeling

The literature on spatial panel data models has been growing rapidly in recent years. In a recent survey on the spatial econometrics literature in the period 2007–11 (Arbia, 2012), this issue is described as the hot topic with the largest numbers of papers published both under the theoretical and the applied point of view. Consequently, the basic framework presented here in a very concise way has been substantially enlarged and improved. Excellent survey papers in this area are provided by the introduction to the special issue of Journal of Applied Econometrics devoted to spatial panels (see Baltagi and Pesaran, 2007), by a review article by Lee and Yu (2010b) and by a long position paper by the same authors (Lee and Yu, 2011). The reader is referred to these surveys for more details on the wide range of topics covered in the recent literature. An important generalization of the topics briefly summarized here concerns the fact that these models are static (even if based on time series data), in the sense that they do not take into account any dynamic effects. In this respect, the literature treats both *stable* dynamic panel data models and *unstable* models to include the treatment of topics such as spatial unit roots, spatial cointegration and explosive roots (Lee and Yu, 2011).

Example 4.5 Munnell's model of public capital productivity (continued)

Although the original focus of Munnell's (1990) paper was on the significance of public capital in the social production area, the researcher might be interested in investigating the spatial properties of the error term in order to detect whether it is governed by a Spatial Error model of either the SEM-RE or the KKP type. In this example we focus on the first typology, by employing a row-standardized neighborhood matrix, and estimating it by Maximum Likelihood on the same data introduced in Example 4.4. The results of the estimation procedure are shown here below:

SEM-RE Model

Parameter		***Standard Error***	***t-test***	***p-value***
Intercept	*1.6220798*	*0.2127472*	*7.6244*	*2.451e–14****
log(pcap)	*0.0382457*	*0.0512507*	*0.7462*	*0.45552*
log(pc)	*0.3970841*	*0.0433498*	*9.1600*	*< 2.2e–16****
log(emp)	*0.6271537*	*0.0395013*	*15.8768*	*< 2.2e–16****
unemp	*–0.0066850*	*0.0026529*	*–2.5199*	*0.01174**
φ variance component	*25.946903*	*6.470930*	*4.0098*	*6.078e–05****
ρ	*0.595470*	*0.065178*	*9.1361*	*< 2.2e–16****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

The estimated variance of the individual effect is much bigger than that of the idiosyncratic error, the latter showing substantial spatial correlation: the evidence in favor of a spatial process in the errors is thus strong. The estimates of the parameters β are not strikingly dissimilar from those obtained with the a-spatial random effects model reported in Example 4.4, but, given the above evidence, we can trust these latter estimates to be more precise as taking a statistically significant component into consideration which was neglected in the a-spatial specification of the model.

If one is interested in the effect of one state's gross social product on neighboring states, then a Random Effect Spatial Lag specification seems more appropriate. The results of the estimation are shown in the following table:

SLM-RE Model

Parameter		***Standard Error***	***t-test***	***p-value***
Intercept	*1.3671510*	*0.1970401*	*6.9384*	*3.964e–12****
log(pcap)	*0.1093629*	*0.0526096*	*2.0788*	*0.0376393**
log(pc)	*0.6066125*	*0.0376993*	*16.0908*	*< 2.2e–16****
log(emp)	*–0.0097658*	*0.0026518*	*–3.6827*	*0.0002307****
unemp	*–0.0097658*	*0.0026518*	*–3.6827*	*0.0002307****
φ variance component	*18.7043*	*4.6822*	*3.9948*	*6.475e–05****
λ	*0.038789*	*0.021832*	*1.7767*	*0.07562*

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

As it turns out, the estimated spatial autoregressive coefficient is low and not significant at the 5% confidence level: the evidence in favor of a SLM-RE model is thus rather weak.

Example 4.6 A spatial panel, fixed effects analysis of convergence in Italian regions

In this example we consider again the regional convergence model à la Barro and Sala-i-Martin (1992) already presented in Example 1.1 revised under a panel context for the 20 Italian regions (see Figure 2.3 for the geographical map). The panel version of the growth equation can be expressed as:

$$\log(y_{i,t+k}) - \log(y_{it}) = \alpha_i + \beta \log(y_{it}) + \varepsilon_{it}$$

where, as already explained in Example 1.1, y_{it} represents the per-capita GDP in year t and region i. Therefore the growth rate over a period of k years for each region i is a function of the initial level of income of the region and of an individual intercept. Looking at the issue from a panel data perspective

enriches the possible interpretations with respect to the purely cross-sectional analysis, in that it allows control for time-persistent, individual heterogeneity in the characteristics of regions. Furthermore, by treating individual effects as parameters to be estimated instead of an error component, a fixed effects specification allows the said heterogeneity to be correlated with a model's regressors. We consider the per-capita GDP data observed each 5 years (k=5) for all 20 Italian regions in the period from 1960 to 1995 (data are not shown for brevity) and we regress the differences between the logarithm of real GDP (say, GDPV) and the log of five-year lagged GDP (say, l5GDPV) on l5GDPV. Let us start as usual, by estimating an a-spatial fixed effect model. The results are shown in the following table:

Parameter		***Standard Error***	***t-test***	***p-value***
log(l5GDPV)	*–0.199311*	*0.014445*	*–13.798*	*< 2.2e–16****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

The model's estimation yields a negative and significant coefficient for β thus confirming convergence.

In order to estimate the various spatial model specifications, we consider the binary neighborhood matrix of Italian regions described in section 2.3.2. We first consider the Spatial Lag Fixed Effect model, estimated through maximum likelihood. We obtain the following results:

SLM-FE

Parameter		***Standard Error***	***t-test***	***p-value***
λ	*0.680944*	*0.053398*	*12.7523*	*< 2.2e–16****
log(l5GDPV)	*–0.068823*	*0.014291*	*–4.8158*	*1.466e–06****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

The spatial lag parameter λ turns out to be very high and statistically significant. Thus growth in neighbouring regions has a strong positive effect on local growth. The estimate of the convergence coefficient β is still statistically significant and has the expected negative sign, but is now much lower in absolute value (although notice that it cannot be directly compared to models without spatial lag). This highlights a lower "speed of convergence" in the adjustment process (see Example 1.1).

Next, we consider a model with a Spatial Error component, assuming that shocks to growth from neighboring regions propagate to neighbors, affecting

their outcome. The results of the Spatial Error, Fixed Effect model estimation are the following:

SEM-FE

Parameter		***Standard Error***	***t-test***	***p-value***
log(l5GDPV)	*–0.271102*	*0.031033*	*–8.736*	*< 2.2e–16****
ρ	*0.780552*	*0.041063*	*19.008*	*< 2.2e–16****

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

The estimate of the spatial parameter of the error is very high and significant. The convergence coefficient is larger than the one obtained with the non-spatial model and it is also statistically significant.

4.4 Non-stationary spatial econometric models

4.4.1 Generalities

All the models that we have considered so far describe the relationships between variables as a mechanism that is stationary over the various geographical units. This approach presents the advantage of synthesis: one single parameter, the regression coefficient, summarizes all the complex links between the dependent variable y and the corresponding independent variable of the model. However, this simplification is sometimes too strong. In many empirical situations, it is unreasonable to believe that a relationship between two variables is constant in the whole geographical area and it is more sensible to conjecture that it varies in the space according to some regular pattern. For instance, if we are dealing with a large sample size observed on a large geographical space (for example, a large country or a continent), it is more sensible to build up a flexible regression mechanism where the relationship between variables is allowed to change smoothly from one place to another. Consider the case reported in Example 3.2, where we studied the relationship between the price of used cars and the taxes on the basis of the data observed in 48 US states. In principle one can imagine that such a relationship should be different in each state or, in other words, that it is spatially *non-stationary*. In many respects the heterogeneity of relationships over space may be regarded more as the rule rather than the exception, first of all because there are geographical variations in people's attitudes and preferences, and secondly because a model is often the result of a misspecification due to the difficulty in

accounting adequately for the observed spatial variations relying only on a relatively small number of explicative variables. Many relevant variables that are at the basis of spatial differentiation can be left out of the model because of the impossibility of measuring them properly or they may be represented in an incorrect functional form.

In this chapter we will introduce an alternative modeling framework that allows the consideration of non-stationary spatial relationships. We will concentrate, in particular, on the technique termed *Geographically Weighted Regression* (GWR). There are two important historical antecedents to this modeling strategy. The first one refers to the concept of *scan statistics* and the second to the notion of *Locally Weighted Regressions*. Scan statistics have a long tradition in the statistical literature (see Glaz et al., 2001, for a review) and are very popular in many applied fields such as, for example, in public health (where researchers look for common causal factors to explain unusual clusters of pathologies), molecular biology (where clusters in DNA suggest possible causes for the origin of the replication of viruses), telecommunications, quality control, and many others. A particular case, relevant for spatial econometric applications, is that of *two-dimensional scan statistics*. In its essence the idea is very simple: rather than considering all available observations together, we select subsamples that are close geographically and we perform some statistical computations within these subsamples. The first examples of two dimensional scan statistics can already be found in chapter 3 of Sir Ronald Fisher's book (Fisher, 1959) who showed applications in the field of astronomy. Arbia (1990) proposes the use of scan statistics (referring to them with the term *moving windows*) to study spatial non-stationarity of the first and second-order moments of a random field. Hoh and Ott (2012) provide an interesting example of their use to genome screening while Paez et al. (2008) show their use in the study of hedonic price estimation. *Locally Weighted Regression* (*LWR*) is a non-parametric methodology (introduced by Cleveland and Devlin, 1988) that can be seen as a scan statistic technique to perform a regression around a point of interest using only a limited number of training data that are somewhat *local* to that point. In Locally Weighted Regression, points are weighted by proximity to the current point using a kernel (see section 4.1.3); a regression is then computed using the weighted points (McMillen, 1996). The output of a Locally Weighted Regression is thus a set of separate regression estimates for each observation but, since the technique makes use of kernel filters, it produces a smooth variation so that nearby observations will tend to display similar

coefficients. Finally, *Geographically Weighted Regression* (*GWR*) is a particular case of the *LWR* (Brundson et al., 1996; Fotheringham et al., 1998, 2002, 2007; McMillen and McDonald, 1997, 2004) which makes use of the geographical space as a selection criterion. *GWR* provides a set of local estimates that can be mapped to produce a parameter surface which varies across the study region. These maps play a paramount role for exploring spatial heterogeneity and for understanding spatial relationships.

4.4.2 Geographically weighted regression

For each location *i,* (*i*=1,2,…,*n*) consider the following regression model (see Wheeler and Paez, 2010):

$$y_i = X_i\beta_i + \varepsilon_i; \quad i = 1,....n \tag{4.89}$$

where y_i is the *i-th* observation of the dependent variable y at location i, X_i represents the *i-th* row of the matrix of observations ${}_nX_k = \begin{bmatrix} 1 & X_{11} & X_{1k-1} \\ \dots & & \dots \\ 1 & & \\ 1 & X_{n1} & X_{nk-1} \end{bmatrix}$, β_i is vector of (geographically varying) unknown parameters to be estimated and ε_i the *i-th* observation of a stochastic disturbance such that $\varepsilon_i \big| X \approx i.i.d.N(0,\sigma^2_{\varepsilon\,n} I_n)$. The estimation strategy consists of estimating n distinct models using the following intuitive *GLS* estimator:

$$\hat{\beta}_i = \left[X^T G_i X\right]^{-1} X^T G_i y \tag{4.90}$$

where y is now a n-by-*1* vector, $\hat{\beta}_i$ is the *OLS* estimator of β_i and $G_i = \begin{bmatrix} g_{i1} & 0 & 0 & \dots & 0 \\ 0 & g_{i2} & 0 & & 0 \\ & & & & \dots \\ 0 & 0 & \dots & 0 & g_{in} \end{bmatrix}$ is an n-by-n diagonal matrix specifying the set of weights for each location. The role of matrix G is similar to that of the familiar weight matrix W used extensively in the previous chapters. In fact, it selects, among the available observations, those that are relevant to forecast the behavior of y at location i by assigning them a specific

weight inversely related to distance. There is, however, an important difference between the two matrices in that the number of observations to be included in each local regression (that is the number of non-zero entries of matrix G_i) has to be large enough to preserve degrees of freedom to the regression and to allow reliable estimation of the parameters. The estimated coefficients (4.90) can be interpreted, as usual, as the (local) marginal effects on the dependent variable of a unitary variation in the independent variable. *GWR* thus semi-parametrically models the essential idea (which is at the basis of all spatial econometric models) that in order to forecast the behavior of one variable in one specific location, the best predictors are the nearby observations. Furthermore, limiting the estimation procedure to only neighboring observations will most likely reduce, if not completely eliminate, most of the negative effects of spatial correlation and heteroscedasticity on parameter estimates.

The estimator reported in Equation (4.90), under some conditions, coincides with the one obtained by the maximization of a weighted pseudo-likelihood derived separately for each observation (see McMillen and McDonald, 2004). The weighting scheme incorporated in matrix G is an essential feature of the method and depends on two crucial elements. The first concerns the choice of the kernel considered and the second the number of observations, close to location *i,* that are relevant to estimate the (spatially changing) parameter at location *i*. This second element is incorporated in a kernel's parameter termed *bandwidth*. The choice of the appropriate bandwidth is an essential element of the procedure. Small bandwidths are preferable in order to identify local patterns, but may produce unreliable estimates being based on a small sample size. On the other hand, large bandwidths, while preserving the reliability of the estimators, will tend to cancel local heterogeneity. When the bandwidth is so large that it includes all observations (equally weighted), clearly *GWR* coincides with the standard *OLS* regression. The simplest choice of the weights is to consider the matrix G to be just the familiar (unstandardized) W matrix defined in accordance to some maximum distance neighboring criterion (see section 2.1) without any specific weighting. More formally:

$$g_{ij} = \begin{cases} 1 & \text{if } d_{ij} < d^* \\ 0 & \end{cases} \tag{4.91}$$

with d_{ij} the distance between observation i and observation j, and d^* the threshold.

While in principle any of the kernels presented in section 4.1.3 could be used, in practice the most common kernels employed in *GWR* are the following:

1. *Gaussian*. This kernel has already been considered in section 4.1.3 and in this new context can be expressed as:

$$g_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{\sigma}\right)^2\right] \tag{4.92}$$

   with σ a parameter, called *bandwidth,* through which we can control the range of observations included in each of the subsamples.
2. *Bi-square* (generalization of bi-weight, see section 4.1.3)

$$g_{ij} = \left(1 - \frac{d_{ij}^2}{\sigma^2}\right)^2 \tag{4.93}$$

3. *Tri-cube*

$$g_{ij} = \left(1 - \frac{d_{ij}^3}{\sigma^3}\right)^3 \tag{4.94}$$

The bandwidth parameter σ can either be chosen *a priori* or estimated from the data, a process often called *calibration*. In this respect, the most popular calibration strategy is the *cross-validation method*, an iterative search process that identifies, through repeated trials, the bandwidth that minimizes the root mean square of the prediction error. (For a different approach, see, for example, Paez et al., 2002. See Wheeler and Paez, 2010 for a review). Once the kernel function is chosen and the model calibrated, Equation (4.90) is fully specified and it can be used to estimate the model parameters. As for hypothesis testing, let us consider, as usual, the test statistics obtained by taking the difference between the value of β_i under the null and under the alternative hypotheses scaled by its standard deviation (Equation 1.21), that is:

$$test = \frac{\hat{\beta}_i}{\sqrt{Var(\hat{\beta}_i)}} \tag{4.95}$$

where the scaling factor in the denominator is obtained using the *GLS* expression (Equation 1.48) which, in this case, leads to:

$$Var(\hat{\beta}_i) = \sigma_\varepsilon^2 A_i A_i^T \tag{4.96}$$

having defined $A_i = \left(X^T G_i X\right)^{-1} X^T G_i$. The test statistics (4.95) can be used to test the validity of the regression hypothesis in each individual observation. In this sense we can observe that the effect of one independent variable on the dependent variable y is significant only in some specific zones of the study area and it is, conversely, irrelevant in others, thus suggesting interesting geographical patterns.

Once the model is estimated and the significance of the independent variables tested in each of the n estimated models, we can proceed, as usual, to test the validity of the hypotheses underlying the model. In particular, in the present context, it is of interest to test if the model's residuals display some geographical regularity (for example, in the form of spatial autocorrelation) that was not adequately captured by the geographically weighted procedure. The theory related to the test of spatial autocorrelation among *GWR* residuals has been developed by Leung et al. (2000). The proposed approach starts from the definition of the pseudo-residual of the global model, defined as the residuals calculated as the difference between the observed value of y_i in each location i and the estimated value on the basis of the (spatially non-stationary) parameters. These residuals are not defined in the conventional way because in each location they are computed starting from a different regression model. Once the pseudo-residuals are calculated they are used to detect possible regularities using a revised version of the familiar Moran's I coefficient. An alternative to Leung et al. (2000) procedure was suggested by Paez et al. (2002).

4.4.3 Further developments

The basic *GWR* methodology presented in this section represents only the starting point for further and more complex modeling strategies. The literature in this field is vast, rapidly-growing and still in the early stages of development in many respects (for a review, see Fotheringham et al., 2002; McMillen and McDonald, 2004; Pace and LeSage, 2004; Wheeler and Paez, 2010). An obvious extension of the basic model presented in section 4.4.2 concerns the use of spatial econometric models belonging to the *SARAR* family in place of the simple *OLS* local regression (see Brunsdon et al., 1998; Paez et al., 2002; Pace and LeSage, 2004; and Mur et al., 2008). However, this extension encounters severe computational challenges when operating with large sample size. In fact, not only does the same model have to be estimated n times (encountering the computational issues raised several times when presenting the various models in chapter 3), but also the iterative search of an optimal bandwidth during the cross-validation phase involves the calculation of the determinant of an $(n–1)$-by-$(n–1)$ matrix repeatedly at each step of the procedure. A second interesting extension is represented by a Bayesian

approach to *GWR* (introduced by LeSage, 2004) in order to solve some of the difficulties emerging in the standard approach in the presence of outliers and strong heteroscedastic phenomena. In its essence the procedure improves the method by imposing an a priori distribution (depending on a set of hyperparameters) to the regression coefficients. A further extension concerns the application of the general *GWR* framework illustrated in this section to deal with discrete choice models (see section 4.2). For instance, McMillen and McDonald (2004) derive a Maximum Likelihood approach to the estimation of Geographically Weighted Probit models showing that through this approach much of the hetorescedasticity and autocorrelation, endemic to spatial discrete choice models (see section 4.2.3), can be eliminated thus reducing the negative effects on consistency and efficiency of the estimators.

Example 4.7 The determinant of educational achievement in Georgia

As an example of Geographical Weighted Regressions let us consider a dataset (presented by Fotheringham et al., 2002) referring to the 159 counties of Georgia. The dataset can be downloaded from the R library spgwr with the command `data(Georgia)` *(see section 4.5.4). The dataset contains the following variables: Proportion of residents with a Bachelor's degree or higher (Bach), Proportion of people living in rural neighborhood (Rural), Proportion of elderly residents (Eld), Proportion of residents who are foreign born (FB), Proportion of residents who are living below the poverty line (Pov) and Proportion of residents who are ethnic black (Black). The dataset also contains information about the total population of each county and a series of other information about the geography of the area, such as latitude and longitude of the centroids of each county. We consider a model that aims to explain the percentage of Bachelor or higher degrees in one county as a function of the other variables described above (Rural, Eld, FB, Pov and Black). The results of the standard a-spatial OLS model are shown in the following table:*

Parameter	*Estimated Value*	*Standard Error*	*t-test*	*p-value*
Intercept	*17.24373*	*1.75329*	*9.835*	*< 2e–16****
Rural	*–0.07032*	*0.01358*	*–5.179*	*6.93e–07****
Eld	*0.01145*	*0.12953*	*0.088*	*0.929693*
FB	*1.85247*	*0.30683*	*6.037*	*1.14e–08****
Pov	*0.25524*	*0.07248*	*–3.522*	*0.000566****
Black	*0.04911*	*0.02648*	*1.854*	*0.065602*

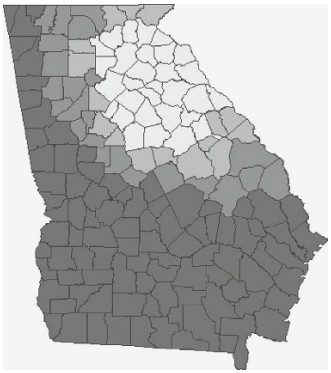
*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

Multiple R-squared: 0.5884, Adjusted R-squared: 0.575
F-statistic: 43.75 on 5 and 153 DF, p-value: < 2.2e–16

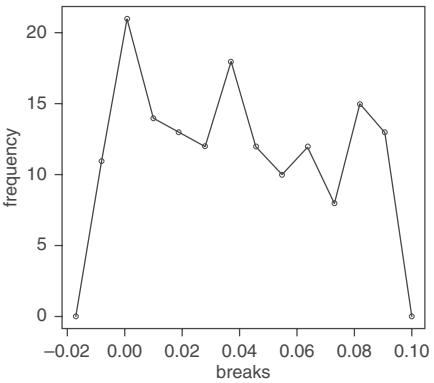
A Geographically weighted regression model can then be estimated using a Gauss kernel and having identified the optimal bandwidth using the cross-validation iterative method. Some summary results are shown in the following table for each of the estimated parameters:

Parameter	***Minimum***	***First Quartile***	***Median***	***Third Quartile***	***Maximum***	***Range***	***Global (OLS)***
Intercept	*14.170000*	*15.350000*	*17.050000*	*18.200000*	*18.860000*	*4.69*	*17.2437*
Rural	*–0.081350*	*–0.073480*	*–0.064850*	*–0.055110*	*–0.051080*	*0.03027*	*–0.0703*
Eld	*–0.191200*	*–0.094630*	*–0.065330*	*–0.032360*	*0.012500*	*0.2037*	*0.0114*
FB	*0.854300*	*1.282000*	*2.031000*	*2.796000*	*3.138000*	*2.2837*	*1.8525*
Pov	*–0.304800*	*–0.258100*	*–0.196100*	*–0.115100*	*–0.034210*	*0.27059*	*–0.2552*
Black	*–0.016900*	*0.006347*	*0.031610*	*0.060620*	*0.087210*	*0.07031*	*0.0491*

In the last column of the previous table we displayed again the results obtained with OLS for the sake of comparison. The results show a high variability of the local coefficients around the global value. Notice that in some cases the parameters can assume, in different locations, the opposite sign (variables "Eld" and "Black"). So in some zones of the study area there is a positive relationship between, e.g. the variable "Bach" and "Black" while in other zones this relationship is positive. As an example, the map shown in the following figure (a) displays the geographical variability of the regression coefficients related to the variable "Black" (darker shades refer to positive coefficients, lighter shades to negative coefficients). The lack of homogeneity of the coefficients is evident, with a clear pattern decreasing from South-West to North-East. Thus the variable "Percentage of Black people" has a positive impact on the variable "Percentage of Bachelor" in the South and in the Western counties, but, in contrast, a negative impact in the North-Eastern counties. Notice the smoothness of the map which is an intrinsic feature of the method originating from to the use of a kernel which is inversely proportional to distance. Figure (b) shows the frequency distribution of the estimated GWR coefficients for the variable "Percentage of Black people" which also highlights the presence of both positive and negative coefficients.



(*a*)



(*b*)

4.5 R Codes

4.5.1 Estimating heteroscedastic linear spatial models

The package `sphet` is designed to estimate and test spatial models with heteroscedastic innovations. It complements and partially overlaps with the econometric features available in the library `spdep`. Remember to type the command `install.packages("sphet")` for the first time and then before each session type `library(sphet)`.

Suppose, to start with, that we are considering a **heteroscedastic SARAR model** expressed as $y = \beta_0 + \beta_1 X + \beta_2 Z + \lambda W y + u;\ \ u = \rho W u + \varepsilon$ with $E(\varepsilon_i) = \sigma_i^2$ and that we want to estimate it using the modified Generalized Spatial Two-Stage Least Squares procedure illustrated in Section 4.1.2. Let us further assume that the observations of the variables y, X and Z are stored in the active `R` session and that a weight matrix W was generated with one of the procedures described in the previous chapters. To estimate the model, type the command:

```
model1<- gstslshet(formula=y~X+Z, date=filename, listw=W)
```

The output of the procedure includes, by default, the Wald test that both spatial coefficients ρ and λ are zero.

If we wish to estimate the **spatial HAC** model (Section 4.1.3), in order to evaluate the variance-covariance matrix contained in Equation (4.8), we need two extra objects with respect to the usual parameters: i) the matrix of the pairwise distances between all the spatial units (the term d_{ij})

and ii) the typology of the kernel function. Let us call D the matrix of distances and let us consider, for example, that we choose the Epanechnikov Kernel function (other possible alternatives supported by the package are the Triangular (Bartlett), the Bisquare (or quadratic, or bi-weight), the Parzen, the TH (Tukey–Hanning) and the QS (Quadratic Spectral) kernels (See Section 4.1.3). With these specifications the command is:

```
model2<-stslshac(formula=y~X+Z,listw=w,distance=D,type=
"Epanechnikov")
```

In this expression, the bandwidth is variable. Conversely, if we wish to fix the bandwidth at a value, e.g. B, we have to add the option `bandwidth="B"`. Furthermore, the command uses by default both WX and W^2X as instruments. If we want to use only WX as a single instrument we have to add the option `W2X=FALSE`.

4.5.2 Estimating spatial probit/logit models

First of all let us introduce the `R` commands to estimate standard, a-spatial Probit and Logit models based on the latent model $y^{\bullet} = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon$. The command is the following:

```
>model0<-glm(y~X+Z, family=binomial(link="probit"))
```

and similarly for the Logit model.

The `R` commands related to the various procedures to estimate the spatial version of the Probit and Logit models are contained in the library `McSpatial`. So remember to type the command `install.packages("McSpatial")` and then before each session type `library(McSpatial)`. In all cases we assume that we want to estimate a Probit model where the latent variable is expressed as: $y^{\bullet} = \beta_0 + \beta_1 X + \beta_2 Z + \lambda W y^{\bullet} + \varepsilon$ and we also assume that the observations of the variables y, X and Z are stored in the active `R` session. As for the weight matrix W, it has to be generated with a different procedure with respect to those described in the previous chapters. In this case, in fact, once the list of neighbors (say `contnb`) is created through the command `poly2nb` or `dnearneigh`, (see section 2.3.3), instead of the command `nb2listw`, we have to use the command:

```
W<-nb2mat(contnb)
```

which transforms a list of neighbors (`nb`) to (`2`) a matrix (`mat`). Let us now consider the various estimators of a spatial Probit model. To start

with, if we want to estimate the parameters of a **spatial Probit** model using the **ML** technique (section 4.2.3.2), type the command:

```
> model1<-spprobitml(y~X+Z,wmat=W,stdprobit=F)
```

If we want estimate the same model using a Generalized Method of Moments approach (**GMM**, see section 4.2.3.3), we need to initialize the parameter ρ before starting the iterative search of a solution. So, before starting the procedure we have to type, e.g.:

```
>rho=rho0
```

rho0 being any value such that |rho0 |<1. The robustness of the results with respect to different initial values should be tested in any practical circumstances. Having done so, now type:

```
> model2 <-gmmprobit(y~X+Z,wmat=W,startrho=rho)
```

Finally, if we want estimate the same model using the Linearized Generalized Method of Moments approach (**LGMM**, see section 4.2.3.4), type instead:

```
> model3<-spprobit(y~X+Z,wmat=W)
```

4.5.3 Estimating spatial panel models

The procedures for estimating spatial panels are contained in the `R` package called `splm` (Millo and Piras 2012), which, as usual, must be installed for all through the command `install.packages("splm")` and can then be loaded at the beginning of each session by typing `library(splm)`. With this command we also automatically load the `plm` package (Croissant and Millo 2008) for the estimation of non-spatial panel models.

Panel datasets in the `R` packages `plm` (Croissant and Millo, 2008) and splm have very few formal requirements. The dataset can be stored in a regular dataset object, provided it contains the necessary pair of indices any panel must have (related to individuals and time) so as to be unambiguously identified. Given these, no particular ordering of the data is required. The software assumes that the individual index is the first column of the dataset and the time index the second. If the data are arranged in a different way, then one must specify the 'index' argument as a pair of strings giving the variable names inside the call to the estimating function.

The two main functions in package `splm` are `spml` (estimation of Spatial Panel models by Maximum Likelihood) and `spgm` (estimation of

Spatial Panel models by Generalized Method of Moments). Munnell's data used in Examples 4.4 and 4.6 are available in the package as a built-in dataset, together with the weights matrix for the 48 US states already considered in Example 3.2 (excluding Alaska and Hawaii and the District of Columbia. See the map in Example 3.2). We can load the dataset through the command:

```
> data(Produc)
```

First of all, to estimate the **OLS model**, pooling all data (spatial and temporal) together, we can use the `plm` function included in the `plm` package adding the option `model="pooling"`

```
> model0<-plm(y ~ X+Z, model="pooling")
```

assuming that the data are in the active dataset.

Let us now present the procedures for the **Random Effect Model**, starting from the case of ***no spatial component***. In this case the ML estimation can be performed using the command `spml` (acronym for Spatial Panel Maximum Likelihood) with the following specifications:

```
> model1 <- spml(y~X+Z, listw=W, model="random", spatial.
error="n", lag=FALSE)
```

where *W* is the usual weight matrix. To add a ***Spatial Lag*** to the specification, one must set the lag argument to TRUE:

```
> model2 <- spml(y~X+Z, listw=W, model="random", spatial.
error="n", lag=TRUE)
```

The ***SEM-RE*** model (according to the Spatial Error specification expressed in (4.79) and (4.80) can be estimated specifying no spatial lag and a spatial error of type "b":

```
> model3 <- spml(y~X+Z, listw=W, model="random", spatial.
error="b", lag=FALSE)
```

while the ***KKP*** specification reported in Equations in (4.83) and (4.84) can be estimated by maximum likelihood setting `spatial.error="kkp"`as an option:

```
> model4 <- spml(y~X+Z, listw=W, model="random", spatial.
error="kkp", lag=FALSE)
```

or, alternatively, using the generalized method of moments using the command:

```
> model5 <- spgm(y~X+Z, listw=W, model="random", spatial.
error="kkp", lag=FALSE)
```

Let us now move on to consider the **Fixed Effect Model** specification.

First of all, the ***a-spatial fixed effects model***, can be simply estimated typing:

```
> model6 <- plm(y~X+Z)
```

If we wish to add a ***Spatial Lag*** to the specification, we can estimate the fixed effect model through the command:

```
> model7 <- spml(y~X+Z, listw=W, spatial.error="none",
lag=TRUE)
```

And, finally, the ***Spatial Error*** counterpart by typing:

```
> model8 <- spml(y~X+Z, listw=W, spatial.error="b", lag=
FALSE)
```

4.5.4 Estimating Geographically Weighted Regression models

To perform a Geographically Weighted Regression estimation, to start, we need to download, the library `spgwr`, using the familiar command `install.packages("spgwr")` and invoking it back at the beginning of each new session by typing the command `library(spgwr)`. Before we can perform a *GWR* estimation, it is necessary to have a system of coordinates of the centroids of each polygon. Let us call it `coords` (for the generation of a system of coordinates see, for example, section 2.3.3). Let us now consider the following *GWR* model: $y_i = \beta_{0i} + X\beta_{1i} + Z\beta_{2i} + \varepsilon_i$. To calibrate the bandwidth, in order to identify the optimal value using the cross-validation method, type the following command:

```
bw <- gwr.sel(y ~ X + Z, coords, gweight=gwr.Gauss, adapt=
TRUE)
```

the Gaussian kernel is the default of the command and can be omitted. If we wish to change the default, we can use different kernels such as,

e.g., `gwr.bisquare`, or `gwr.tricube`. Finally, to fit the GWR model, type:

```
modelgwr <- gwr(y ~ X + Z, coords, adapt=bw, hatmatrix=
TRUE)
```

and to see the results, type:

```
modelgwr
```

The results of the modeling procedure are now all stored in an `R` object called `Spatial Data Frame (SDF)`. So, for instance, the various coefficients of variable `X` estimated in each location are stored in the object `modelgwr.$SDF$X`. To visualize them on a map, in order to identify possible regularities, type the command:

```
plot(modelgwr.$SDF,   col=cols[findInterval(gwr.model
$SDF$X, brks, all.inside=TRUE)])
```

Finally, to compute *Moran's I* test on residuals, using the Leung et al. (2006) procedure (and having available a weight matrix, say *W*) we can type the command:

```
gwr.morantest(modelgwr, W)
```

Key Terms and Concepts Introduced

- Feasible GS2SLS
- Matrix estimation
- HAC estimators
- Spatial HAC procedures
- Nuisance parameters
- Kernel estimation
- Kernel functions
- Dichotomous variable
- Discrete choice models
- Binary models
- Latent variable
- Latent regression

- Index function
- Standardized logistic distribution
- A-spatial Probit and Logit model
- Marginal effects in a Logit and in a Probit model
- Reduced form of a model
- Structural form of a model
- Spatial Lag Probit model
- EM algorithm
- ML estimator of Spatial Lag Probit model
- GMM estimator of Spatial Error Probit model
- Generalized Probit and Logit disturbances
- Linearized GMM estimator
- Gradient terms
- Sort and long panels
- Pooled time-series
- Individual and idiosyncratic error term
- Unobserved heterogeneity
- Kronecker product
- Fixed effects
- Random effects
- Spatial Error model with Random Effects
- KKP model
- Spatial Lag model with Random Effects
- Spatial Error model with Fixed Effects
- Spatial Lag model with Fixed Effects
- Time-demeaning
- Within residuals
- Stable dynamics spatial panels
- Unstable dynamics spatial panels
- Non-stationary spatial relationships
- Scan statistics
- Moving windows
- Locally weighted regression
- Geographically weighted regression
- Kernel function
- Bandwidth
- Gaussian, Bi-square and Tri-cube kernels
- Calibration
- Cross-validation
- Bayesian GWR
- Discrete choice GWR

Questions

1. Why is the hypothesis of errors homoscedasticity often implausible in the case of spatial regressions?
2. Why, in a spatial HAC estimation, do we need to consider a kernel smoothing function?
3. Why in the case of a SARAR model with heteroscedastic disturbances, do we need to consider the error variances as nuisance parameters?
4. What is the difference between a Probit and a Logit model for binary choices?
5. What are the reasons why a standard *ML* approach cannot be used to estimate a Probit model on spatial data?
6. Why is a linearized version of the Generalized Method of Moments estimation introduced in spatial Probit/Logit estimation?
7. What is the difference between the fixed effects and the random effects in panel data modeling? How can we choose between the two models?
8. Describe the specificity of the two error components in a panel data framework. How can we interpret the "individual" and the "idiosyncratic" error component?
9. Describe the operation of time-demeaning. What is the effect of demeaning the data in a spatial panel data model?
10. What are the differences between the *SEM-RE* and the *KKP* specification?
11. What is the advantage of using a Geographically Weighted Regression model with respect to a traditional regression model? In what circumstances is it more appropriate to use? In what sense can it be seen as an alternative to Spatial Econometric models to account for spatial effects?

Exercises

Exercise 4.1 You are given a simple linear regression model $y=\beta_0 +\beta_1 x+\varepsilon \;\; \varepsilon \,|X \approx i.i.d.N(0,\sigma^2_{\varepsilon\,n} I_n)$ observed at certain level of aggregation (e.g. regions). Suppose further that data are aggregated at a higher level of aggregation (e.g. countries). Call $\bar{y}$ the dependent variable at this level of aggregation, similarly for x and ε. Suppose further that we have m countries and $n_1, n_2, ..., n_m$ regions respectively in each country. Let us now specify the model at the higher level of aggregation as: $\overline{y} = \beta_0 + \beta_1 \overline{x} + \overline{\varepsilon}$. Prove that even if the regression model is homescedastic at a lower level of aggregation, this property is generally lost at a higher level of

aggregation. [Hint: Use the aggregation matrix ${}_mG_n$ such that $\bar{y} = Gy$ and similarly for the other variables].

Exercise 4.2 One property of the kernel functions is that they integrate to 1 in their domain. Prove such a property for the Uniform, the Epanechnikov and the Quadratic (Bi-weight) kernels.

Exercise 4.3 Consider again the dataset used in Exercise 3.1 related to the 27 European Member States. Estimate a heteroscedastic SARAR model which explains the *growth* in one country as a function of the *education expenses* in the same country plus additional spatial lag and spatial error terms. Estimate the model using the parametric approach (section 4.1.2).

Exercise 4.4 You are given the following latent regression $y^{\bullet} = X\beta + u$ of variable $y = I(y^{\bullet} > 0)$ and the following Spatial Error model specification $u = PWu + \varepsilon$, with $\varepsilon | X \approx n.i.d.N(0, I)$. Derive the spatial Probit log-likelihood function for a sample of dimension *n*.

Exercise 4.5 Consider again the data related to the 27 EU Member States used in Exercises 3.1 and 4.3. Consider also some additional data related to the intensity of *Hi-tec* export (downloadable from the website http://epp.eurostat.ec.europa.eu/portal/page/portal/region_cities/regional_statistics/data/main_tables). In particular we classified each Member State as "high-intensity" if the percentage of hi-tec exports is greater than 18 per cent. The data are shown in the following table.

COUNTRY CODE	COUNTRY	% of Hi-tec Exports	Hi-tec Intensity	COUNTRY CODE	COUNTRY	% of Hi-tec Exports	Hi-tec Intensity
BE	Belgium	8.8	0	AT	Austria	11.7	0
BG	Bulgaria	4.6	0	PL	Poland	5.7	0
CZ	Czech Republic	15.2	0	PT	Portugal	3.7	0
DK	Denmark	12.3	0	RO	Romania	8.2	0
DE	Germany	14.0	0	SI	Slovenia	5.5	0
EE	Estonia	6.9	0	SK	Slovakia	5.9	0
IE	Ireland	22.1	1	FI	Finland	13.9	0
ES	Spain	4.8	0	SE	Sweden	14.6	0
FR	France	19.7	1	UK	United Kingdom	19.0	1
IT	Italy	6.8	0	EL	Greece	6.6	0
CY	Cyprus	20.1	1	LU	Luxembourg	41.9	1
LT	Lithuania	5.8	0	LV	Latvia	5.3	0
HU	Hungary	22.2	1	MT	Malta	35.2	1
NL	Netherlands	18.4	1				

On the basis of the results of Exercise 3.1 and of these new data, estimate a Probit model explaining *Hi-tec intensity* as a factor of *Education* in both the a-spatial and its spatial version using all three estimators (ML, GMM and LGMM) discussed in section 4.2.

Exercise 4.6 In a spatial panel constituted by $n = 3$ individuals, we have T = 2 time observations. Given the following W matrix: $W = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ and the parameters $\sigma_\mu^2 = 1$ and $\lambda = 0.3$, derive the explicit form of the matrix $I_T \otimes A$ and of the matrix $\sigma_\mu^2 \left(J_T \otimes I_n \right)$.

Exercise 4.7 From the `splm` package, load the dataset `Insurance` and the binary contiguity matrix for the Italian provinces (`itaww`). See Millo (2014) for details. Transform the contiguity matrix into a `listw` object as described in 2.3.2. Estimate a Spatial Error panel data model explaining the variable *real per-capita insurance premia* (coded as `ppcd`) as a function of the variable *real per-capita GDP* (`rgdp`) using the Maximum Likelihood strategy, with random effects, both in the *SEM-RE* and in the *KKP* specifications. Discuss the results with particular attention to the spatial parameter estimates.

Exercise 4.8 Consider again the `R` dataset `Boston` already presented in Example 3.3 and Exercise 3.6 and study the relationship between the variable "House median value" (MEDV) and the variable "number of rooms" (RM). Estimate an a-spatial model using the *OLS* criterion, then estimate a Spatial Lag Model with the Maximum Likelihood procedure, finally estimate a Geographically Weighted Regression model using a Gaussian kernel and the bandwidth specified using the cross-validation method. Compare the results obtained with the various methods. In particular, compare the estimated regression coefficients obtained with OLS and SLM with the median value of the GWR coefficients. Compute Moran's *I* statistics for the residuals of the GWR model.

References to the Chapter

Amemiya, T. (1985) *Advanced Econometrics*, Oxford: Basil Blackwell.

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

Anselin, L. (1990) 'Some Robust Approaches to Testing and Estimation in Spatial Econometrics', *Regional Science and Urban Economics*, 20, 141–63.

Anselin, L. (2002) 'Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models', *Agricultural Economics,* 27, 3, 247–67.

Anselin L., Le Gallo J., Jayet H. (2008) 'Spatial Panel Econometrics', in L. Matyas, P. Silvestre (eds), *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*, 3rd edition, Heidelberg: Springer-Verlag, pp. 625–60.

Arbia, G. (1989) *Spatial Data Configuration in the Statistical Analysis of Regional Economic Data and Related Problems*, Dordrecht: Kluwer Academic Press.

Arbia, G. (1990) 'On Second Order Non-Stationarity in Two-Dimensional Lattice Processes', *Computational Statistics and Data Analysis*, 9, 147–60.

Arbia, G. (2012) 'A Lustrum of SEA: Recent Research Trends Following the Creation of the Spatial Econometrics Association (2007–2011)', *Spatial Economic Analysis*, 6, 4, 377–96.

Baltagi, B. (2008) *Econometric Analysis of Panel Data*, 4th edition, New York: Wiley.

Baltagi B. and Pesaran, H. (2007) 'Heterogeneity and Cross-Section Dependence in Spatial Panel Data Models: Theory and Applications Introduction', *Journal of Applied Econometrics,* 22, 2, 229–32.

Baltagi, B.H. and Pinnoi, N. (1995) 'Public Capital Stock and State Productivity Growth: Further Evidence from an Error Components Model', *Empirical Economics,* 20, 2, 351–9.

Baltagi, B., Egger, P. and Pfaffermayr, M. (2013) 'A Generalized Spatial Panel Data Model with Random Effects', *Econometric Reviews*, 32, 5, 650–85.

Barro, R.J. and Sala-i-Martin, X. (1992) 'Convergence', *Journal of Political Economy*, 100, 2, 223–51.

Beron, K.J. and Vijverberg, W.P. (2004) 'Probit in a Spatial Context: A Monte Carlo Approach', in L. Anselin, R. Florax and S. Rey (eds) *Advances in Spatial Econometrics,* Heidelberg: Springer-Verlag, pp. 169–95.

Brundson, C., Fotheringham, A.S. and Charlton, M. (1996) 'Geographically Weighted Regression: A Method for Exploring Spatial Non-Stationarity', *Geographical Analysis*, 28, 4, 281–98.

Brundson, C., Fotheringham, A.S. and Charlton, M. (1998) 'Spatial Non-Stationarity and Autoregressive Models', *Environment and Planning, A,* 30, 6, 957–73.

Case, A. (1992) Spatial pattern in household demand, *Econometrica,* 59, 4, 953–65.

Cleveland, W. and Devlin, S. (1988) 'Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting', *Journal of the American Statistical Association,* 82, 596–610.

Conley, T. (1999) 'GMM Estimation with Cross-Sectional Dependence', *Journal of Econometrics,* 92, 1–45.

Croissant, Y. and Millo, G. (2008) 'Panel Data Econometrics in R: The plm Package', *Journal of Statistical Software,* 27, 2, 1–43.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) 'Maximum Likelihood from Incomplete Data via the EM Algorithm', *Journal of the Royal Statistical Society. Series B,* 39, 1, 1–38.

Driscoll J.C. and and Kraay, A.C. (1998) 'Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data', *The Review of Economics and Statistics,* 80, 549–60.

Dubin, R.A. (1992) 'Spatial Autocorrelation and Neighborhood Quality', *Regional Science and Urban Economics,* 22(3), 433–52.

Elhorst, J.P. (2003) 'Specification and Estimation of Spatial Panel Data Models', *International Regional Sciences Review*, 26, 3, 244–68.
Elhorst, J.P. (2009) 'Spatial Panel Data Models', in M.M. Fischer and A. Getis (eds), *Handbook of Applied Spatial Analysis,* Heidelberg: Springer-Verlag.
Elhorst, J.-P., and Fréret, S. (2009) 'Evidence of Political Yardstick Competition in France Using a Two-Regime Spatial Durbin Model with Fixed Effects', *Journal of Regional Science*, 49, 5, 931–51.
Fisher, R.A. (1959) *Statistical Methods and Statistical Inference*, New York: Hafner.
Fleming M.M. (2004) 'Techniques for Estimating Spatially Dependent Discrete Choice Models', in L. Anselin, R. Florax and S. Rey (eds), *Advances in Spatial Econometrics,* Heidelberg: Springer-Verlag, pp. 145–68
Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. (1998) 'Geographically Weighted Regression: A Natural Evolution of the Expansion Method of Spatial Data Analysis', *Environment and Planning, A,* 30, 1905–27.
Fotheringham, A.S., Brunsdon, C. and Charlton, M.E., (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships,* Chichester: Wiley.
Fotheringham, A.S., Brunsdon, C. and Charlton, M.E. (2007) 'Geographically Weighted Discriminant Analysis', *Geographical Analysis,* 4, 376–96
Glaz J., Naus, J. and Wallenstein, S. (2001) *Scan Statistics*, Springer Series in Statistics, Heidelberg: Springer-Verlag.
Greene, W. (2011) *Econometric Analysis,* 7th edition, Upper Saddle River, NJ: Pearson Educational.
Grenander, U. and Rosenblatt, M. (1957) *Statistical Analysis of Stationary Time Series*. Wiley, New York. [2nd ed. (1984), Chelsea, New York.].
Hausman, J. (1978) 'Specification Tests in Econometrics', *Econometrica*, 46, 1251–71.
Hoh, J. and Ott, J. (2012) 'Scan Statistics in Human Gene Mapping', *Am J Hum Genet,* 91, 970.
Kapoor, M., Kelejian, H. and Prucha, I. (2007) 'Panel Data Model with Spatially Correlated Error Components', *Journal of Econometrics*, 140, 1, 97–130.
Kelejian, H. and Prucha, I. (1999) 'A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model', *International Economic Review*, 40, 2, 509–33.
Kelejian, H. and Prucha, I. (2007) 'HAC Estimation in a Spatial Framework', *Journal of Econometrics*, 140, 1, 131–54.
Kelejian, H. and Prucha, I. (2010) 'Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances', *Journal of Econometrics*, 157, 1, 53–67.
Klier, T. and McMillen, D.P. (2008) 'Clustering of Auto Supplier Plants in the United States: Generalized Method of Moments Spatial Logit for Large Samples', *Journal of Business and Economic Statistics,* 26, 4, 460–71.
Lee, L. and Yu, J. (2010a) 'Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects', *Journal of Econometrics*, 154, 2, 165–85.
Lee, L. and Yu, J. (2010b) 'Some Recent Developments in Spatial Panel Data Models', *Regional Science and Urban Economics*, 40, 5, 255–71.
Lee, L.F. and Yu, J. (2011) 'Estimation of Spatial Panels', *Foundation and Trends in Econometrics*, 4, 1/2, 1–164.
Lee, L.F. and Yu, J. (2012) 'Spatial Panels: Random Components Versus Fixed Effects', *International Economic Review*, 53, 4, 1369–412.

LeSage, J. (2004) 'A Family of Geographically Weighted Regression Models', in L. Anselin, R.J. Florax, and S. Rey (eds), *Advances in Spatial Econometrics: Methodology, Tools and Applications,* Springer Verlag, Berlin, pp. 241–64.

LeSage, J. and Pace, K. (2009) *Introduction to Spatial Econometrics*, Boca Raton, FL: Chapman & Hall/CRC Press.

Leung, Y., Mei, C.L. and Zhang, W.X. (2000) 'Testing for Spatial Autocorrelation Among the Residuals of Geographically Weighted Regression', *Environment and Planning, A,* 32, 5, 871–90.

McMillen, D.P. (1992) 'Probit with Spatial Autocorrelation', *Journal of Regional Science,* 32, 3, 335–48.

McMillen, D. (1996) 'One Hundred Fifty Years of Land Values in Chicago: A Non-Parametric Approach', *Journal of Urban Economics,* 40, 100–24.

McMillen, D. and McDonald, J.F. (1997) 'A Nonparametric Analysis of Employment Density in a Polycentric City', *Journal of Regional Science,* 37, 591–612.

McMillen, D. and McDonald, J.F. (2004) 'Locally Weighted Maximum Likelihood Estimation: Monte Carlo Evidence and an Application', in L. Anselin, R.J.G.M. Florax and S. Rey (eds), *Advances in Spatial Econometrics: Methodology, Tools and Applications,* Berlin: Springer Verlag, pp. 225–39.

Millo, G. (2014) 'Maximum Likelihood Estimation of Spatially and Serially Correlated Panels with Random Effects', *Computational Statistics & Data Analysis,* 71, 914-33

Millo, G. and Pasini, G. (2010) 'Does Social Capital Reduce Moral Hazard? A Network Model for Non-Life Insurance Demand', *Fiscal Studies*, 31, 3, 341–72.

Millo, G. and Piras, G. (2012) 'splm: Spatial Panel Data Models in R', *Journal of Statistical Software*, 1–43. URL http://www.jstatsoft.org/v47/i01/.

Mundlak, Y. (1978) 'On the Pooling of Time Series and Cross-Section Data', *Econometrica*, 46, 69–85.

Munnell, A.H. (1990) 'Why Has Productivity Growth Declined? Productivity and Public Investment', *New England Economic Review*, 30, 3–22.

Mur, J., Lopez, F. and Angulo, A. (2008) 'Symptoms of Instability in Models of Spatial Dependence', *Geographical Analysis,* 40, 2, 189–211.

Mutl, J. and Pfaffermayr, M. (2011) 'The Hausman Test in a Cliff and Ord Panel Model', *Econometrics Journal*, 14, 1, 48–76.

Newey, W. and West, K. (1987) 'A Simple Positive Semi-Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix', *Econometrica,* 55, 703–8.

Oberhofer, W. and Kmenta, J. (1974) 'A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models', 42, 3, *Econometrica*, 579–90.

Pace K. and LeSage, J. (2004) 'Spatial Autoregressive Local Estimation', in A. Getis, J. Mur and H. Zoller (eds), *Recent Advances in Spatial Econometrics*, New York: Palgrave Macmillan, pp. 31–51.

Paez, A., Long, F. and Faber, S. (2008) 'Moving Windows Approaches for Hedonic Price Estimation: An Empirical Comparison of Modeling Techniques', *Urban Studies,* 45, 8, 1565–81.

Paez, A., Uchida, T. and Miyamoto, K. (2002) 'A General Framework for Estimation and Inference of Geographically Weighted Regression Models', *Environment and Planning, A,* 34, 4, 733–54.

Pinkse, J. and Slade M.E. (1998) 'Contacting in Space: An Application of Spatial Statistics to Discrete-Choice Models', *Journal of Econometrics,* 85, 125–54.

Pinkse, J., Slade, M. and Shen, L. (2006) 'Dynamic Spatial Probit with Fixed Effects Using One Step GMM: An Application to Mine Operating Decisions', *Spatial Economic Analysis*, 1, 53–99.

Pinkse, J., Slade, M. E. and Brett, C. (2002) 'Spatial Price Competition: A Semi-parametric Approach', *Econometrics,* 70, 1111–53

Piras, G. (2011) 'Estimation of Random Effects Spatial Panel Data Models: Some Additional Evidence', *Working Paper*.

Smirnov, O. (2010) 'Modelling Spatial Discrete Choice', *Regional Science & Urban Economics,* 40, 5, 292–8.

Wang, H., Iglesias, E.M. and Wooldridge, J.M. (2013) 'Partial Maximum Likelihood Estimation of Spatial Probit Models', *Journal of Econometrics*, 172, 77–89.

Wang, X. and Kockelman, K.M. (2009) 'Bayesian Inference for Ordered Response Data with a Dynamic Spatial Ordered Probit Model', *Journal of Regional Science*, 49, 5, 877–913.

Wheeler, D. C. and Paez, A. (2010) 'Geographically Weighted Regression', in M. Fischer and A. Getis (eds), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, Berlin: Springer Verlag.

Wooldridge, J. (2002) *Econometric Analysis of Cross-Section and Panel Data*, Cambridge, MA: MIT Press.

5
Alternative Model Specifications for Big Datasets

5.1 Introduction

In Chapters 3 and 4 we have presented a series of estimation techniques for the general spatial econometric linear model which can basically be traced back to the two paradigms of the Maximum Likelihood and the Generalized Two-Stage Least Squares. In particular, when discussing the Maximum Likelihood approach we pointed out that the likelihood functions cannot be maximized analytically due to a high degree of non-linearity in the parameters and so we have to use numerical approximations. However, the likelihood function involves the calculation of the determinant of a matrix whose dimension depends on the sample size and that has to be evaluated repeatedly for each trial value of the spatial correlation parameter in a numerical search. If n is very large, as often happens in many empirical applications with a massive quantity of data, this operation may be highly demanding, if not prohibitive. The eigenvalues decomposition suggested by Ord (1975) and used for many years in the literature (see section 3.4.3) also has some limitations. Indeed, Kelejian and Prucha (1998) report that the computation of eigenvalues by standard subroutines for general non-symmetric matrices, are also approximate and may already be highly inaccurate for W matrices that are of the order 400-by-400, that is, for relatively small sample sizes. The accuracy improves if the weight matrix is symmetric, but unfortunately this is not the case when dealing with row-standardized versions. Many other approximations were proposed in the literature (see Arbia, 2006 for a review). Such approximations are usually accurate if the weight matrix is sparse (that is, it contains a high percentage of zero entries), but not if they are very large. In this respect, Bell and Bockstael (2000) report accuracy

problems in determining the eigenvalues of sparse matrices of the order of 2000-by-2000. The problem is exacerbated if the matrices are very dense (that is, if they contain a high percentage of non-zero entries) like those frequently encountered, for example, in social interaction applications.

The computational issues associated with the Likelihood maximization was a relevant topic in the 1970s, at times of low-speed and small-memory computers, but even in the current situation of increasing computing power, the burden of calculation can become unbearably demanding in terms of both computing time and also of the required computer storage because the availability of very large databases has also been increasing at an accelerated rate. For example, many health data provided by the US National Center for Health Statistics are available at a county level involving, therefore, more than 3,000 observations and thus requiring the inverse of a W matrix of dimension 3000-by-3000 in any spatial econometric modeling of them. However, this is still nothing compared to many other geo-referenced economic data such as, for instance, the observations related to US establishments provided by the US Census Bureau's Longitudinal Business Database (LBD) which refers to some four million individual, geo-located firms. Other examples include satellite images used in land cover assessment or high resolution medical images where spatial econometric techniques have to accommodate millions of spatially dependent observations composed of pixilated imagery. Furthermore, the data available in human genome mapping display spatial dependence and involve millions of observations for which a joint likelihood approach is completely unfeasible. We have quoted only a few of the many possible examples that are emerging in different fields. It is easy to predict that the future demand for big spatial data will increase further and so will the demand for developing appropriate methods for analyzing these new data sources.

The continuing relevance of computational issues in spatial econometric modeling is witnessed by the large number of approximate solutions suggested in the literature in recent years (for example, Smirnov and Anselin, 2001; Griffith, 2000, 2004; Pace and LeSage, 2004). These contributions have been introduced to speed up the calculation of the procedures introduced in Chapters 3 and 4, but some of the most recent literature has concentrated, instead, on the specification of alternative models which depart from the conventional spatial autoregressive class with the specific aim of reducing the computational obstacles. These contributions have some common characteristics: they are theoretically

very simple, they produce closed form solutions for the *ML* estimators and they improve the numerical performance dramatically.

In this chapter we will review some of these alternative specifications. In particular, in section 5.2 we will consider the *Matrix Exponential Spatial Specification* (*MESS*), section 5.3 is devoted to the *unilateral approximation* of spatial models and, finally, in section 5.4 we will present the *bivariate coding technique* and the associate bivariate marginal maximum likelihood estimation.

5.2 The MESS specification

5.2.1 A MESS Spatial Lag specification

To introduce the MESS specification, let us consider again the Spatial Lag model defined in scetion 3.5:

$$y = \lambda W y + Z\beta + u \qquad |\lambda| < 1 \tag{5.1}$$

with $u|X \approx i.i.d.N(0, \sigma^2_{\varepsilon\, n} I_n)$ and W non-stochastic and often (although not necessarily) row-standardized in the empirical applications. Let us re-write expression (5.1) in the following alternative way:

$$(I - \lambda W)y = Z\beta + u \tag{5.2}$$

As said, the computational problems emerging when analyzing very large databases derive mainly from the inversion of the variance-covariance matrix of the model $(I - \lambda W)^{-1}$ (see Equation 3.47). Now let us generalize Equation (5.2) and specify a new generalized Spatial Lag model as:

$$Sy = Z\beta + u \tag{5.3}$$

with S a real positive definite matrix. The specification of the matrix S in different ways has an effect on the variance-covariance matrix and, as a consequence, produces the specification of different spatial econometric models. Following a procedure suggested by Chiu et al. (1996), LeSage and Pace (2007), proposed the following matrix exponential specification for the matrix S:

$$S = e^{\alpha W} \tag{5.4}$$

Before reviewing the computational advantages of this specification let us rewrite matrix *S* exploiting the following power series expansion:

$$S = e^{\alpha W} = \sum_{i=0}^{\infty} \frac{\alpha^i W^i}{i!} \tag{5.5}$$

In this expression we find the terms $W^1, W^2, \ldots\ldots$ which require explanation. In spatial econometrics, by analogy with time series analysis, we can define higher-order levels of neighborhood by extending the notion of a spatial lag introduced in section 2.1. In particular, we have defined the *W* matrix in such a way that its non-zero entries refer to pairs of neighboring spatial units. We can refer to them as to "first-order neighbors". The non-zero elements of the matrix W^2 similarly refer to those pairs that are neighbors of the first-order neighbors. In this sense they can be considered *second-order neighbors*. Higher-order *W* matrices are similarly defined. As an alternative, higher-order neighbors can also be defined by considering various levels of distance.

Chiu et al. (1996) list a series of advantages of transformation (5.5). In particular the following useful properties hold:

Property 1. For any real positive definite matrix S, there always exists a real symmetric matrix αW such that $S = e^{\alpha W}$
Property 2. For any symmetric real matrix W, *S* is a positive definite matrix
Property 3. The inverse of *S* is $S^{-1} = e^{-\alpha W}$
Property 4. The determinant of *S* is $|S| = e^{tr(\alpha W)}$. Since all the diagonal elements of a *W* matrix are zero by definition ($W_{ii} = 0$ see section 2.1), this expression simplifies further as $|S| = e^{tr(\alpha W)} = e^0 = 1$.

The above properties ensure that:

1. It is always possible to utilize this approach.
2. The approach leads to well-defined variance-covariance matrices.
3. The inverse of the *S* matrix is very simple to calculate.
4. The determinant of the S matrix is always equal to 1.

As a consequence, the log-likelihood of the MESS model will not require the calculation of the log-determinant which is the main source of computational problems in the case of the Spatial Lag model.

The parameter α in Equation (5.5) is related to the parameter λ in the Spatial Lag specification (5.1) and controls for the level of spatial

correlation. In fact, from the equation, $S = (I - \lambda W) = e^{-a W}$ by taking the maximum row sum norm of the two matrices and equating them we obtain:

$$1 - \lambda = e^{a} \tag{5.6}$$

or inversely:

$$\lambda = \ln(1 - \lambda) \tag{5.7}$$

From Equation (5.7), given that $|\lambda| < 1$, we derive that the range of a is $-\infty < a \leq 0$, for positive λ and $0 < a \leq 0.693147$ for negative λ. In particular, when $\lambda = 0$, then also $a = 0$, and when $\lambda \to 1$, then $a \to -\infty$ although we can achieve values of λ very close to 1 already for $a = -5$ ($\lambda = 0.99$).

For estimation and hypothesis testing purposes, let us now consider the log-likelihood of this new specification. First of all notice that in Equation (3.47) if we replace the term $(I - \lambda W)$ with S, we have:

$$E\left(yy^{T}\right) = \sigma_{\varepsilon}^{2}\Omega = \sigma_{\varepsilon}^{2}\left(I - \lambda W\right)^{-1}\left(I - \lambda W\right)^{-T} = \sigma_{\varepsilon}^{2}S^{-1}S^{-T} \tag{5.8}$$

So the log-likelihood in Equation (3.49) now can be expressed as:

$$\begin{aligned} l(\sigma^{2}, \rho, \beta; y) &= const - \frac{1}{2}\ln\left|\sigma_{\varepsilon}^{2}S^{-1}S^{-T}\right| - \frac{1}{2\sigma_{\varepsilon}^{2}}\left[y - S^{-1}X\beta\right]^{T} SS^{T}\left[y - S^{-1}Z\beta\right] \\ &= const - \frac{1}{2}\ln\left|\sigma_{\varepsilon}^{2}S^{-1}S^{-T}\right| - \frac{1}{2\sigma_{\varepsilon}^{2}}\left[Sy - Z\lambda\right]^{T}\left[Sy - Z\lambda\right] \end{aligned} \tag{5.9}$$

Notice that from Equation (3.50), we have $\left|S^{-1}S^{-T}\right| = \left|S^{-1}\right|\left|S^{-T}\right|$ and, from Property 4 of the exponential matrix reported above, this simplifies as $\left|S^{-1}S^{-T}\right| = 1$. As a consequence we have:

$$l(\sigma^{2}, \lambda, \beta; y) = c - \frac{1}{2}\ln\sigma_{\varepsilon}^{2} - \frac{1}{2\sigma_{\varepsilon}^{2}}\left[Sy - Z\beta\right]^{T}\left[Sy - Z\beta\right] \tag{5.10}$$

with S a function of λ and hence of a. Since no determinant is present in Equation (5.10), when estimating the parameters β and a, maximizing the log-likelihood is equivalent to minimizing the term $\left[Sy - Z\beta\right]^{T}\left[Sy - Z\beta\right]$, which corresponds to the sum-of-squared errors of the transformed model.

From a practical point of view, the infinite expansion shown in (5.5) has obviously to be truncated to, say, q terms leading to the approximation:

$$S = e^{a\,W} = \sum_{i=0}^{\infty} \frac{a^{\,i} W^i}{i!} \approx \sum_{i=0}^{q} \frac{a^{\,i} W^i}{i!} \tag{5.11}$$

LeSage and Pace (2007) have shown that the stopping role q depends on the amount of (positive or negative) spatial correlation while it does not depend on the sample size. They also show, through simulated data, that when the absolute value of the spatial correlation is not very high (lower than 0.95, as usually happens in practical circumstances), a truncation at $q=16$ produces a very good approximation. For higher levels of λ more terms in the expansion are needed. If the W matrix is sparse, the calculations are rather simple and do not require particular computational effort. However, if the weight matrix is very dense (as happens, for example, in social networks applications) then the transformation $\sum_{i=0}^{\infty} \frac{a^{\,i} W^i}{i!}$ will also be dense and the calculation of matrix S could be prohibitive in terms of the required memory and of the computing time, thus nullifying the computational benefits of the procedure. However, notice that the procedure does not require the calculation of S, but only of the matrix product Sy (see Equation (5.10)), and this operation is significantly simpler.

LeSage and Pace (2007) show that another advantage of the *MESS* specification is the possibility of deriving closed-form solutions for the parameter estimations, something which is not possible for the traditional autoregressive models examined in Chapter 3 due to the high degree of nonlinearity of the likelihood function. To derive such closed-form solutions, we need to express the transformation (5.5) in matrix form. In order to achieve this aim, let us introduce some definitions. First of all define a matrix containing all the values of the dependent variable y and of its lagged terms of various orders:

$$Y = \left[y, Wy, W^2y, \ldots., W^{q-1}y\right] \tag{5.12}$$

where q is the stopping value chosen in the power expansion approximation (5.11). If y is, say, an n-by-*1* vector, then the matrix Y is an n-by-q matrix. Secondly, let us define a diagonal (q-by-q) matrix,

say G, containing the first q coefficients of the power expansion in Equation (5.11):

$$G = \begin{bmatrix} 1/0! & 0 & \dots & 0 \\ 0 & 1/1! & \dots & 0 \\ \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 1/(q-1)! \end{bmatrix} \tag{5.13}$$

Finally, define a vector, say $\upsilon(\alpha)^T$, containing the power of the parameter α which controls for spatial correlation:

$$\upsilon(\alpha)^T = \left[1, \alpha, \alpha^2, \dots, \alpha^{q-1}\right] \tag{5.14}$$

Using the definitions contained in Equation (5.12), (5.13) and (5.14) we can re-express Equation (5.5) as:

$$Sy \approx YG(\alpha) \tag{5.15}$$

the approximation being due to the truncation in the power expansion. Premultiplying both sides of (5.15) by the idempotent projection matrix $P = I - X(X^TX)^{-1}X^T$, we can derive the sums of squares of the residuals expressed as a function of the transformed elements, that is:

$$u^Tu = \upsilon(\lambda)^T G(Y^TP^TPY)G\upsilon(\alpha) \tag{5.16}$$

or, more simply, as:

$$u^Tu \approx \upsilon(\alpha)^T Q\upsilon(\alpha) \tag{5.17}$$

having defined $Q = G(Y^TP^TPY)G$, so that in the end:

$$\min(u^Tu) \approx \min(\upsilon(\alpha)^T Q\upsilon(\alpha)) \tag{5.18}$$

LeSage and Pace (2007) have shown that Equation (5.18) is a polynomial in α that admits a closed-form solution, and they proved that such a solution is unique and it represents a minimum. Furthermore, the second-order conditions provide an alternative way of evaluating the Hessian and, as a consequence, the standard errors on which we can base inference and hypothesis testing on the parameters.

The Matrix Exponential Spatial Specification is an approximate technique (due to the truncation in Equation (5.11)) which dramatically reduces the computational time and the computer storage required for estimating spatial econometric models. However, the proponents of the methodology show through some empirical analyses (LeSage and Pace, 2009) that this approximation is very good. Using a dataset on the house prices of 506 census tracts in the Boston area (see Examples 3.3. and 3.4) the authors estimate a spatial model using both a full likelihood Spatial Lag and the *MESS* specification and then compared the results obtained in terms of the estimates of the parameters σ^2, β and λ (comparing it with the value of a implied by Equations (5.6) and (5.7)). The empirical results provide a clear indication that the inference is identical, both in terms of the regression parameters' point estimation and in terms of the *p-values* of the hypothesis testing procedures. Furthermore, the estimation of λ in the full likelihood method applied to the Spatial Lag model is coincident with the one derived from the estimation of a in the MESS specification. So the computational benefits are not at the expenses of estimation precision, at least in the relatively small dataset examined.

5.2.2 A MESS Spatial Error Specification and further extensions

In a similar fashion, we can consider a *MESS* specification corresponding to a Spatial Error autoregressive model. Let us now assume, as in Equation (3.13) in Chapter 3.4, the model:

$$y = Z\beta + u \tag{5.19}$$

but now, instead of modeling the residuals like in Equation (3.14) as $u = \rho Wu + \varepsilon, \quad |\rho| < 1$, let us consider the following more general model:

$$\varepsilon \approx N(0, \Omega) \tag{5.20}$$

with Ω a generic variance-covariance matrix for the regression disturbances.

If in Equation (5.20) we assume $\Omega = e^{aW}$, then we have a *MESS* specification of a Spatial Error, that can be expressed as $\Omega = e^{aW} = \sum_{i=0}^{\infty} \frac{a^i W^i}{i!}$ or:

$$\Omega^{-1} = e^{-aW} = \sum_{i=0}^{\infty} \frac{-a^i W^i}{i!} \tag{5.21}$$

If, conversely, $\Omega = (I - \rho W)^{-1}(I - \rho W)^{-T}$, or

$$\Omega^{-1} = (I - \rho W)(I - \rho W)^{T} \tag{5.22}$$

we again have the conventional Spatial Error model. Notice that we can express Equation (5.22) as:

$$\begin{aligned} \Omega^{-1} &= (I - \rho W)(I - \rho W)^{T} = (I - \rho W)\left(I - \rho W^{T}\right) \\ &= I - \rho W - \rho W^{T} + \rho^{2} W W^{T} \end{aligned} \tag{5.23}$$

or

$$\Omega^{-1} = I - 2\rho W + \rho^{2} W^{2} \tag{5.24}$$

which can be interpreted as a particular case of Equation (5.21) when $-\alpha = -2\rho$, $-\frac{\alpha^{2}}{2} = \rho^{2}$ and the expansion is truncated at $q = 2$.

Real data analysis based on the 3,107 US counties and the analysis of simulated data (laid on the same geographical partition) were used by LeSage and Pace (2009) to show that the MESS specification provides a very good approximation of the estimated parameters to the full likelihood Spatial Error model estimators.

More general spatial weights can be considered extending the flexibility of the MESS specification. This is achieved by including an extra parameter in the matrix exponential transformation which controls for the speed with which correlation decays in the higher order neighborhoods. In particular LeSage and Pace (2009) suggest the following expression for the matrix W to be used in Equation (5.4):

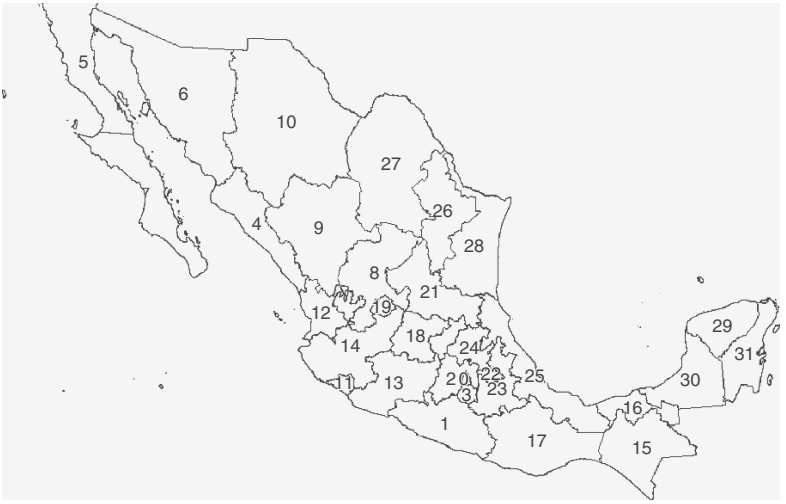
$$W = \sum_{i=1}^{m} \frac{\phi^{i} W_{i}}{\sum_{i=1}^{m} \phi^{i}} \tag{5.25}$$

In this expression, W_i is a spatial weight matrix containing non-zero elements for the i-th closest neighbor, ϕ is a distance decay parameter such that $0 \leq \phi \leq 1$ and the term in the denominator is a normalizing factor. By definition, the new matrix W has zero elements in the main diagonal and is such that the sum of each row is equal to 1. Of course

the greater flexibility of this specification is the price paid for higher computational effort, also due to the presence of two more parameters to be estimated (m and ϕ). For this reason the proponents of the methodology suggest a Bayesian approach to the phase of estimation and hypothesis testing by specifying appropriate priors for the parameters.

Example 5.1 Health planning in Mexico

Let us now consider, as an example, the case of a Spatial Lag model estimated via the MESS approximation. The dataset we are using in this example refers to some data on Mexican states in 2010. Without the aim of contributing substantially to the problem, but just for the sake of illustrating the MESS procedure, let us imagine that, for planning purposes, we wish to estimate a relationship which explains the number of doctors per 1,000 inhabitants in each state as a function of the number of health staff per 1,000 inhabitants, the percentage of population over 65 and the per capita health expenditure in 2000 and in 2010. The shapefiles of the 32 Mexican states can be downloaded at http://www.gadm.org/mexico, and the map is shown here below together with the data necessary for the analysis.



Code	States	Number of Doctors per 1,000 Inhabitants	Per-Capita Health Expenditure 2010	Per-Capita Health Expenditure 2000	Health Staff per 1,000 Inhabitants	% of Population over 65
0	Distrito Federal	3.1	9.437522	0.7521	15.3	7.771425
1	Guerrero	1.4	2.572524	0.427321	5.5	6.917765
2	México	1	2.780234	0.435988	4.7	4.911075
3	Morelos	1.4	2.974519	0.629462	6.1	6.992579
4	Sinaloa	1.7	3.271296	0.700082	7.1	6.616576
5	Baja California	1.2	3.420365	0.574503	5.5	4.457112
6	Sonora	1.9	3.424168	0.770872	8.6	5.950505
7	Baja California Sur	2.1	5.02135	0.667388	9.3	4.272824
8	Zacatecas	1.7	2.897978	0.640509	6.4	7.450485
9	Durango	1.8	3.45163	0.556376	7.2	6.429409
10	Chihuahua	1.2	3.622121	0.691472	6.2	5.67107
11	Colima	2,2	3,983012	0,672654	9	6,205932
12	Nayarit	2,2	3,329873	0,732461	8.1	7,12659
13	Michoacán de Ocampo	1,4	2,164095	0,642803	5.1	7,266681
14	Jalisco	1,5	2,974531	0,702626	6,4	6,269595
15	Chiapas	1	2,139129	0,496826	4.1	4,898949
16	Tabasco	2.2	4.03974	0.493174	9.3	5.190782
17	Oaxaca	1.4	2.532036	0.526933	5.3	7.787479
18	Guanajuato	1.4	2.532267	0.529055	5.4	6.045926
19	Aguascalientes	1.9	3.549239	0.529501	8.3	5.092591
20	Querétaro	1.3	2.726577	0.48921	5.1	5.109038
21	San Luis Potosí	1.5	2.654342	0.594948	5.5	7.155858
22	Tlaxcala	1.4	2.75405	0.399939	5.5	5.957505
23	Puebla	1.3	2.180045	0.501114	4.9	6.295532
24	Hidalgo	1.5	2.566565	0.477279	6.3	6.613051
25	Veracruz de Ignacio de la Llave	1.5	2.88041	0.646695	5.9	7.311852
26	Nuevo León	1.4	3.52745	0.689333	7	5.902514
27	Coahuila de Zaragoza	1.7	3.364743	0.691015	7.9	5.691221
28	Tamaulipas	1.7	3.452093	0.669398	7.6	5.976068
29	Yucatán	1.6	3.754511	0.63864	6.9	6.898322
30	Campeche	2.5	4.689265	0.51398	9.3	5.653901
31	Quintana Roo	1.2	3.595703	0.299441	6.2	2.978851

Source: Population data:http://www.inegi.org.mx/est/contenidos/proyectos/ccpv/cpv2010/Default.aspx Health data: http://www.sinais.salud.gob.mx/estadisticasportema.html.

We considered a (row-standardized) distance-based weight matrix, considering neighbors two states if the distance between their centroids was less than a threshold that guarantees that all states have at least one neighbor.

We estimate a Spatial Lag model using both the exact Maximum Likelihood and the MESS approximation. The specification of the Spatial Lag model used was the following $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \lambda W y + \varepsilon$, *with* y *= number of doctors per 1,000 inhabitants,* X_1 *being the number of health staff per 1,000 inhabitants,* X_2 *and* X_3 *the per capita health expenditure in 2000 and, respectively, in 2010 and* X_4 *the percentage of population over 65.*

	Maximum Likelihood			***MESS Specification***		
	Estimated Value	***Standard Error***	***p-value***	***Estimated Value***	***Standard Error***	***p-value***
β_0	*0.139859*	*0.446938*	*0.75434*	*0.124292*	*0.200922*	*0.54136*
β_1	*0.186960*	*0.012195*	*0.000****	*0.187352*	*0.012738*	*0.000****
β_2	*0.058434*	*0.031370*	*0.06250*	*0.058685*	*0.033965*	*0.09545*
β_3	*–0.049681*	*0.034338*	*0.14794*	*–0.049863*	*0.037131*	*0.19049*
β_4	*0.064011*	*0.025163*	*0.01096**	*0.064794*	*0.024870*	*0.01475**
λ	*–0.14109*	*0.19208*	*0.46263*	*–0.1365203*	*0.16331*	*0.43326*

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

Comparing the results obtained with the two estimation methods we notice that the MESS specification substantially confirms the conclusions of the ML estimation in terms of both the sign and the significance of the parameters. The health staff per 1,000 inhabitants and the percentage of population over 65 appear to be the most significant factors in explaining the geographical distribution of the doctors in the 32 Mexican states. The point estimates of the parameters are all very similar with a small underestimation of the MESS procedure only for the spatial dependence parameter λ *and for the intercept. In both cases the parameter* λ *is negative and not significantly different from zero. The standard errors are all also remarkably similar. With a very small sample size like the one employed in this example, the difference in computing times is obviously negligible. However, the example shows that in big datasets the MESS specification can substantially reduce the computational burden without loss of accuracy of the estimates.*

5.3 The unilateral approximation approach

5.3.1 The importance of asymmetries and anisotropies in spatial econometrics

The autoregressive models that we have discussed in the previous chapters 3 and 4 have a common characteristic. They are based on the

general definition of spatially lagged variable given in section 2.1 where we defined the spatially lagged value of, say, variable y in location i as:

$$L(y_i) = \sum_{j=1}^{n} w_{ij}^{*} y_j \tag{5.26}$$

w_{ij}^{*} being the standardized weight matrix (see Equation (2.3)). In other terms, the spatial lag is defined as the average of the value y_j observed in all the locations that are neighbors to location i. This expression implies that the *direction does not matter* in that all the neighbors (whatever their position with respect to location i) contribute in the same way to the lagged variable. This hypothesis, implicitly assumed by all the models presented so far, is known in the literature as *isotropy* (for a formal definition, see, for example, Cressie, 1993).

The concept of isotropy derives from physics and implies that the dependence structure does not present *directional biases* or *preferred directions*. The hypothesis of isotropy is commonly discussed in many branches of spatial statistics. For instance, when dealing with spatial data in meteorology, geology or in other physical phenomena (where the direction is of paramount importance; see, for example, Schabenberger and Gotway, 2002) or in medical imaging applications where, for instance, anisotropy is proved to be a good predictor of breast cancer risk (Heine and Mahorta, 2002).

Economists have always been aware of this problem. For instance the Nobel laureate, Clive Granger (1974), noted that the assumption of absence of directional biases is particularly strong and unrealistic in spatial econometrics. He states that "if direction did not matter, the degree to which these variables were related would depend only on the distance between the points. The relationship between values measured at Oxford and London will be the same as between values measured at two Lincolnshire villages 55 or so miles apart. The correlation between unemployment figures in New York and Philadelphia will be the same as between two small mid-Western towns roughly a hundred miles apart. This assumption of stationarity on the plane is completely unrealistic for economic variables" (Granger, 1974, p. 15).

A particularly common manifestation of anisotropy is the asymmetry of spatial relationships. Indeed, in the founding book of the discipline, Paelinck and Klaassen (1979) identify asymmetry as one of the five fundamental characteristics of spatial econometrics. A good example of

asymmetry is represented by the *core-periphery* model (see, for example, Paelinck and Nijkamp, 1975) which assigns stronger dependence from the center to the periphery than on the reverse direction. Directional biases can be observed in many empirical circumstances, such as in the dynamic pattern of house prices (Holly et al., 2010), in the many economic variables observed in the United States with reference to dependencies along the coasts (North–South) as opposed to that from the coasts to the internal states (East–West), or in the EU in the different dependencies between the center and the periphery and vice versa, only to suggest a few examples.

Even if asymmetry is the most intuitive aspect of anisotropy, it is only one of its many manifestations. To clarify the relationship between the asymmetry and anisotropy concepts, let us consider a study area divided into only three spatial regions (called *i*, *l* and *m*) arranged in a regular, square lattice grid as in Figure 5.1, and consider, without loss of generality, the hypothesis that the topology of the system can be captured by a simple rook's neighboring structure.

Let us now define in general terms two different dependence structures among the regions along the vertical (*V)* and the horizontal (*H*) directions respectively. According to the rook's definition of neighbors, we have the following system of equations:

$$\begin{aligned} y_i &= \vec{\rho}_H y_m + \vec{\rho}_V y_l + \varepsilon_i \\ y_l &= \overleftarrow{\rho}_V y_i + \varepsilon_l \\ y_m &= \overleftarrow{\rho}_H y_i + \varepsilon_m \end{aligned} \tag{5.27}$$

with $\vec{\rho}_H$ and $\vec{\rho}_V$ representing the spatial dependence parameters along the horizontal and, respectively, the vertical line, and the superscript arrows indicating the direction of such dependence. Equation (5.27) express the fact that location *i* depends horizontally (left-to-right) on location *m* and vertically (top-to-bottom) on location *l*; location *l* depends vertically (bottom-to-top) on location *i* and location *m* depends horizontally (right-to-left) on location *i*. With reference to this simple situation, asymmetry implies that $\vec{\rho}_V \neq \overleftarrow{\rho}_V$ and $\vec{\rho}_H \neq \overleftarrow{\rho}_H$

	l
m	i

Figure 5.1 Three regions arranged on a regular lattice grid

whereas anisotropy (still in the presence of symmetry) implies that $\vec{\rho}_V = \overleftarrow{\rho}_V = \rho_V \neq \rho_H = \overleftarrow{\rho}_H = \vec{\rho}_H$. Thus asymmetry implies anisotropy, but not vice-versa. Furthermore, isotropy implies symmetry, but symmetry does not necessarily imply isotropy.

5.3.2 Testing isotropy in Spatial Lag models

In order to illustrate a testing procedure for the hypothesis of isotropy, let us consider n observations of a variable y and a (non row-standardized) weight matrix W (however it is defined).

Let us also define two different non-overlapping weight matrices W_1 and W_2 referring to two different directions and such that $W_1 + W_2 = W$. The two matrices incorporate two different preferred directions of spatial dependence. For instance in a regular square lattice grid, we can assume a rook's case definition of neighbors and two matrices describing the topology shown in Figure 5.2.

Let us now define a two-parameter, anisotropic Spatial Lag model in which dependence develops along the two different specified directions:

$$y = Z\beta + \lambda_1 W_1 y + \lambda_2 W_2 y + u \qquad u \mid Z \approx i.i.d.N(0,\sigma^2 I) \tag{5.28}$$

In Equation (5.28) $Z = [X, WX]$ is a non-stochastic matrix of independent variables that may include lagged terms, W_1 is a (non row-standardized) weight matrix that incorporates dependency in one direction, W the corresponding full (non row-standardized) weight matrix derived under the assumption of isotropy, and $W_2 = W - W_1$ represents the second direction of dependency. If $\lambda_1 = \lambda_2 = \lambda$, model (5.28) reduces to the familiar one-parameter isotropic Spatial Lag Model (see Equation (3.54)) defined as:

$$y = Z\beta + \lambda W y + u \qquad u \mid Z \approx i.i.d.N(0,\sigma^2 I) \tag{5.29}$$

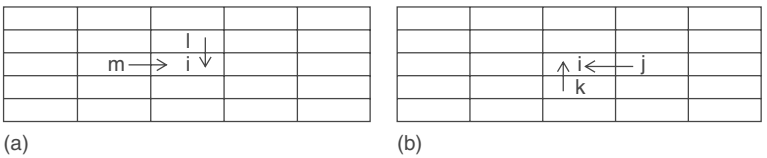


Figure 5.2 Two different W matrices incorporating two different directions of spatial dependence. (a) W_1 = north-west to south-east dependence; (b) W_2 = south-east to north-west dependence. Cell i depends only on cells l and m following the W_1 weighting matrix, and only on cells j and k following the W_2 weighting matrix.

Let $u^{(ANI)} = y - Z\beta - \lambda_1 W_1 y - \lambda_2 W_2 y$ and $u^{(ISO)} = y - Z\beta - \lambda W y$ be the disturbances in the two competing models. The log-likelihood functions for the anisotropic and isotropic models are then respectively given by:

$$l_{ANI}(\sigma^2, \lambda_1, \lambda_2, \beta) = const - \frac{1}{2}\ln\left|I - \lambda_1 W_1 - \lambda_2 W_2\right| - \frac{1}{2\sigma^2} u^{(ANI)^T} u^{(ANI)} \tag{5.30}$$

and

$$l_{ISO}(\sigma^2, \lambda, \beta) = const - \frac{1}{2}\ln\left|I - \lambda W\right| - \frac{1}{2\sigma^2} u^{(ISO)^T} u^{(ISO)} \tag{5.31}$$

Given these definitions, Arbia et al. (2013) developed a test of isotropy by considering the Likelihood Ratio test statistics (see Equation (1.30)):

$$A - test = -2\left[l_{ANI}(\hat{\sigma}^2, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\beta}) - l_{ISO}(\hat{\sigma}^2, \hat{\lambda}, \hat{\beta})\right] \tag{5.32}$$

where $l_{ANI}(\hat{\sigma}^2, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\beta})$ is the maximized log-likelihood of the unrestricted model and $l_{ISO}(\hat{\sigma}^2, \hat{\lambda}, \hat{\beta})$ is the maximized log-likelihood of the restricted model. In both expressions, the unknown parameters are replaced by the corresponding *ML* estimators, denoted by the hat. Standard asymptotic theory guarantees that, under the null hypothesis of isotropy, the test statistics converges in distribution to the χ^2_1 random variable:

$$A - test \xrightarrow{d} \chi^2_1 \tag{5.33}$$

In Arbia et al. (2013) the authors report the results of a Monte Carlo study which shows that neglecting the problem of isotropy can lead to serious biases and losses of efficiency in the estimation of the

regression parameters in spatial econometric models. In particular, they showed that in wrongly identified models (that is, when we assume isotropy when this hypothesis is not true), if the absolute difference between λ_1 and λ_2 is large (and therefore we are very distant from the hypothesis of isotropy) both the relative bias and the standard errors of the estimates are larger. Furthermore, while the *ML* estimators still obviously enjoy the usual asymptotic properties, the rate of convergence is slower.

Example 5.2 Anisotropies in the Barro and Sala-i-Martin convergence model

Let us consider again the Barro and Sala-i-Martin model of regional convergence introduced in Example 1.1. In their paper Arbia et al. (2013) focus on the detection of possible directional biases in the pattern of spatial dependence in regional convergence modeling. In particular, they consider the following anisotropic Spatial Lag specification of the traditional model:

$$y = Z\beta + \lambda_1 W_1 y + \lambda_2 W_2 y + u \qquad u \mid Z \approx i.i.d.N(0,\sigma^2 I)$$

with $y = \ln \frac{X_{it}}{X_{i0}}$; $Z_i = X_{i0}$, X_{it} *the per capita income in region i at time t and u the error component.*

The empirical data considered referred to the per capita income in the 20 Italian regions already considered in Example 1.1 in a period of time ranging from year 2000 to year 2008. To identify a significant directional bias the authors defined 20 different W_1 *matrices by considering in turn each region as the one originating the directional bias and the other regions following in a time-like fashion satisfying the contiguity constraint. So, for instance, the* W_1 *matrix referring to the region Lombardy as originating the directional bias is built up according to the scheme shown in the next figure. Thus in the graph, Lombardy is coded as "region 1" and does not depend on any other region, the 4 regions coded as "region 2" depend only on Lombardy, the 6 regions coded as "region 3" depend only on the regions coded as "region 2" and so on.*

Source: Author's creation using software "R" and shapefiles available on ISTAT that can be downloaded at the webpage http://www.istat.it/it/archivio/44523
Note: Anisotropic spatial relationship originating from the region Lombardy. Each region depends only on those regions with a lower code number. Region 1 (Lombardy) is the one where the directional bias originates and does not depend on any other regions.

The region originating the directional bias plays a role analogous to the initial observations in a time series. The W_2 matrices are then derived as $W_2 = W - W_1$ in order to satisfy the requirements of the test. The A-test was then calculated in the 20 different specifications. The results of this analysis are shown in the following table:

Region	***Anisotropy***	***p-value***
	A-test	
Lombardy	***8.894***	***0.002***
Friuli Venezia-Giulia	*8.502*	*0.003*
Trentino Alto Adige	*9.483*	*0.003*
Tuscany	*6.736*	*0.009*
Aosta Valley	*6.437*	*0.011*
Liguria	*5.666*	*0.017*

(*continued*)

Continued

Region	***Anisotropy***	***p-value***
	A-test	
Umbria	*2.793*	*0.094*
Marche	*2.772*	*0.095*
Emilia Romagna	*1.301*	*0.254*
Sardinia	*1.189*	*0.275*
Campania	*1.032*	*0.309*
Calabria	*0.719*	*0.396*
Latium	*0.497*	*0.480*
Veneto	*0.337*	*0.561*
Basilicata	*0.316*	*0.573*
Puglia	*0.315*	*0.574*
Sicily	*0.207*	*0.648*
Abruzzo	*0.065*	*0.798*
Molise	*0.003*	*0.954*
Piedmont	*0.002*	*0.956*

(Numbers in italics refer to the detection of significant anisotropy).

The analysis reveals that there are significant anisotropies in regional growth and highlights significant directional biases originating from six regions, namely Lombardy, Friuli, Trentino-Alto Adige, Tuscany, Aosta Valley and Liguria. The largest significance of the A-test is obtained for Lombardy (the region of Milan), which therefore appears as a leading region in the process of growth diffusion. This result is consistent with common knowledge and with the empirical evidence since Lombardy is the region with the highest concentration of industrial activity in Italy.

5.3.3 Inference for a unilateral Spatial Lag model

If the hypothesis of isotropy is rejected using the *A*-test, then appropriate anisotropic models of the kind presented in the previous section have to be considered. Conversely, if this hypothesis can be accepted on the basis of empirical data, then this property can be exploited to simplify the calculations and to circumvent some of the computational difficulties connected with the use of Maximum Likelihood in conventional spatial econometric modeling. This aim can be achieved by considering the so-called *unilateral approximations* introduced in the literature by Besag (1974). Indeed, if a model is isotropic, then direction does not matter and we can approximate its properties by specifying a *unilateral model* which is simpler to treat while providing the same inferential information. In order to identify such a model, we first have to select one conventional preferred direction and then build

up a time-like structure of dependence for the variables involved. The choice can be purely subjective since direction does not matter. In this case, similar to what happens in time series analysis, the likelihood of the model can be easily factorized as the product of the conditional densities of each of the variables, conditional upon a set of variables termed the *predecessors-neighbors* (denoted as *PN*, see Besag, 1974 and Arbia, 2006) defined as the neighbors along one designated preferred direction (for instance, one of the two directions described in Figure 5.2). As previously mentioned, if the model is isotropic, this procedure is inferentially equivalent to a full likelihood approach. To introduce this method let us consider the following *unilateral spatial lag model*:

$$y = Z\beta + \lambda_1 W_1 y + \lambda_2 W_2 y + u \tag{5.34}$$

where y is a vector of observations of the dependent variable, X a matrix of non-stochastic exogenous regressors; β a vector of unknown parameters to be estimated and u a vector of stochastic disturbances such that $u \mid Z \approx i.i.d.N(0,\sigma^2 I)$. Furthermore let W_1 be the (non row-standardized) associated unilateral weight matrix such that $w_{ij} \in W_1$; $w_{ij} = 1$ if $i \in PN(j)$ and 0 otherwise. In other words, w_{ij} =1 if j is a neighbor of i in the preferred direction chosen. In this case, referring, without loss of generality, to the scheme of three locations shown in Figure 5.1, and to the case of one single regressor, the likelihood of the unilateral model can be expressed as:

$$L(\sigma^2,\lambda,\beta) = \prod_{i=1}^{n} f_{y_i|y_l,y_m}(y_i \mid y_l, y_m) \tag{5.35}$$

where $l,m \in PN(i)$ so that each density in the right-hand side is only conditional on the predecessors-neighbors.

Using the standard results on multinormal conditional distributions (see, for example, Anderson, 2003) and assuming no cross-covariance between the variables Z and y in different locations (that is, $Cov(Z_i y_l) = Cov(Z_i y_m) = 0$), the generic expression for the conditional density function on the right-hand side of (5.35) will be $N(\mu_c,\sigma_c^2)$ with conditional mean and variances given by:

$$\begin{aligned} \mu_c &= \mu + \frac{\lambda}{(1+\lambda')}\left[\sum_{j=1}^{n} w_{ij}(y_j - \mu)\right] + \frac{\beta}{(1-\lambda'^2)}(z_i - \mu_z) \\ \sigma_c^2 &= \sigma^2 - \frac{2\lambda^2}{(1-\lambda'^2)} - \frac{\beta^2\sigma_z^2}{(1-\lambda'^2)} \end{aligned} \tag{5.36}$$

where λ' denotes the correlation between y_l and y_m, which is the correlation between the predecessors of y_i which are mutually second-order neighbors (see section 5.2.1).

Finally, the likelihood associated with the distribution (5.36) can be expressed as:

$$\begin{aligned} L(\sigma^2,\lambda,\lambda',\beta) &= \prod_{i=1}^{n} f_{y_i|y_l,y_m}(y_i \mid y_l, y_m) \\ &= const\left(\sigma_c^2\right)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_c^2}\left[\sum_{i=1}^{n}\left(y_i-\mu_c\right)^2\right]\right\} \end{aligned} \tag{5.37}$$

and the log-likelihood as:

$$l(\sigma^2,\lambda,\lambda',\beta) = const - \frac{n}{2}\ln\left(\sigma_c^2\right) - \frac{1}{2\sigma_c^2}\left[\sum_{i=1}^{n}\left(y_i-\mu_c\right)^2\right]$$

and, substituting Equations (5.36) into (5.37), we finally have:

$$\begin{aligned} l(\sigma^2,\lambda,\lambda',\beta) = const &- \frac{n}{2}\ln\left[\sigma^2 - \frac{2\lambda^2}{(1-\lambda'^2)} - \frac{\beta^2\sigma_z^2}{(1-\lambda'^2)}\right] \\ &- \frac{1-\lambda'^2}{2\left[\sigma^2(1-\lambda'^2) - 2\lambda^2(1-\lambda') - \beta^2\sigma_z^2\right]} \\ &\times\left[\sum_{i=1}^{n}\left(y_i-\mu+\frac{\lambda}{(1+\lambda')}\left[\sum_{j=1}^{n} w_{ij}(y_j-\mu)\right]\right.\right. \\ &\left.\left.+\frac{\beta}{(1-\lambda'^2)}(z_i-\mu_z)\right)^2\right] \end{aligned} \tag{5.38}$$

This expression does not contain any matrix inversion and so its calculation can be executed straightforwardly, even with a very large sample size.

Example 5.3 Health planning in Mexico (continued)

In this example, we wish to illustrate the performances of the unilateral approximation with respect to a full likelihood approach. We will also compare the results with those obtained with the MESS specification discussed in section 5.1. For this reason, we will examine again the dataset on health data in the 32 Mexican states. Since in an isotropic model the direction of proximity does

	Maximum Likelihood			*MESS Specification*			*Unilateral Approximation*		
	Estimated value	***Standard error***	***p-value***	***Estimated value***	***Standard error***	***p-value***	***Estimated value***	***Standard error***	***p-value***
β_0	*0.139859*	*0.446938*	*0.75434*	*0.124292*	*0.200922*	*0.54136*	*0.036752*	*0.229270*	*0.872646*
β_1	*0.186960*	*0.012195*	*0.000****	*0.187352*	*0.012738*	*0.000****	*0.184956*	*0.011789*	*0.000****
β_2	*0.058434*	*0.031370*	*0.06250*	*0.058685*	*0.033965*	*0.09545*	*0.050829*	*0.031676*	*0.108572*
β_3	*–0.049681*	*0.034338*	*0.14794*	*–0.049863*	*0.037131*	*0.19049*	*–0.047028*	*0.034094*	*0.167774*
β_4	*0.064011*	*0.025163*	*0.01096**	*0.064794*	*0.024870*	*0.01475**	*0.066109*	*0.023256*	*0.004473****
λ	*–0.14109*	*0.19208*	*0.46263*	*–0.136520*	*0.16331*	*0.43326*	*–0.069582*	*0.046159*	*0.1317*

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

not matter, a particular interpretation of the unilateral approximation method is to consider a neighborhood in which each spatial unit has only one neighbor, defined according to the nearest neighbor criterion (see Example 2.1 in section 2.1). The model presented in Example 5.1 is thus estimated again using this particular specification of the unilateral model. The results are shown in the following table where, to facilitate the comparison, we also display again the results of the full likelihood and that of the MESS approximation.

The three methods lead to remarkably similar results in terms of the inferential conclusions associated with the parameters, all agreeing in identifying the health staff and the percentage of elderly people (variables x_1 and x_4) as significant variables. The sign of the estimates are always the same and the absolute values are very similar using the three methods, with the only exception being the intercept and the dependence parameter λ which is underestimated by the unilateral approximation more than by the MESS specification. As already remarked in Example 5.1 the differences in terms of the computing time and storage required may become a very relevant factor in choosing one of the two approximated techniques in very large samples.

5.4 A composite likelihood approach

5.4.1 Generalities

In this section we will consider an inferential approach based on a particular form of composite likelihood termed *pairwise likelihood* (Lindsey, 1988; Varin et al., 2011), a method that is becoming increasingly popular in the statistical literature as a viable solution to those cases where a joint likelihood approach is computationally unfeasible. A composite likelihood (a subclass of the more general *pseudo-likelihood,* see Pace and Salvan, 1997) is defined as a function that, although not being a fully specified likelihood, enjoys some of its properties. The literature reports some examples of composite likelihood inference in time series econometrics (Davis and Yau, 2011) and in spatial econometrics (see Arbia, 2012 for spatial cross-sectional models and Wang et al., 2013 for the spatial Probit models).

5.4.2 A bivariate Marginal Likelihood approach to Spatial Error Model estimation

Let us consider a linear regression model:

$$y_i = \beta Z_i + \varepsilon_i \tag{5.39}$$

i=1…n, as in model (3.13), but now, instead of specifying the error in autoregressive form as in (3.14), let us assume the following joint bivariate Gaussian distribution for each pair of disturbances, say *i* and *l*:

$$\begin{pmatrix}\varepsilon_i \\ \varepsilon_l\end{pmatrix} \approx MVN\left(0\mathbf{1}, \sigma^2\Omega\right) \quad \forall i,l \;\; l \in N(i) \qquad \Omega = I \;\; \forall \; l \notin N(i) \tag{5.40}$$

This definition is very general and similar to the one adopted in Equation (5.20) in section 5.2.2; however, here we do not aim to model simultaneously the joint random behavior of all *n* disturbances as in the conventional spatial econometric models and in the *MESS* approximation, but only their *2*-by-*2* relationships. Notice that $N(i)$, as usual, represents the set of neighbors of location *i* and that the choice of the neighborhood criterion is not essential to the method. In Equation (5.40), $\Omega = \begin{pmatrix}1 & \psi \\ \psi & 1\end{pmatrix}$ represents the correlation matrix and ψ is a parameter controlling for the error spatial correlation such that $\psi \in (-1;+1)$. The symbol ψ is used intentionally in place of the more commonly used symbol ρ in a Spatial Error context to highlight the differences in their meaning. The (apparently restrictive) hypothesis that the correlation between all units in a neighborhood is the same no matter what the direction, derives from the assumption of isotropy (implicitly) made in all the spatial econometrics literature (see section 5.3).

Before introducing a composite *ML* procedure for the estimation of the unknown parameters of model (5.39)–(5.40), let us first introduce the definition of a *bivariate coding*. This concept was introduced by Arbia (2012, 2014) extending the work of Besag (1974). Just for the purposes of illustration, let us assume that the *n* observations are available on a regular square lattice grid and let us label the interior cells of such a grid with a cross "x" as indicated in Figure 5.3.

	x	x		x	x		x	x	
	x	x		x	x		x	x	

Figure 5.3 Bivariate coding pattern in a regular square lattice grid

In particular, let us code with a cross, a subset of q of the n available spatial observations, ($q \in Q$) and let us also code with a cross, a set of q further locations chosen randomly in the neighborhood of location i. The disturbances ε_i and ε_l, with $l \in N(i)$, are assumed to be spatially dependent due to their proximity while the pairs $\{\varepsilon_i, \varepsilon_l\}$ and $\{\varepsilon_j, \varepsilon_k\}$ (with $k \in N(j)$) are assumed to be stochastically independent provided $j, k \notin N(i,l)$, having defined $N(i,l) = \{N(i) \cup N(l)\}$ as the joint neighborhood of i and l. The above classification scheme defines a bivariate coding pattern.

Similar coding schemes can be easily introduced in irregular spatial schemes with an appropriate definition of neighborhood. The aim of this procedure is rather clear: by selecting pairs of observations following the criterion of including two neighboring spatial units, we are able to retain the spatial information contained in the sample. However, by selecting pairs of spatial units that are, by definition, independent of one another we avoid incurring problems in the estimation, typical of the spatial econometric models introduced in Chapter 3. The hypothesis of independent pairs, which may seem too restrictive at a first sight, has to be considered with reference to the definition of a neighborhood. Indeed, we can always define a neighborhood in such a way that those pairs that do not belong to the neighborhood are far enough to be independent. Finally, considering pairs of random disturbances is not essential to the method and one could equally consider triplets and higher-order groups, the only theoretical justification for this choice being the Hammersley–Clifford theorem and the consequent restriction to pairwise interaction assumed by the models discussed by Besag (1974).

Under these assumptions, we have that the joint density of any of the q pairs of disturbances included in the bivariate coding is given by:

$$f_{\varepsilon_i \varepsilon_l}(\varepsilon_i \varepsilon_l) = \frac{1}{2\pi\sigma^2\sqrt{1-\psi^2}} \exp\left\{-\frac{1}{2\sigma^2(1-\psi^2)}\left[\varepsilon_i^2 - 2\psi\varepsilon_i\varepsilon_l + \varepsilon_l^2\right]\right\} \quad (5.41)$$

$$\text{if } l \in N(i),\ i=1,\ldots,q$$

A composite likelihood can then be derived as the product of the q bivariate density functions reported in Equation (5.41) that are assumed to be independent. We can then treat such a composite likelihood as a

proper likelihood and base our inference on it. In particular, multiplying the q Equations (5.41), we have:

$$\begin{aligned} L(\beta,\sigma^2,\psi) &= \prod_{i=1}^{q} f_{\varepsilon_i\varepsilon_l}(\varepsilon_i\varepsilon_l) \\ &= \prod_{i=1}^{q} \frac{1}{2\pi\sigma^2\sqrt{1-\psi^2}} \exp\left\{-\frac{1}{2\sigma^2(1-\psi^2)}\left[\varepsilon_i^2 - 2\psi\varepsilon_i\varepsilon_l + \varepsilon_l^2\right]\right\} \\ &= (2\pi)^{-q}(\sigma^2)^{-q}(1-\psi^2)^{-\frac{q}{2}} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2(1-\psi^2)}\sum_{i=1}^{q}\left[\varepsilon_i^2 - 2\psi\varepsilon_i\varepsilon_l + \varepsilon_l^2\right]\right\} \end{aligned} \tag{5.42}$$

and correspondently the log-likelihood is equal to:

$$\begin{aligned} l(\beta,\sigma^2,\psi) = const - q\ln(\sigma^2) - \frac{q}{2}\ln(1-\psi^2) \\ - \frac{1}{2\sigma^2(1-\psi^2)}\sum_{i=1}^{q}\left[\varepsilon_i^2 - 2\psi\varepsilon_i\varepsilon_l + \varepsilon_l^2\right] \end{aligned} \tag{5.43}$$

Let us now introduce some new symbols. Let a_1, a_2, a_3, a_4, a_5 and a_6, be a set of statistics defined as follows:

$$\begin{aligned} a_1 &= \sum_{i=1}^{q} Z_i^2 + \sum_{l=1}^{q} Z_l^2 = \sum_{j=1}^{2q} Z_j^2 & a_2 &= \sum_{i=1}^{q} y_i^2 + \sum_{l=1}^{q} y_l^2 = \sum_{j=1}^{2q} y_j^2 \\ a_3 &= \sum_{i=1}^{q} Z_i y_i + \sum_{l=1}^{q} Z_l y_l = \sum_{j=1}^{2q} Z_j y_j & a_4 &= \sum_{i=1}^{q} Z_i y_l + \sum_{l=1}^{q} Z_l y_i \\ a_5 &= \sum_{i=1}^{q} Z_i Z_l & a_6 &= \sum_{i=1}^{q} y_i y_l \end{aligned} \tag{5.44}$$

Notice that in the above definitions, for notational convenience, with the expression $\sum_{i=1}^{q} Z_i Z_l$ we indicate for brevity $\sum_{i=1}^{n}\sum_{l\in N(i)} Z_i Z_l$ which is the sum of products between the two observations belonging

to the same pair (and similarly for the other expressions in Equations (5.44)). We will consistently employ this notational simplification. Arbia (2014) has shown that, under the assumptions of the model, Equation (5.43) admits a maximum, that this maximum is unique and that it is achieved in the points obtained as the solution of the following system of non-linear equations:

$$\begin{cases} \hat{\beta}_{BML} = \dfrac{a_3 - \hat{\psi}_{BML} a_4}{a_1 - 2\hat{\psi}_{BML} a_5} & (5.45) \\ \hat{\sigma}_{BML}^2 = \dfrac{\sum_{i=1}^{q} \left(\varepsilon_i^2 - 2\hat{\psi}_{BML} \varepsilon_i \varepsilon_l + \varepsilon_l^2\right)}{2q(1 - \hat{\psi}_{BML}^2)} \\ \qquad = \dfrac{a_2 + \hat{\beta}_{BML}^2 a_1 - 2\hat{\beta}_{BML} a_3 - 2\hat{\psi}_{BML} a_6 - 2\hat{\psi}_{BML} \hat{\beta}_{BML}^2 a_5 + 2\hat{\psi}_{BML} \hat{\beta}_{BML} a_4}{2q\left(1 - \hat{\psi}_{BML}^2\right)} & (5.46) \\ \hat{\psi}_{BML} = \dfrac{\sum_{i=1}^{q} \varepsilon_i \varepsilon_l}{q\hat{\sigma}_{BML}^2} = \dfrac{a_6 - \hat{\beta}_{BML} a_4 + \hat{\beta}_{BML}^2 a_5}{q\hat{\sigma}_{BML}^2} & (5.47) \end{cases}$$

Arbia (2012) termed this estimator the Bivariate marginal Maximum Likelihood estimator (or *BML*).

Notice that, if $\psi = 0$ (case of pairwise bivariate spatial independence of the regression disturbances), in Equation (5.45) we have:

$$\hat{\beta}_{BML} = \frac{\psi a_4 - a_3}{2\psi a_5 - a_1} = \frac{a_3}{a_1} = \frac{\sum_{j=1}^{2q} x_j y_j}{\sum_{j=1}^{2q} x_j^2} = \hat{\beta}_{ML} \tag{5.48}$$

whereas in Equation (5.46) we have:

$$\hat{\sigma}_{BML}^2 = \frac{1}{2q} \sum_{i=1}^{q} \left[\varepsilon_i^2 + \varepsilon_l^2\right] = \frac{1}{2q} \sum_{j=1}^{2q} \varepsilon_j^2 = \hat{\sigma}_{ML}^2 \tag{5.49}$$

so that the solutions correspond to the familiar *ML* estimators of, respectively, β and σ^2 in the case of independent errors (see Chapter 1). Notice also that the estimator $\hat{\psi}_{BML}$ derived in Equation (5.47)

corresponds to the intuitive estimator of the spatial correlation among the disturbances.

In the quoted paper, Arbia (2014) proves that the BML estimators are normally distributed, unbiased in small samples and weakly consistent. The exact Fisher Information matrix and the exact standard errors of the estimators are also derived to be used in confidence interval estimation and hypothesis testing. Apart from the analytical and computational advantages, the *BML* estimators also provide some interesting interpretative advantages. In fact, consider again the formal expression of the *BML* estimator of the regression parameter β derived in Equation (5.45): $\hat{\beta} = \frac{a_3 - \psi a_4}{a_1 - 2\psi a_5}$. The numerator in this expression represents the covariance between Z and y (the term a_3) augmented with the extra term $-\psi a_4$ which represents the spatial spillover of the variable Z in one location onto the variable y in a neighboring location belonging to the same pair (the term a_4), weighted with the spatial correlation parameter ψ. Similarly, the denominator represents the variance of the independent variable (the term a_1) augmented with the term $-2\psi a_5$ representing the spatial autocovariance of variable Z (the term a_5), weighted with the spatial correlation of the error term. The interpretation is rather straightforward. In the case of positive error spatial correlation ($\psi > 0$), if the spatial spillover between Z and y and the spatial autovariance of Z are of different sign, we can monitor the multiplicative effect of a change in variable Z on variable y. In particular, if the spatial spillover (a_4) is negative and the spatial autocovariance of Z (a_5) is positive, the multiplicative effect will be emphasized. The formal expression of the *BML* estimator $\hat{\beta}$ also shows that, in the presence of a strong positive spatial correlation of the independent variable, the effect on y of a variation in the independent variable is more pronounced in the presence of a positive spillover between the two variables. This result has an intuitive explanation. In fact, one location benefits not only from an increase of Z in the same location, but also for the increase of Z in the neighboring locations. Similarly, the formal expression of the *BML* estimator of β (Equation (5.45)) also shows that we can have a higher impact of the independent variable on the dependent variable even if there is no spatial spillover between the two. In fact, when $a_4 = 0$, the effect on y of a variation in the independent variable will be more pronounced if $a_5 > 0$, that is, in the presence of a positive spatial correlation in the independent variable.

As already mentioned, given that the exact Fisher Information matrix can be formally derived, standard likelihood-based hypothesis testing

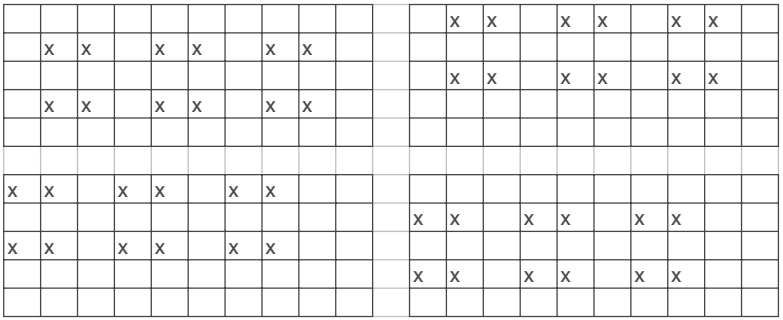


Figure 5.4 Four different bivariate coding schemes for the data in Figure 5.3

procedures can be applied to this new specification. However, this framework also allows a further approach to standard error evaluation and to hypothesis testing based on the idea of resampling. Resampling methods for sets of dependent random variables have a long tradition in statistics dating back to the earlier contributions of Solow (1985), Künsch (1989), Arbia (1990) and Sherman (1996). The bivariate coding technique presented in this section is based on the identification of a subsample of pairs of units. However, in any given empirical situation the bivariate coding scheme is non-unique. Even in the very simple example shown in Figure 5.3 we can, in fact, produce four different codings (and consequently four different estimations of the model's parameters) as shown in Figure 5.4.

The number of possible configurations could be even larger when dealing with irregular spatial data such as those encountered in practical applications. Dealing with very large sample size, we can therefore derive many possible bivariate coding schemes, and correspondently many different estimates of the parameters, allowing the derivation of a resampling distribution. In this respect, the *bivariate coding* approach suggests a formal way of bootstrapping spatial data in a regression context preserving the condition of independence between the subsamples while not destroying the features of spatial dependence of the data.

Example 5.4 Monte Carlo assessment of BML estimation method

Arbia (2014) evaluate the performances of the BML method by examining both artificially generated data and real datasets. In particular, in Arbia (2014), the author shows, by means of some Monte Carlo experiments, that the BML estimates are very accurate while the computing time is negligible: with an Intel

core i7 processor working at 2.7 GHz, the time required for the calculations ranges from between 0.9 milliseconds when the sample size is n = 100 to 831 milliseconds when n = 2500.

5.5 R Codes

When dealing with the **MESS specification**, the R library `Matrix`, contains the function `expm` which allows computation of the exponential of a matrix (see Equation (5.5)). First of all, as usual, we need to download the package with the command `install.packages("Matrix")` and then we need to call it back in each new session with the command `library(Matrix)`. The exponential of a matrix A, say, can be executed through the command

```
>expm(A)
```

with A a real positive definite matrix. Even if the exponential of a matrix is defined as the infinite Taylor series, it is approximated using Ward's diagonal Padé approximation (Moler and Van Loan, 2003).

The R package `spdep` also contains a routine for estimating a Spatial Lag Model using the *MESS* specification. The command is very simple and closely mirrors the command for the corresponding Spatial Lag model presented in section 3.8. The syntax of the command is the following:

```
> model1<-lagmess(formula=y ~ X+Y, data=filename, listw=W)
```

with W a weight matrix.

For the **unilateral approximation**, once a likelihood is specified explicitly (such as, for example, in Equation (5.38) for a unilateral Spatial Lag model), only a routine is necessary to maximize it with respect to the parameters. To this end, we can use the specific R library `maxLik` which can be installed by typing the command `install.packages("maxLik")` the first time and then, at the beginning of each session, the command `library (maxLik)`

Once the library is installed, we have to write a log-likelihood function (call it, for example, `loglik`) as a function of a set of parameters (say `beta, sigma, lambda, rho`, etc.). The R command for likelihood maximization is:

```
> mle <- maxLik(logLik = logLikFun, start = c(beta = 0,
sigma = 1, lambda=0.2, rho=0.1))
```

with `beta` = 0, `sigma` = 1, `lambda` = 0.2, `rho` = 0.1 arbitrary initial values for the numerical search of the maximum and `mle` a conventional name for the model. The results of the maximization procedure can be then displayed by typing:

```
> summary(mle)
```

Furthermore, one possible interpretation of the unilateral approximation is to consider, for each unit, only its first nearest neighbor (see Example 2.1) in the definition of the *W* matrix (see Example 5.3). In this case, having in hand a shapefile (say `poly`) we start deriving the coordinates of the centroids (as we have illustrated in section 2.3.3):

```
coords<-coordinates(poly)
```

then we identify for each unit the first nearest neighbor through the command:

```
knn<-knearneigh(coords, k=1)
```

Notice that the command can be generalized to the search of a number *k* of neighbors. In the next step, we derive a system of neighborhood on the basis of our definition of neighborhood through the command:

```
nn<-knn2nb(knn)
```

and, finally, we derive the *W* matrix, as usual, through the command:

```
W<-nb2listw(nn)
```

Applying the usual commands for inference on a spatial model will lead to a unilateral approximation which will produce the output in a much shorter time.

No specific routine is required for the **bivariate coding approach** in the estimation of a Spatial Error Model. To implement the procedure, we only need a program to randomly select pairs of non-adjacent spatial units and then use the information contained in this subsample to calculate the closed-form solutions of the estimators reported in Equation (5.45) to (5.47).

Key Terms and Concepts Introduced

- Big data
- Sparse and dense weight matrices
- Computing time and storage limitations
- Matrix exponential transformation
- First-order neighbors
- Higher-order neighbors
- Matrix power expansion
- Matrix exponential spatial specification
- Isotropy and anisotropy
- Directional bias of spatial relationships
- Asymmetric spatial relationships
- Core–periphery models
- Test of isotropy
- Anisotropic spatial lag model
- Unilateral approximation
- Unilateral spatial lag model
- Predecessors-neighbors
- Composite likelihood
- Pairwise likelihood
- Pseudo-likelihood
- Bivariate coding technique
- Bivariate Marginal Maximum Likelihood estimator
- Spatial resampling and spatial bootstrap

Questions

1. What problems emerge when using a Maximum Likelihood approach with very large sample size?
2. Why are the problems more dramatic in the case of a dense $\boldsymbol{W}$ matrix and, conversely, less dramatic when considering a sparse $\boldsymbol{W}$ matrix?
3. What are the advantages of using the MESS specification of a Spatial Lag model with respect to a Maximum Likelihood estimation approach?
4. Illustrate the differences between the concept of asymmetry and the concept of anisotropy.
5. Given the computational advantages, what is the major pitfall of using the Bivariate Marginal Maximum Likelihood approach with respect to the full Maximum Likelihood approach?

Exercises

Exercise 5.1 Consider again the boundaries of the eight Romanian NUTS 2 regions reported in Exercise 2.1 and derive the second-oder and the third-order W matrices, say W^2 and W^3. Using these matrices, derive the transformation (5.11) $S = e^{a\,W}$ and develop the approximation $S \approx \sum_{i=0}^{q} \frac{a^{\,i} W^i}{i!}$ for $\lambda = 0.9502$, the spatial correlation parameter and $q = 3$.

Exercise 5.2 Consider the case of an isotropic model and consider a neighborhood in which each spatial unit has only one neighbor which is not a neighbor of any other unit. Derive the $\boldsymbol{W}$ matrix in this hypothesis for a sample of dimension n. Derive also the matrix $(I - \lambda W)$ and its inverse. Recall the formal expression of the variance-covariance matrix Ω of a unilateral Spatial Lag model and derive its determinant $|\Omega|$.

Exercise 5.3 Given the data set used in Examples 3.3 and 3.4 and in Exercises 3.6 and 4.8, and using a row-standardized W matrix, re-estimate a Spatial Lag model using the MESS specification. On the same datasets define a unilateral row-standardized W matrix based on the first nearest neighbor and estimate again a unilateral Spatial Lag model. Compare the results obtained with the three methods.

Exercise 5.4 Consider a system of four spatial units arranged on a 2-by-2 regular square lattice grid and a Spatial Lag model described by the following system of equations:

$$
\begin{aligned}
y_1 &= a_1 w_{12} y_2 + b_1 w_{13} y_3 + \varepsilon_1 \\
y_2 &= a_2 w_{21} y_1 + b_2 w_{24} y_4 + \varepsilon_2 \\
y_3 &= a_3 w_{31} y_1 + b_3 w_{34} y_4 + \varepsilon_3 \\
y_4 &= a_4 w_{43} y_3 + b_4 w_{42} y_2 + \varepsilon_4
\end{aligned}
$$

Derive the parameter conditions under which the model is symmetric and those under which the model is isotropic without symmetry. Write the equations in a compact matrix notation in the case of isotropy and symmetry.

Exercise 5.5 The hypothesis of normality is not essential to the Bivariate Marginal Maximum Likelihood estimator (*BML*) illustrated in section 5.4.2. Consider, as an alternative, the hypothesis that the errors

follow a bivariate exponential distribution Type II (Gumbel, 1960). In this case, we have the following model:

$$y_i = \beta \mathrm{x}_i + \varepsilon_i$$

and, for each pair of disturbances, say i and l, such that $l \in N(i)$:

$$f_{\varepsilon_i \varepsilon_l}(\varepsilon_i \varepsilon_l) = e^{-\varepsilon_i - \varepsilon_l}\left[1 + \alpha\,(2e^{-\varepsilon_i} - 1)(2e^{-\varepsilon_l} - 1)\right], i = 1, \cdots, q;\ \rho = \frac{\alpha}{4}$$

with ρ the spatial error parameter. Under these hypotheses, derive the likelihood function and the conditions for the *BML* estimators of the vector of parameters β.

References to the Chapter

Anderson, T. (2003) *An Introduction to Multivariate Statistical Analysis*, 3rd edition, Hoboken, NJ: Wiley.

Arbia G., (1990) 'On Second Order Non-Stationarity in Two-Dimensional Lattice Processes', *Computational Statistics and Data Analysis*, 9, 147–60.

Arbia, G. (2006) *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Heidelberg: Springer-Verlag.

Arbia, G. (2012) 'Simplifying the Estimation and Hypothesis Testing phases in Spatial Econometric Modeling of large Databases', Keynote speech of the 6th Annual World Conference of the Spatial Econometrics Association, Salvador-Bahia, Brazil, 11–13 July.

Arbia, G. (2014) 'Pairwise Likelihood Inference for Spatial Regressions Estimated on Very Large Datasets', *Spatial Statistics*, 7, 21–39.

Arbia, G., Bee, M. and Espa, G. (2013) 'Testing Isotropy in Spatial Econometric Models', *Spatial Economic Analysis*, 8, 3, 228–40.

Bell, K.P. and Bockstael, N.E. (2000) 'Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data', *Review of Economics and Statistics*, 82, 72–82.

Besag, J. (1974) 'Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion)', *Journal of the Royal Statistical Society B*, 36, 192–236.

Chiu, Y. M., Leonard, T. and Tsui, K. (1996) 'The Matrix-Logarithmic Covariance Model', *Journal of The American Statistical Association*, 91, 198–210.

Cressie, N. (1993) *Statistics for Spatial Data*, New York: Wiley.

Davis, R.A. and Yau, C.Y. (2011) 'Comments on Pairwise Likelihood in Time Series', *Statistica Sinica,* 21, 255–77.

Granger, C.W.J. (1974) 'Spatial Data and Time Series Analysis', in A. Scott (ed.), *Studies in Regional Science*, London: Pion.

Griffith, D.A. (2000) 'Eigenfunction Properties and Approximations of Selected Incidence Matrices Employed in Spatial Analysis', *Linear Algebra and its Applications,* 321, 95–112.

Griffith, D.A. (2004) 'Extreme Eigenfunctions of Adjacency Matrices for Planar Graphs Employed in Spatial Analysis', *Linear Algebra and its Applications*, 388, 201–19.

Gumbel, E.J. (1960) 'Bivariate Exponential Distribution', *Journal of the American Statistical Association,* 55, 698–707.
Heine J. and Mahorta, T. (2002) 'Mammography Tissue, Breast Cancer Risk, Serial Image Analysis and Digital Mammography: Tissue and Related Risk Factors', *Acad. Radiol.,* 9, 298–316.
Holly, S., Pesaran, M.H. and Takashi Y. (2010) 'A Spatio-Temporal Model of House Prices in the USA', *Journal of Econometrics*, 158, 1, 160.
Kelejian, H. and Prucha, I. (1998) 'A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances', *Journal of Real Estate Finance and Economics*, 17, 99–121.
Künsch, H. (1989) 'The Jackknife and Bootstrap for General Stationary Observations', *The Annals of Statistics*, 17, 1217–41.
Lesage J. and Pace, K. (2007) 'A Matrix Exponential Spatial Specification', *Journal of Econometrics,* 140, 1, 190–214.
LeSage, J. and Pace, K. (2009) *Introduction to Spatial Econometrics*, Boca Raton, FL: Wiley.
Lindsay, B.G. (1988) 'Composite Likelihood Methods', *Contemporary Mathematics*, 80, 221–39.
Moler C. and Van Loan, C. (2003) 'Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later', *SIAM Review*, 45, 1, 3–49.
Ord, J.K. (1975) 'Estimation Methods for Spatial Interaction', *Journal of the American Statistical Association*, 70, 120–6.
Pace, K. and LeSage, J. (2004) 'Chebyshev Approximation of Log-Determinants of Spatial Weight Matrices', *Computational Statistics and Data Analysis*, 45, 179–96.
Pace, L. and Salvan, A. (1997) *Principles of Statistical Inference*, Singapore: World Scientific.
Paelinck, J.H.P. and Nijkamp, P. (1975) *Operational Theory and Methods in Regional Analysis*, Farnborough: Saxon House.
Paelinck, J.H.P. and Klaassen, L.H. (1979) *Spatial Econometrics*, Aldershot, UK: Gower.
Schabenberger, O. and Gotway, C.A. (2002) *Statistical Methods for Spatial Data Analysis,* Boca Raton, FL: Chapman & Hall/CRC.
Sherman, M. (1996) 'Variance Estimation for Statistics Computed from Spatial Lattice Data', *Journal of the Royal Statistical Society, Series* B, 58, 509–23.
Smirnov O. and Anselin, L. (2001) 'Fast Maximum Likelihood Estimation of Very Large Spatial Autoregressive Models: A Characteristic Polynomial Approach', *Computational Statistics and Data Analysis*, 35, 301–19.
Solow, A. (1985) 'Bootstrapping Correlated Data', *Journal of the International Association for Mathematical Geology*, 17, 769–75.
Varin, C. and Vidoni, P. (2009) 'Pairwise Likelihood Inference for General State Space Models', *Econometric Reviews*, 28, 1–3, 170–85.
Varin, C. (2008) 'On Composite Marginal Likelihood', *Advances in Statistical Anal*., 92, 1–28.
Varin, C., Reid, N. and Firth, D. (2011) 'An Overview of Composite Likelihood Methods', *Statistica Sinica*, 21, 5–42.
Wang, H., Iglesias, E.M. and Wooldridge, J.M. (2013) 'Partial Maximum Likelihood Estimation of Spatial Probit Models', *Journal of Econometrics*, 172, 77–89.

6
Conclusions: What's Next?

As mentioned in the preface, we conceived this text not as a comprehensive book, but as a bridge between the unspecialized econometrics textbook literature and the more advanced spatial econometrics textbooks. So we imagine our typical reader as someone who has studied a general econometrics textbook (such as, for example, Greene, 2011), has devoted some time to the study of the present monograph and is now eagerly awaiting directions as to where to go next. In this respect, we can suggest different (although not necessarily mutually exclusive) directions.

Similar to what occurs in time series econometrics with the theory of random processes (Hamilton, 1994), at the basis of the spatial econometric models treated in this book, we find the notion of random fields. If the reader wants to go into more depth regarding the inferential statistical foundations and the probabilistic roots of the models treated here, they can refer to Arbia (2006) which contains an exhaustive account of the random field models which are the basis of the *SARAR* paradigm, plus a number of other random fields which may be potentially very useful in spatial econometrics and are still not adequately explored in the current literature. One example is represented by an alternative specification of the Spatial Lag and the Spatial Error models based on conditional, rather than marginal, probability distribution. This approach was originally introduced by Besag (1974) and has been recently considered in the spatial econometrics literature by Sain and Cressie (2007) and Ippoliti et al. (2013). For the reader who is interested in exploring the deep roots of spatial econometric modeling, these can be found in the spatial statistical literature. Good introductions to the topic can be found in books like Ripley (1981), Cressie (1993), Gaetan and Guyon (2009) and Cressie and Wikle (2011).

In this respect, in spatial statistics it is traditional to distinguish between methods designed for data observed on points, lines and areas. Point data methods are designed to study the observed regularities in the distribution of individuals on the plane; line data methods are defined to explain the spatial interactions between individuals that are observed along networks (such as flows of goods, individuals and information in the geographical space); finally, areal methods refer to data observed as aggregates within portions of space, generally countries or regions. In the present book we have chosen to concentrate mainly on the treatment of regional data while neglecting the other two typologies of spatial data. Readers who, after consulting this book, are interested in deepening their knowledge in these areas, can refer to the existing literature (still scattered in econometrics) on the so-called *point pattern analysis* (see, for example, papers such as Marcon and Puesch, 2003; Duranton and Overman, 2005 and Arbia et al., 2008, 2010, 2012) for point data and to books like LeSage and Pace (2009) and Patuelli and Arbia (2014) for *spatial interaction models*.

Even if this textbook covers a substantial part of the spatial econometric models, there are still a number of topics intentionally omitted to keep the discussion limited to the essentials and as simple as possible. If the readers wish to enlarge their view on the variety of modeling alternatives available, they can consult the book by Anselin (1988) which, although dated, still represents a source of references in the subject. Here the reader will find references, amongst others, to topics such as the analysis of spatial heterogeneity, robust estimation methods, seemingly unrelated spatial regressions, pre-testing and bootstrapping in a spatial context, non-nested tests, spatial expansion methods, edge effects, and others. Furthermore, throughout this book we consistently followed a frequentist-Fisherian approach to model inference. The Bayesian approach is not treated here and the reader can refer again to Anselin (1988) and to the extensive account given in LeSage and Pace (2009). Finally, an interesting emerging topic in spatial econometrics, omitted in the present context, is represented by spatial quantile regression models for which the reader is referred to the recent comprehensive book by McMillen (2013).

If the interest is mainly practical, the reader will have to develop a deeper knowledge and practice in the available software to treat spatial data and spatial relationships. As clarified in the preface, even if this textbook presents a series of computer tutorials and practical exercises on R, it certainly cannot be considered a comprehensive course on the use of R in spatial econometrics. A rich source of references in this

respect is the textbook by Bivand et al. (2008). Furthermore commercially available software which are very useful in spatial econometric modeling are Matlab, STATA and GeoDa. Matlab contains toolboxes for spatial econometric analysis developed by Pace and by LeSage (see the website http://www.mathworks.com/products/matlab/ and the website www.spatialeconometrics.com). Similarly, the software STATA has recently developed a very rich set of spatial econometrics procedures (see the website http://www.stata.com/ and papers like Drukker et al., 2013a, 2013b, 2013c). Finally, the software GeoDA and PySAL, developed by Luc Anselin and his co-workers, contain one of the more extensive libraries of procedures in spatial analysis and spatial econometrics (see https://geodacenter.asu.edu/software).

Finally, if the aim of our readers is to have familiar grasp of the ongoing, cutting-edge research in the subject, under both the theoretical and the applied points of view, they can refer to the paper by Arbia (2012) (who surveys 230 papers published in the period 2007–11) and to a series of special issues of various scientific journals devoted to the subject which appeared in recent years, such as those hosted in journals like *Journal of Econometrics* (see Baltagi et al., 2007), *Empirical Economics* (see Arbia and Baltagi, 2008), *Papers in Regional Science* (see Arbia and Fingleton, 2008), *Regional Science and Urban Economics* (see Arbia and Kelejian, 2010), *Journal of Regional Science* (see Partridge et al., 2012), *Economic Modelling* (see Arbia et al., 2012), *Spatial Economic Analysis* (see Arbia and Prucha, 2013) and *Geographical Analysis* (see Arbia and Thomas, 2014).

References to the Chapter

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Dordrecht: Kluwer Academic Publishers.

Arbia, G. (2006) *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*, Heidelberg: Springer-Verlag.

Arbia, G. (2012) 'A Lustrum of SEA: Recent Research Trends Following the Creation of the Spatial Econometrics Association (2007–2011)', *Spatial Economic Analysis*, 6, 4, 377–96.

Arbia, G. and Baltagi, B. (2008) 'Introduction to the Special Issue on Spatial Econometrics', *Empirical Economics*, 34, 1–4.

Arbia, G. and Fingleton, B. (2008) 'New Spatial Econometric Techniques and Applications in Regional Science', *Papers in Regional Sciences*, 87, 3, 311–17.

Arbia, G. and Kelejian, H. (2010) 'Advances in Spatial Econometrics', *Regional Science and Urban Economics*, 40, 5, 253–4.

Arbia, G. and Prucha, I. (2013) 'Introduction to the Special Issue on Spatial Econometrics', *Spatial Economic Analysis*, 8, 3, 215–17.

Arbia, G. and Thomas, C. (2014) 'New Developments in Spatial Econometrics', *Geographical Analysis* (forthcoming).

Arbia, G., Espa, G. and Giuliani, D. (2013) 'Conditional vs. Unconditional Industrial Agglomeration: Disentangling Spatial Dependence and Spatial Heterogeneity in the Analysis of ICT Firms' Distribution in Milan (Italy)', *Journal of Geographical Systems*, 15, 31–50.

Arbia, G., Espa, G. and Giuliani, D. (2014) 'Weighting Ripley's K-function to Account for the Firm Dimension in the Analysis of Spatial Concentration (to appear in *International Regional Science Review*)', first published on 9 November 2012 as doi:10.1177/0160017612461357.

Arbia, G., Espa, G. and Quah, D. (2008) 'A Class of Spatial Econometric Methods in the Empirical Analysis of Clusters of Firms in Space', *Empirical Economics,* 34, 1, 81–103.

Arbia, G., Lopez-Bazo, E. and Moscone, F.(2012) 'Frontiers in Spatial Econometrics Modelling (with Enrique Lopez Bazo and Francesco Moscone)', *Economic Modelling*, 29, 1, 1–2.

Arbia, G., Espa, G., Giuliani, D. and Mazzitelli, A. (2010) 'Detecting the Existence of Space-Time Clusters of Firms', *Regional Science and Urban Economics*, 40, 5, 311–23.

Arbia, G., Espa, G., Giuliani, D. and Mazzitelli, A. (2012) 'Clusters of Firms on an Inhomogeneous Space: The Case of High-Tech Industries in Milan in 2001', *Economic Modelling*, 29, 1, 3–11.

Baltagi, B., Kelejian, H. and Prucha, I. (2007) 'Analysis of Spatially Dependent Data', *Journal of Econometrics,* 140, 1, 1–4.

Besag, J. (1974) 'Spatial Interaction and the Statistical Analysis of Lattice Systems', *Journal of the Royal Statistical Society, Series B,* 36, 192–236.

Bivand, R., Pedesma, J. and Gomèz-Rubio, V. (2008) *Applied Spatial Data Analysis with R*, Heidelberg: Springer-Verlag.

Cressie, N. (1993) *Statistics for Spatial Data*, revised edition of the 1991 original, New York: Wiley.

Cressie, N. and Wikle, C. (2011) *Statistics for Spatio-Temporal Data*, Hoboken, NJ: Wiley.

Drukker, D.M., Prucha, I. and Raciborski, R. (2013a) 'Maximum Likelihood and Generalized Spatial Two-Stage Least-Squares Estimators for a Spatial-Autoregressive Model with Spatial-Autoregressive Disturbance', *Stata Journal,* 13, 2, 221–41.

Drukker, D.M., Prucha, I. and Raciborski, R. (2013b) 'A Command for Estimating Spatial-Autoregressive Models with Spatial-Autoregressive Disturbances and Additional Endogenous Variables', *Stata Journal*, 13, 2, 287–301.

Drukker, D.M., Peng, H., Prucha, I. and Raciborski, R. (2013c) 'Creating and Managing Spatial-Weighting Matrices with the spmat Command', *Stata Journal,* 13, 2, 242–86.

Duranton G., Overman H.G. (2005) 'Testing for Localisation Using Micro-Geographic Data', *Rev Econ Stud* 72, 4, 1077–106.

Gaetan, C. and Guyon, X. (2009) *Spatial Statistics and Modeling*, New York: Springer-Verlag.

Greene, W. (2011) *Econometric Analysis*, 7th edition, Upper Saddle River, NJ: Pearson Education.Hamilton, J.D. (1994) *Time Series Analysis*, Princeton, NJ: Princeton University Press.

Ippoliti, L., Romagnoli, L. and Arbia, G. (2013) 'A Gaussian Markov Random Field Approach to Convergence Analysis', *Spatial Statistics*, 6, 78–90.

LeSage, J. and Pace, K. (2009) *Introductory Spatial Econometrics*, Boca Raton, FL: Chapman & Hall/CRC Press.

Marcon E. and Puech, F. (2003) 'Evaluating the Geographic Concentration of Industries Using Distance-Based Methods', *Journal of Economic Geography,* 3, 4, 409–28.

Marcon E. and Puech, F. (2010) 'Measures of the Geographic Concentration of Industries: Improving Distance-Based Methods', *Journal of Economic Geography,* 10, 5, 745–62.

McMillen, D. (2013) *Quantile Regression for Spatial Data*, SpringerBriefs in Regional Science, Heidelberg: Springer-Verlag.

Partridge, M. D., Boarnet, M., Brakman, S. and Ottaviano G. (2012) 'Introduction: Whither Spatial Econometrics?', *Journal of Regional Science*, 52, 2, 167–71.

Patuelli, R. and Arbia, G. (2014) *The Spatial Econometrics of Spatial Interaction Models*, Heidelberg: Springer-Verlag (forthcoming).

Ripley, B. (1981) *Spatial Statistics*, New York: Wiley.

Sain, S.R. and Cressie, N. (2007) 'A Spatial Model for Multivariate Lattice Data', *Journal of Econometrics*, 140, 1, 226–59.

Solutions to the Exercises

*In all exercises, Signif. Codes are: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Exercise 1.1

1. Model 1

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*
(Intercept)	*–22.31118*	*3.38594*	*–6.589*	*0.000100****
Labor	*0.27750*	*0.05346*	*5.191*	*0.000571****
business	*0.42239*	*0.47243*	*0.894*	*0.394567*
F-statistic	*43.99*			*2.59e–06****

Multiple R-squared: 0.9072, Adjusted R-squared: 0.8866

2. Model 2

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*
(Intercept)	*–21.38866*	*3.19237*	*–6.700*	*5.36e–05****
Labor	*0.31460*	*0.03335*	*9.432*	*2.71e–06****
**F-statistic*	*88.97*			*2.71e–06****

Multiple R-squared: 0.8431, Adjusted R-squared: 0.8274

Model 3

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*
(Intercept)	*–16.0547*	*5.9992*	*–2.676*	*0.02325**
business	*2.3264*	*0.5646*	*4.121*	*0.00208***
F-statistic	*16.98*			*0.00208****

Multiple R-squared: 0.6293, Adjusted R-squared: 0.5923

3. The best model is model 2 which attains the highest significance level of the F-statistics with a high value of the adjusted R-square.

4. Model 4

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*
(Intercept)	*22.546*	*18.718*	*1.205*	*0.25613*
business	*6.861*	*1.761*	*3.895*	*0.00298***
F-statistic:	*15.17*			*0.00298***

Multiple R-squared: 0.6027, Adjusted R-squared: 0.563

5. Residuals of model 1

Wales	*−0.6329899*
Scotland	*−0.8822571*
Northern Ireland	*−1.1388039*
North of England	*−3.1397582*
North West of England	*2.5364883*
Yorkshire & Humberside	*1.2721572*
East Midlands	*−0.5921008*
West Midlands	*0.4511717*
East Anglia	*−0.2855529*
Greater London	*−1.0219450*
South East England	*2.3965191*
South west England	*1.0370716*

6. Breusch–Pagan test: BP = 1.5183, df = 2, p-value = 0.4681
Jarque Bera Test: JB= 0.092, df = 2, p-value = 0.9551

7. Model 1 overerestimates the GVA (negative residuals) in the East regions (East of England, East Midlands and London) and in the North (North of England, North West England, Scotland and Northern Ireland) and underestimates it in the center. Thus there appears to be a distinct geographical pattern of residuals.

Exercise 2.1

We order the 8 regions as follows: (RO11,RO12,RO21,O22,RO31,RO32,RO41,RO42)

Unstandardized W Matrix:

0	1	1	0	0	0	0	1
1	0	1	1	1	0	1	1
1	1	0	1	0	0	0	0
0	1	1	0	1	0	0	0
0	1	0	1	0	1	1	0
0	0	0	0	1	0	0	0
0	1	0	0	1	0	0	1
1	1	0	0	0	0	1	0

Standardized W Matrix:

0	0,333333	0,333333	0	0	0	0	0,333333
0,166667	0	0,166667	0,166667	0,166667	0	0,166667	0,166667
0,333333	0,333333	0	0,333333	0	0	0	0
0	0,333333	0,333333	0	0,333333	0	0	0
0	0,25	0	0,25	0	0,25	0,25	0
0	0	0	0	1	0	0	0
0	0,333333	0	0	0,333333	0	0	0,333333
0,333333	0,333333	0	0	0	0	0,333333	0

Percentage of non-zero entries =26/64= 40.625 %

Exercise 2.2 *Spatially lagged variable:*

Ly=(9.6999903 9.7666862 10.0333233 10.1666565 9.1 10.3 9.7666569 9.3666573)

Exercise 2.3

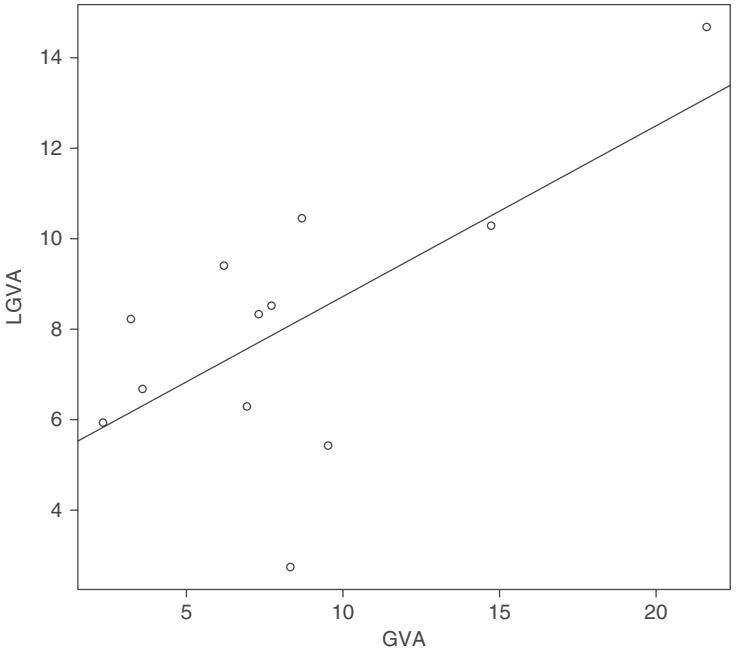
L(X)=(26, 15.67, 23.67, 17, 24.5, 25.33, 27.25, 14, 26, 29, 25, 16, 24, 21.25, 20.33, 14.67, 17.5, 14.5, 20.5, 16, 14.5, 15, 20.67, 14, 21.5)

Exercise 2.4 *Here below we display the .GAL file*

```
0 12 UK UK_reg
  1 4
  3 5 8 12
  2 2
  3 4
  3 2
  1 2
  4 3
  2 5 6
  5 5
  1 4 6 7 8
  6 3
  4 5 7
  7 5
  5 6 8 9 11
  8 5
  1 5 7 11 12
  9 2
  7 11
  10 1
  11
  11 5
  7 8 9 10 12
  12 3
  1 8 11
```

Exercise 2.5 *L(GVA)=(6.7, 2.75, 5.95, 8.233, 5.44, 6.3, 9.42, 8.34, 10.45, 14.7, 10.3, 8.53)*

Exercise 2.6 *Moran scatterplot:*



There is a positive relationship between the GVA and its spatially lagged value (LGVA) showing spatial correlation for this variable.

Exercise 2.7

Observed Moran's I = –0.17589503
Expectation = –0.16886659
Variance = 0.03307571
Standardized Moran I = –0.0386
p-value = 0.5154

Spatial uncorrelation of residuals should be accepted at 49 % confidence level.

Exercise 2.9

Exercise 3.1 *Here below we display the .GAL file*

0 27 eu eu_countries
1 4
5 9 14 25
2 2
18 24
3 4
5 15 16 20
4 2
5 22
5 8
1 3 4 9 14 15 16 25
6 2
21 26
7 1
23
8 2
9 17

9 6
1 5 8 10 23 25
10 4
9 15 19 27
11 1
24
12 2
16 26
13 4
15 18 19 20
14 2
1 5
15 6
3 5 10 13 19 20
16 4
3 5 12 20
17 1
8
18 2
2 13
19 3
10 13 15
20 4
3 13 15 16
21 2
6 22
22 2
4 21
23 2
7 9
24 2
2 11
25 3
1 5 9
26 2
6 12
27 1
10

SARAR model

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	*0.31168102*	*0.19459931*	*1.6017*	*0.1092*
educ	*0.00094883*	*0.00055511*	*1.7093*	*0.0874 .*
Lambda	*0.6819*		*3.6192*	*0.00029551****
Rho:	*–0.47147*		*–1.6595*	*0.097007 .*
LR test:	*4.2047*			*0.12217*

AIC: –91.606

SEM

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	*1.0294194*	*0.0286529*	*35.9273*	*< 2e–16****
educ	*0.0014198*	*0.0007987*	*1.7776*	*0.07546 .*
Rho	*0.32931*		*1.7495*	*0.080209 .*
LR test value:	*2.6041*			*0.10659*
Wald statistic:	*3.0607*			*0.080209 .*
AIC: –92.005				

SLM

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	*0.66359077*	*0.19079632*	*3.4780*	*0.0005052****
educ	*0.00132915*	*0.00071032*	*1.8712*	*0.0613180 .*
Lambda	*0.34291,*		*1.8959*	*0.057977 .*
LR test value:	*3.1194*			*0.077367 .*
Wald statistic:	*3.5943*			*0.057977 .*
AIC: –92.521				

The best model is SLM. Rho is non-significant in SARAR and SEM and SLM achieves the highest AIC.

Exercise 3.2 *Writing the model as:*

$$y_i = \beta X_i + u_i;\ u_i = \rho L(u_i) + \varepsilon_i;$$

the second equation can be written as

$$u_i - \rho L(u_i) = \varepsilon_i$$

$$(1 - \rho L)u_i = \varepsilon_i$$

and the first equation as

$$(1 - \rho L)y_i = (1 - \rho L)\beta X_i + (1 - \rho L)u_i$$

$$y_i = \rho L y_i + \beta X_i - \rho\beta L X_i + \varepsilon_i$$

$$y_i = \rho L y_i + \beta X_i - \gamma L X_i + \varepsilon_i$$

This is a Spatial Lag model with a spatially lagged independent variable L(X).

A simple OLS strategy cannot be employed because the coefficient γ is subject to the constraint $\gamma = \rho\beta$.

Exercise 3.3 *Writing the model as:*

$$y = X\beta + u$$

$$u = \rho W u;$$

the second equation can be written as

$$(1 - \rho W)u = \varepsilon$$

and the first equation as

$$(1 - \rho W)y = \beta(1 - \rho W)X + (1 - \rho W)u$$

or

$$y^{\star} = \beta X^{\star} + \varepsilon$$

which can estimated with the OLS by minimizing the expression

$$\varepsilon^T \varepsilon = \left(y^{\star} - \beta X^{\star}\right)^T \left(y^{\star} - \beta X^{\star}\right) = \min$$

or

$$\left(y - \beta X\right)^T (I - \rho W)^T (I - \rho W)\left(y - \beta X\right) = \min$$

with $y^{\star} = (I - \rho W)y$ *and* $X^{\star} = (I - \rho W)X$.

Consider now the likelihood of a SEM model (Equation 3.23)

$$l(\rho, \sigma_{\varepsilon}^2, \beta) = const - \frac{n}{2}\ln(\sigma_{\varepsilon}^2) - \frac{1}{2}\ln\left|(I - \rho W)^{-1}(I - \rho W)^{-T}\right| - \frac{1}{2}(y - X\beta)^T\left[(I - \rho W)^{-1}(I - \rho W)^{-T}\right]^{-1}(y - X\beta)$$

For known ρ *and* σ_{ε}^2 *this expression is a maximum for* β *if*

$$(y - X\beta)^T\left[(I - \rho W)^{-1}(I - \rho W)^{-T}\right]^{-1}(y - X\beta) = \min$$

or

$$(y - X\beta)^T (I - \rho W)^T (I - \rho W)(y - X\beta) = \min$$

which coincides with the previous expression obtained for OLS.

From the GLS solution we have from Equation (1.47):

$$(X^{*T}\ X^{*})^{-1}X^{*T}\ y^{*} = (X^{T}P^{-T}PX)^{-1}X^{T}P^{-T}Py$$
$$= (X^{T}\Omega^{-1}X)^{-1}X^{T}\Omega^{-1}y = \hat{\beta}_{GLS}$$

having set $P^{-1} = (I - \rho W)$ *and as a consequence.* $\Omega^{-1} = P^{-T}P^{-1} = (I - \rho W)^{T}(I - \rho W)$. *So the ML estimator of the* β *coefficient in a SEM model can be expressed as:*

$$\hat{\beta}_{GLS} = \left[X^{T}(I - \rho W)^{T}(I - \rho W)X\right]^{-1}$$
$$\times\left[X^{T}(I - \rho W)^{T}(I - \rho W)\right]y$$

with error variances that can be obtained from the main diagonal of the matrix (see Equation 1.48)

$$Var(\hat{\beta}_{GLS}) = \sigma_{\varepsilon}^{2}(X^{T}\Omega^{-1}X)^{-1} = \sigma_{\varepsilon}^{2}\left[X^{T}(I - \rho W)^{T}(I - \rho W)X\right]^{-1}$$

Exercise 3.4 *Since the variable X is non-stochastic, we can rewrite the model as* $y_i = \lambda \dfrac{\sum_{i=1}^{n} w_{ij}y_i}{\sum_{j=1}^{n} w_{ij}} + \vartheta_i$; *with* $\vartheta_i = \beta X_i + u_i$. *Furthermore, given that* $u_i|X \approx i.i.d.N(0,\sigma_{u\,n}^{2}I_n)$ *then:*

$$E(\vartheta_i) = E(\beta X_i + u_i) = \beta E(X_i) + E(u_i) = \beta X_i$$

and

$$Var(\vartheta_i) = Var(\beta X_i + u_i) = Var(u_i) = \sigma_u^2$$

So the equation describes a purely autoregressive model with an error $\vartheta_i|X \approx i.i.d.N(\beta X_i, \sigma_u^2 I)$.

Exercise 3.5 *The model can be re-written as:*

$$y = -\rho Wy + u$$
$$u = \rho Wu + \varepsilon$$

or

$$(I + \rho W)y = u$$
$$(I - \rho W)u = \varepsilon$$

So that

$$y = (I + \rho W)^{-1} u$$
$$u = (I - \rho W)^{-1} \varepsilon$$

and therefore:

$$y = (I + \rho W)^{-1}(I - \rho W)^{-1}\varepsilon = [(I - \rho W)(I + \rho W)]^{-1}\varepsilon$$
$$= \left(I - \rho^2 W^2\right)^{-1}\varepsilon = \left(I - \rho^* W^*\right)^{-1}\varepsilon$$

This is a SLM

$$y = \rho^* W^* y + Z\beta + u \qquad \left|\rho^*\right| < 1$$

with weight matrix $W^* = W^2$ *and parameter* $\rho^* = \rho^2$ *. Since* $|\rho| < 1$*, then also* $\left|\rho^*\right| < 1$*. From (Equation 3.47) the variance-covariance matrix of y is:*

$$E\left(yy^T\right) = \sigma_\varepsilon^2 \left(I - \rho W\right)^{-1}\left(I - \rho W\right)^{-T}$$
$$= \sigma_\varepsilon^2 \left(I - \rho^2 W^2\right)^{-1}\left(I - \rho^2 W^2\right)^{-T}$$

and from Equation (3.48) its likelihood:

$$L(\sigma^2, \lambda, \beta; y) = const\left|\sigma_\varepsilon^2 \Omega\right|^{-\frac{1}{2}}$$
$$\times \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\left[y - (I - \rho^2 W^2)^{-1} Z\beta\right]^T \Omega^{-1}\left[y - (I - \rho^2 W^2)^{-1} Z\beta\right]\right\}$$

Exercise 3.6

ML estimation

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	*40.6458123*	*5.2937581*	*7.6781*	*1.621e–14****
CRIM	*–0.1188867*	*0.0324540*	*–3.6632*	*0.0002491****
RM	*3.8507202*	*0.4062432*	*9.4789*	*< 2.2e–16****
INDUS	*–0.0059026*	*0.0618481*	*–0.0954*	*0.9239678*
NOX	*–20.4193252*	*4.0011873*	*–5.1033*	*3.338e–07****
AGE	*–0.0195181*	*0.0140191*	*–1.3923*	*0.1638451*
DIS	*–1.4560840*	*0.2717635*	*–5.3579*	*8.419e–08****
RAD	*0.3219532*	*0.0732975*	*4.3924*	*1.121e–05****
PTRATIO	*–1.0407873*	*0.1366262*	*–7.6178*	*2.576e–14****
B	*0.0098856*	*0.0026439*	*3.7391*	*0.0001847****
LSTAT	*–0.5149470*	*0.0496189*	*–10.3780*	*< 2.2e–16****
TAX	*–0.0112409*	*0.0038685*	*–2.9058*	*0.0036637****

Rho:	*0.57191*	*6.3752*	*1.8267e–10****
LR test value:	*25.792*		*3.8025e–07****
Wald statistic:	*40.644*		*1.8267e–10****
AIC: 3021.4			

FGLS estimation

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	*40.5245582*	*5.2994923*	*7.6469*	*2.065e–14****
CRIM	*–0.1180917*	*0.0325913*	*–3.6234*	*0.0002907****
RM	*3.8591297*	*0.4082023*	*9.4540*	*< 2.2e–16****
INDUS	*–0.0044561*	*0.0620706*	*–0.0718*	*0.9427686*
NOX	*–20.2981368*	*4.0071806*	*–5.0654*	*4.075e–07****
AGE	*–0.0186184*	*0.0140273*	*–1.3273*	*0.1844113*
DIS	*–1.4431412*	*0.2636080*	*–5.4746*	*4.386e–08****
RAD	*0.3217374*	*0.0731746*	*4.3968*	*1.098e–05****
PTRATIO	*–1.0462088*	*0.1365008*	*–7.6645*	*1.799e–14****
B	*0.0098673*	*0.0026561*	*3.7149*	*0.0002033****
LSTAT	*–0.5156174*	*0.0498751*	*–10.3382*	*< 2.2e–16****
TAX	*–0.0112381*	*0.0038747*	*–2.9004*	*0.0037267****
Rho:	*0.53872*		*0.60881*	*0.88488****

Exercise 4.1 *The error variance-covariance matrix can be written as:*

$$Var(\bar{\varepsilon}) = \bar{\varepsilon}\bar{\varepsilon}^T = \bar{\varepsilon} G G^T \bar{\varepsilon}^T = \sigma_\varepsilon^2 G G^T.$$

Now define the aggregation matrix G as:

$$G = \begin{bmatrix} 1 \underbrace{\cdots}_{n_1 \text{ times}} 1 & & & \\ & 1 \underbrace{\cdots}_{n_2 \text{ times}} 1 & & \\ & & \ddots & \\ & & & 1 \underbrace{\cdots}_{n_m \text{ times}} 1 \end{bmatrix} \text{ and } GG^T = \begin{bmatrix} n_1 & & & \\ & n_2 & & \\ & & \ddots & \\ & & & n_m \end{bmatrix}$$

So $Var(\bar{\varepsilon})$ *has non-constant diagonal terms and* $\bar{\varepsilon}$ *is heteroscedastic.*

Exercise 4.2

Uniform*:*

$$\int_{-\infty}^{\infty} K(v)\, dv = \int_{-1}^{1} 0.5\, dv = [0.5v]_{-1}^{1} = 1$$

Epanechnikov *(quadratic)*

$$\int_{-\infty}^{\infty} K(v)dv = \int_{-1}^{1} \frac{3}{4}\left(1-v^2\right)\, dv = \frac{3}{4}\int_{-1}^{1}\left(1-v^2\right)\, dv = \frac{3}{4}\left[v - \frac{v^3}{3}\right]_{-1}^{1} = 1$$

Quadratic *(bi-weight)*

$$\begin{aligned}\int_{-\infty}^{\infty} K(v)dv &= \int_{-1}^{1} \frac{15}{16}\left(1-v^2\right)^2\, dv = \frac{15}{16}\int_{-1}^{1}\left(1-v^2\right)^2\, dv \\ &= \frac{15}{16}\int_{-1}^{1}\left(1-2v^2+v^4\right)\, dv \\ &= \frac{15}{16}\left[v - \frac{2}{3}v^3 + \frac{v^5}{5}\right]_{-1}^{1} = 1\end{aligned}$$

Exercise 4.3

	Estimate	*Std. Error*	*t-value*	*Pr(>|t|)*
(Intercept)	*0.4923310*	*0.6317384*	*0.7793*	*0.4358*
educ	*0.0014920*	*0.0012071*	*1.2361*	*0.2164*
lambda	*0.4967469*	*0.6222546*	*0.7983*	*0.4247*
rho	*–0.5638543*	*0.4352350*	*–1.2955*	*0.1951*
Wald test that rho and lambda are both zero	*0.013056*			*0.90903*

Exercise 4.4 *Consider the model:*

$$y^{\bullet} = X\beta + u$$

$$u = \rho W u + \varepsilon$$

We have

$$u = (I - \lambda W)^{-1}\varepsilon$$

and, consequently:

$$E(u) = 0$$

$$E(uu^T) = \Omega = \left[(I - \lambda W)^T (I - \lambda W)\right]^{-1}$$

The likelihood of the single observation is

$$P(y_i = 1) = P(y_i^{\bullet} > 0 \,|\, X_i, w_{ij} y_i^{\bullet})$$

$$= P(X_i \beta + u_i > 0 \,|\, X_i, w_{ij}, y_i^{\bullet}) = P(u_i \leq X_i \beta \,|\, X_i, w_{ij}, y_i^{\bullet}) \cong \Phi(X_i \beta)$$

with $\Phi(\cdot) MVN\ (0, \Omega)$
and the likelihood, from Equation (4.30):

$$L = \int_{-\infty}^{\mu_1} \int_{-\infty}^{\mu_2} \cdots \int_{-\infty}^{\mu_n} \varphi(\mu) d\mu$$

with $\phi(\mu) = (2\pi)^{-\frac{n}{2}} |\Omega| \exp\left[-\frac{1}{2}(\mu^T \Omega \mu)\right]$ *where* $\mu_i \in \mu$, $\mu_i = X_i \beta$ *and* $\Omega = \left[(I - \lambda W)^T (I - \lambda W)\right]^{-1}$.

Exercise 4.5

Standard Probit

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	*–2.88609*	*1.14867*	*–2.513*	*0.0120**
educ	*0.06724*	*0.03067*	*2.192*	*0.0284**

AIC: 31.126

Probit ML

	Estimate	*Std. Error*	*z-value*	*Pr(>|z|)*
(Intercept)	*–3.22362474*	*1.96301263*	*–1.6421824*	*0.1005522*
educ	*0.07365255*	*0.04548824*	*1.6191556*	*0.1054138*
rho	*–0.16447635*	*0.65983415*	*–0.2492692*	*0.8031525*

Probit GMM

	Estimate	*Std. Error*	*z-value*	*Pr(>|z|)*
(Intercept)	*–3.43547076*	*1.63015676*	*–2.1074481*	*0.03507876**
educ	*0.07741729*	*0.03519487*	*2.1996756*	*0.02782992**
WXB	*–0.26053582*	*0.58076446*	*–0.4486084*	*0.65371417*

Probit LGMM

	Estimate	*Std. Error*	*z-value*	*Pr(>|z|)*
(Intercept)	*–3.37481*	*1.45691*	*–2.31641*	*0.02054**
educ	*0.07637*	*0.03393*	*2.25085*	*0.02439**
WXB	*–0.27239*	*0.65782*	*–0.41408*	*0.67882*

Exercise 4.6

$$I_T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix};\ A = (I_n - \lambda W) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -\lambda & -\lambda \\ -\lambda & 1 & 0 \\ -\lambda & 0 & 1 \end{bmatrix}$$

$$I_T \otimes A = \begin{bmatrix} 1 & -\lambda & -\lambda & 0 & 0 & 0 \\ -\lambda & 1 & 0 & 0 & 0 & 0 \\ -\lambda & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -\lambda & -\lambda \\ 0 & 0 & 0 & -\lambda & 1 & 0 \\ 0 & 0 & 0 & -\lambda & 0 & 1 \end{bmatrix}$$

$$J_T = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\ J_T \otimes I_n = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Exercise 4.7

SEM-RE

Error variance parameters:

	Estimate	*Std. Error*	*t-value*	*Pr(>|t|)*
phi	*39.838568*	*8.586543*	*4.6397*	*3.49e–06****
rho	*0.094122*	*0.066619*	*1.4128*	*0.1577*

Coefficients:

(Intercept)	*–5.855669*	*0.640199*	*–9.1466*	*< 2.2e–16****
log(rgdp)	*1.126815*	*0.065639*	*17.1669*	*<2.2e–16****

KKP

Error variance parameters:

	Estimate	*Std. Error*	*t-value*	*Pr(>|t|)*
phi	*28.727941*	*5.486499*	*5.2361*	*1.64e–07****
rho	*0.499248*	*0.059529*	*8.3866*	*< 2.2e–16****

Coefficients

(Intercept)	*–1.57211*	*0.78284*	*–2.0082*	*0.04462**
log(rgdp)	*0.68644*	*0.08021*	*8.5580*	*< 2e–16****

Only the KKP-like term is significant: hence, in the random effects approach, we conclude against spatial correlation in the idiosyncratic part of the error, but in favour of spatial correlation of individual effects.

Exercise 4.8

OLS a-spatial model

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*
(Intercept)	*–34.671*	*2.650*	*–13.08*	*<2e–16****
RM	*9.102*	*0.419*	*21.72*	*<2e–16****
F-statistic:	*471.8*			*< 2.2e–16****

Multiple R-squared: 0.4835, Adjusted R-squared: 0.4825

Spatial Lag Model

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	*–37.75130*	*2.45486*	*–15.378*	*< 2.2e–16****
RM	*7.60451*	*0.41218*	*18.450*	*< 2.2e–16****
Lambda:	*0.56252,*	*9.8236*		*< 2.22e–16****
LR test value:	*76.532*			*< 2.22e–16****
Wald statistic:	*96.504*			*< 2.22e–16****

AIC: 3277.6
LM test for residual autocorrelation *0.013478* *0.90758*

GWR

Summary of GWR coefficient estimates:

	Min.	*1st Qu.*	*Median*	*3rd Qu.*	*Max.*	*Global*
(Intercept)	*–120.600*	*–38.360*	*–23.130*	*–5.059*	*52.540*	*–34.6706*
RM	*–5.974*	*3.626*	*7.443*	*9.838*	*21.320*	*9.1021*

Moran's I
statistic = 12.578, p-value = 0.000597

Exercise 5.1

$$W^2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}; \quad W^3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

For $\lambda = 0.9502$, *then* $\alpha = -3$, *then* $S \approx \sum_{i=0}^{3} \frac{\alpha^i W^i}{i!} = 1 - 3W + \frac{9}{2}W^2 - \frac{27}{6}W^3$

Exercise 5.2 *Rearranging the units without loss of generality:*

$$W = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & & \\ 0 & 0 & \dots & \dots & 0 \end{bmatrix}; \quad (I - \lambda W) = \begin{bmatrix} 1 & -\lambda & 0 & \dots & 0 \\ 0 & 1 & -\lambda & \dots & 0 \\ & & & & \\ 0 & 0 & \dots & \dots & 1 \end{bmatrix}$$

$$(I - \lambda W)^{-1} = \begin{bmatrix} 1 & \lambda & \lambda^2 & \dots & \lambda^{n-1} \\ 0 & 1 & \lambda & \dots & \lambda^{n-2} \\ 0 & 0 & 1 & & \lambda^{n-3} \\ & & & & \\ 0 & 0 & \dots & \dots & 1 \end{bmatrix}; \quad |I - \lambda W| = 1 .$$

Furthermore, the variance-covariance matrix of a Spatial Lag model is:

$\Omega = (I - \lambda W)^{-1}(I - \lambda W)^{-T}$, and, since $|AB| = |A||B|$:

$$\left|(I - \lambda W)^{-1}(I - \lambda W)^{-T}\right| = \left|(I - \lambda W)^{-1}\right|\left|(I - \lambda W)^{-T}\right| = 1$$

Exercise 5.3

Full likelihood

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	*2.8378e+01*	*5.8225e+00*	*4.8739*	*1.094e–06****
CRIM	*–9.7501e–02*	*3.2606e–02*	*–2.9902*	*0.0027877****
RM	*3.8432e+00*	*4.1351e–01*	*9.2941*	*< 2.2e–16****
INDUS	*–7.1563e–04*	*6.0617e–02*	*–0.0118*	*0.9905805*
NOX	*–1.3602e+01*	*4.0537e+00*	*–3.3555*	*0.0007921****
AGE	*1.6953e–03*	*1.3242e–02*	*0.1280*	*0.8981255*
DIS	*–1.1782e+00*	*1.8339e–01*	*–6.4249*	*1.320e–10****
RAD	*2.9274e–01*	*6.5501e–02*	*4.4693*	*7.848e–06****
PTRATIO	*–9.7610e–01*	*1.3042e–01*	*–7.4845*	*7.172e–14****
B	*9.8041e–03*	*2.6659e–03*	*3.6776*	*0.0002354****
LSTAT	*–5.2343e–01*	*5.0249e–02*	*–10.4167*	*< 2.2e–16****
TAX	*–1.0491e–02*	*3.6223e–03*	*–2.8962*	*0.0037769****

Lambda	*0.22019*		*3.6276*	*0.00028602****
LR test value: 12.492				*0.00040864****
Wald statistic: 13.1				*0.00028602****
AIC: 3034.7				

MESS

	Estimate	*Std. Error*	*t value*	*Pr(>|t|)*
(Intercept)	*2.8457e+01*	*5.1346e+00*	*5.5421*	*4.870e–08****
CRIM	*–9.7111e–02*	*3.2922e–02*	*–2.9497*	*0.0033314****
RM	*3.8443e+00*	*4.1538e–01*	*9.2549*	*< 2.2e–16****
INDUS	*–6.7227e–04*	*6.1239e–02*	*–0.0110*	*0.9912456*
NOX	*–1.3590e+01*	*3.8410e+00*	*–3.5381*	*0.0004411****
AGE	*1.4406e–03*	*1.3186e–02*	*0.1093*	*0.91304500*
DIS	*–1.1763e+00*	*1.8379e–01*	*–6.3998*	*3.618e–10****
RAD	*2.9290e–01*	*6.6042e–02*	*4.4351*	*1.135e–05****
PTRATIO	*–9.7675e–01*	*1.2450e–01*	*–7.8453*	*2.692e–14****
B	*9.8306e–03*	*2.7004e–03*	*3.6405*	*0.0003009****
LSTAT	*–5.2282e–01*	*5.0898e–02*	*–10.2719*	*< 2.2e–16****
TAX	*–1.0484e–02*	*3.6634e–03*	*–2.8619*	*0.0043895****
Alpha	*0.24305*	*0.072939*	*–3.3323*	*0.00086147****
Implied Lambda:	*0.2157694*			
LR test value:	*12.741*			*0.00035776****

Unilateral approximation

	Estimate	*Std. Error*	*z value*	*Pr(>|z|)*
(Intercept)	*18.5128197*	*4.6841857*	*3.9522*	*7.744e–05****
CRIM	*–0.0819921*	*0.0288223*	*–2.8447*	*0.0044446****
RM	*3.9028307*	*0.3648848*	*10.6961*	*< 2.2e–16****
INDUS	*0.0118638*	*0.0534946*	*0.2218*	*0.8244892*
NOX	*–11.4646211*	*3.3830255*	*–3.3889*	*0.0007018****
AGE	*–0.0050442*	*0.0115172*	*–0.4380*	*0.6614070*
DIS	*–0.8719433*	*0.1636842*	*–5.3270*	*9.986e–08****
RAD	*0.2530190*	*0.0578713*	*4.3721*	*1.231e–05****
PTRATIO	*–0.7405545*	*0.1126720*	*–6.5727*	*4.943e–11****
B	*0.0078935*	*0.0023660*	*3.3362*	*0.0008494****
LSTAT	*–0.3659074*	*0.0461518*	*–7.9284*	*2.220e–15****
TAX	*–0.0096598*	*0.0032010*	*–3.0177*	*0.0025467****
Lambda	*0.30525*	*12.336*		*< 2.22e–16****
LR test value:	*111.79*			*2.22e–16****
Wald statistic:	*152.17*			*2.22e–16****

AIC: 2935.4

Exercise 5.4

Symmetry conditions:

$$a_1 w_{12} = a_2 w_{21}\,;\; b_1 w_{13} = a_3 w_{31};\; b_2 w_{24} = b_4 w_{42};\; b_3 w_{34} = a_4 w_{43}$$

Isotropy conditions:

$$a_i = b_i \quad \forall \mathrm{i} = 1,\ldots,4 \quad \textit{and} \quad w_{12} = w_{13};\ w_{21} = w_{24};\ w_{31} = w_{34};\ w_{43} = w_{42}$$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix};\ W = \begin{bmatrix} 0 & w_{12} & w_{13} & 0 \\ w_{21} & 0 & 0 & w_{24} \\ w_{31} & 0 & 0 & w_{34} \\ 0 & w_{42} & w_{43} & 0 \end{bmatrix};\ \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix};\ a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix};\ a_i = b_i \quad \forall \mathrm{i} = 1,\ldots,4$$

Compact expression:

$$y = a^T W y + \varepsilon$$

Exercise 5.5

Likelihood

$$L = \prod_{i=1}^{q} e^{-\varepsilon_i - \varepsilon_l} \left[1 + \alpha\,(2e^{-\varepsilon_i} - 1)(2e^{-\varepsilon_i} - 1)\right]$$

Log-likelihood

$$\begin{aligned} l &= \sum_{i=1}^{q} \ln\left\{e^{-\varepsilon_i - \varepsilon_l}\left[1 + \alpha\,(2e^{-\varepsilon_i} - 1)(2e^{-\varepsilon_l} - 1)\right]\right\} \\ &= -\sum_{i=1}^{q} \varepsilon_i - \sum_{i=1}^{q} \varepsilon_l + \sum_{i=1}^{q} \ln\left[1 + \alpha\,(2e^{-\varepsilon_i} - 1)(2e^{-\varepsilon_l} - 1)\right] \end{aligned}$$

Conditions

$$\frac{\partial}{\partial \beta}\left\{-\sum_{i=1}^{q}(y_i - X_i\beta) - \sum_{i=1}^{q}(y_l - X_l\beta) + \sum_{i=1}^{q} \ln\left[1 + \alpha\,(2e^{-(y_i - X_i\beta)} - 1)(2e^{-(y_l - X_l\beta)} - 1)\right]\right\} = 0$$

and

$$\frac{\partial^2}{\partial \beta^2}\left\{-\sum_{i=1}^{q}(y_i - X_i\beta) - \sum_{i=1}^{q}(y_l - X_l\beta) \times \sum_{i=1}^{q} \ln\left[1 + \alpha\,(2e^{-(y_i - X_i\beta)} - 1)(2e^{-(y_l - X_l\beta)} - 1)\right]\right\} < 0$$

Index

Average Direct Impact (ADI), 89
adjusted R^2, 21
aggregation matrix, 161
Akaike information criterion (AIC), 6, 22
anisotropic spatial lag model, 181–5, 196
anisotropy, 179–85, 198
approximate solutions, 168
a-spatial
  logit model, 112–15
  probit model, 112–13
asymmetric spatial relationships, 179–80, 198
asymmetry, 179–81, 198
asymptotic
  efficiency, 80
  normality, 4, 115, 123
  unbiasedness, 4
autocorrelation, 11–12, 15, 22–3, 26–7, 33, 36–9, 46–7, 49, 51, 73, 81–2, 84, 92–3, 104, 110, 120, 150–1
Average Direct Impact (ADI), 89
Average Total Impact (ATI), 89
Average Total Impact From (ATIF), 89
Average Total Impact To (ATIT), 89

bandwidth, 148–52, 154, 159, 162
Barro and Sala-i-Martin Model, 9, 15, 143, 183
Bartlett-triangular kernel, 105, 154
Bayesian,
  approach, 6, 22, 112, 127, 150, 159, 176, 203
  approach to geographically weighted regression (GWR), 150, 159
  information criterion (BIC), 6, 22, 115
best feasible generalized spatial two-stage least squares estimators, 80
best linear unbiased estimator (BLUE), 2, 22
big data, 167, 178, 198
binary models, 99, 111–18, 121, 127, 158
bi-square kernel, 149, 159
bivariate,
  coding technique, 169, 190–1, 195–8
  marginal maximum likelihood estimator, 127
bootstrap, 195, 198, 203
breast cancer risk, 179
Breusch–Pagan test of homoscedasticity, 12, 22

calibration, 149, 159
censoring, 127
chi-square distribution, 5, 182
closed-form solution, 169, 172–3, 197
Cobb–Douglas production function, 10
Cochrane–Orcutt transformation, 77, 93, 140, 141
coefficient of determination (R^2), 6
composite likelihood, 189–91
computational issues, 58, 78, 112, 120–7, 134, 141, 150, 168–78, 185, 189, 194, 198
computing time, 127, 168, 172, 189, 198
conditional probability distribution, 2, 186, 203
connectivity matrix, 28, 47
consistency, weak, 2, 4, 22, 84, 106, 123, 151
consumer choice models, 111–12
core–periphery models, 180, 198
correlation matrix, 11, 13–14, 62, 111, 198
cosine kernel, 106
count data, 111, 127
cross-sectional data, 124, 140, 144, 189
cross-validation, 149, 150, 157, 159, 162

data
  crime, Columbus, 106
  economic, 23, 129, 168
  educational achievements, Georgia, 151
  house price, Boston, 70
  luxury houses, Baltimore, 46
  production function, Munnel's, 10, 18, 131, 142, 156
  public capital productivity, 130
  regional convergence, Italy, 9, 15, 136, 143, 183
  used car price, US, 46, 63, 66, 145
dense weight matrices, 122, 168, 172, 198
determinants of crime, 106
diagnostic checking, 21
dichotomous variable, 113, 158
directional bias, 179–80, 183–5
discrete choice,
  models, 111–12, 115, 127, 158
  models in geographically weighted regression, 151, 159
distribution,
  normal, 22, 24, 34–5, 47, 53, 57–8, 76, 86, 113, 120, 122, 124–5, 134, 141
  chi-square, 5, 182
  t-Student, 5
  F-Fisher, 6
  logistic, 113–14, 124, 125, 159
duration models, 127
Durbin–Watson test, 12, 23, 33, 47

edge effects, 203
efficiency, 2, 4, 22, 76, 80, 100, 118, 120, 127, 130, 151, 182
EM algorithm, 121, 159
endogenity, 16, 56, 66, 68–9, 73, 78–9, 100, 102, 118, 120
Epanechnikov (quadratic) kernel, 105, 154, 161
estimation methods,
  best feasible generalized spatial two stage least square (BFGS2SLS), 80
  best linear unbiased estimator (BLUE), 2, 22
  feasible generalized least square (FGLS), 56, 59, 62, 66
  feasible spatial two-stage least square (FS2SLS), 102
  generalized least squares (GLS), 14–5, 18, 22–3, 26, 91, 96, 138–9, 147, 215
  generalized method of moments (GMM), 59, 78, 85, 101–4, 120–4, 134, 139–41, 155, 159, 162
  instrumental variables (IV), 17–18, 22–3, 56, 76, 93
  linearized GMM, 124
  maximum likelihood (ML), 3, 7, 14, 22, 53, 57, 83, 97, 113, 161, 167, 172, 196, 200
  method of moments (MM), 4, 18, 22
  ordinary least squares (OLS), 4, 11, 14, 16–17, 22
  pseudo-maximum likelihood, 58, 148, 189, 198
  seemingly unrelated, 203
  two-stage least square (2SLS), 17–18, 22, 66, 68, 73, 76, 78, 92, 100, 126, 141
  within, 130
exogeneity, 2, 18, 22, 68

F test, 6
feasible GS2SLS, 56, 59, 62, 66
first-order neighbors, 170, 198
Fischerian approach, 203
Fisher's information matrix, 4, 194
fixed effects, 133–5, 141, 143, 157, 159–60
frequents approach, 203

Gaussian,
  distribution, 22, 24, 34–5, 47, 53, 57–8, 76, 86, 113, 120, 122, 124–5, 134, 141
  kernel, 106
generalized
  least squares (or Aitken) estimation criterion (GLS), 91
  method of moments estimation, 4, 18, 22
  probit and logit disturbances, 123

spatial two-stage least estimation, 78, 92, 153
geographical information systems (GIS), 43, 50
geographically weighted regression, 146–51
geology, 179
geo-reference economic data, 168
Gerschgorin theorem, 52
Gibbs sampler, 127
gradient terms, 125–6, 159

HAC estimators, 104–5, 153, 158, 160
health planning, 111, 146, 168, 176
heteroscedasticity, 15, 21–2, 55, 99, 100–4, 118–22, 125, 148, 151–3, 160–1, 217
high-resolution medical images, 168
higher-order neighbors, 170, 191, 198
homogeneity, 11
house price, 70, 80, 97, 115, 127, 174, 180
human genome, 146

idempotent matrix, 17, 173
idiosyncratic, error term, 159–60
impact measures, 87–90, 92–3, 115
  ADI, 89
  ATI, 89
  ATIT, 89
  ATIF, 89
  in a Logit and in a Probit model, 115
index function, 112, 159
individual error term, 130, 132, 139
information matrix, 4, 7, 22, 83, 115, 194
instrumental variables, 17
instruments, 17
isotropy test, 179–83
iterative procedure, 121, 135, 138
iterative search, 149, 150, 155

Jarque–Bera test, 8

kernel function,
  Bartlett-triangular, 105, 154
  bi-square, 149, 154
  cosine, 106
  Epanechnikov (quadratic), 105, 154
  estimation, 158
  Gaussian, 106, 149, 157
  Partzen, 106
  quadratic (bi-weight), 105, 154
  quadratic (triweight), 106
  quadratic spectral, 106, 154
  tri-cube, 149
  Tukey–Hanning, 106, 154
  uniform, 105
KKP model, 15–16, 132–4, 139, 157, 159
Kronecker product, 132, 159
kurtosis, 8

lagged independent variables model, 52, 93
Lagrange Multiplier (LM), test procedure, 7, 12, 35, 83, 92
land cover assessment, 168
large sample, 4, 58, 76, 78, 80, 86, 112, 145, 150, 187, 195, 198
latent regression, 112, 121, 154, 158
latent variable, 112, 121–2, 154, 158
lattice grid, 80, 100
least square method, 4, 11, 14, 16–17, 22
Lee Instrumental Variable estimators, 78, 92, 153
Likelihood factorization, 186
Likelihood function, 7, 14, 53, 57, 83, 97, 113, 119–22, 133, 137, 161, 167, 172, 196, 200
Likelihood ratio test, 7, 22, 183
line data, 203
linearized GMM estimator, 124
locally weighted regression, 146, 113–14, 124–5, 159
logistic distribution, 129, 159
long panels, 129, 158

marginal effects, 87–90, 92–3, 115, 194
matrix,
  exponential spatial specification, 169–75, 190, 196, 198–9
  exponential transformation, 169
  power expansion, 170

maximum distance neighborhood criterion, 28, 47, 118
Maximum likelihood,
  estimation method 3, 7, 14, 22, 53, 57, 83, 97, 113, 161, 167, 172, 196, 200
  solution for SARAR model paramters estimation, 76
  solution for Spatial Error model parameters estimation, 57
  solution for Spatial Lag model parameters estimation, 66
medical images, 168
meteorology, 179
  method of moments estimation method, 4, 18, 22
misspecification, 100, 145
models,
  anisotropic spatial lag, 181–5, 196
  geographically weighted regression (GWR), 146–51
  KKP, 133
  locally weighted regression, 146
  logit, 112
  panel, 129
  probit, 113
  pure spatial autoregression, 52
  SARAR with heteroscedastic disturbance, 100
  SARAR, 75
  spatial error model (SEM), 55
  spatial error with fixed effects, 141
  spatial error with random effects, 132
  spatial HAC, 104
  spatial lag model (SLM), 66
  spatial lag with fixed effects, 135
  spatial lag with random effects, 137
  spatial logit, 124
  spatial probit, 127
  spatially lagged independent variable, 55
  unilateral spatial lag, 185
Monte Carlo Markov Chain (MCMC), 127
Moran's I,
  modified test, 84–6
  moments of, 34–5
  scatterplot, 49
  test for residual spatial autocorrelation, 33–5, 46–7, 82–3
moving windows, 146
multivariate,
  normal distribution, 119, 186
  probit, 127

nearest neighbor, criterion, 28, 47, 197, 199
neighborhood criteria,
  higher-order, 170, 175
  maximum distance, 28–9
  queen's case, 27–8, 40, 47
  rook's case, 27, 40, 47, 128, 180–1
non-linear models, 122, 124, 167
non-nested test, 203
non-parametric estimation, 100–4
non-stationary spatial relationships, 145
normal distribution, 22, 24, 34–5, 47, 53, 57–8, 76, 86, 113, 120, 122, 124–5, 134, 141
nuisance parameters, 101, 158, 160

Okun's Law, 35, 87, 90
Ord's decomposition, 58
Ordinary Least Squares estimation method, 4, 11, 14, 16–17, 22

Pairwise likelihood, 189, 193, 198
Panel,
  data, 128–9
  long, 129, 158
  short, 129
parameter of interest, 101
parameter space, 96
Partzen kernel, 106
Phillips curve, 37
point data, 203
point pattern analysis, 203
pooled time-series, 129
power series expansion, 170
predecessors-neighbors, 186–7
preferred direction, 179, 181, 185–6
pre-testing, 203
price index, 37–8
properties of estimators,
  asymptotic,
    consistency, 2, 4, 22, 106, 123, 133, 151

efficiency, 80
normality, 4, 18, 34, 58, 115, 123
unbiasedness, 4
small sample,
efficiency, 76, 80, 101–2, 120, 130, 134–6
unbiasedness, 4, 12, 22, 194
pseudo-likelihood, 5, 148, 189, 198
pure spatial autoregression, 52

quadratic kernel,
bi-weight, 105, 154
spectral, 106, 154
triweight, 106
queen's case neighborhood, 27, 40, 47, 128, 180–1

R^2 (coefficient of determination), 21
random,
effects, 130, 132–7, 140–1, 156, 159–62
fields, 146, 202
Rao score test (Lagrange multiplier), 7
reduced form of a model, 88–9, 118, 122, 124, 158
regional convergence, 9, 15, 136, 143, 183
relevant instruments, 17
resampling, 195, 198
robust,
estimation, 100, 134, 155, 203
Lagrange multiplier test, 84, 92–4
rook's case, 27, 40, 47, 128, 180–1

sample,
analogue, 123
large, 4, 58, 78, 80, 86, 167–8
moment, 102, 104, 123, 141
selection, 127
scan statistics, 146
Schwartz information criterion, 6, 22, 115
score function, 3, 7, 22, 83, 114
seemingly unrelated regression, 203
skewness, 8
social,
distances, 28
networks, 172, 203
software,
GeoDA, 204
Matlab, 204
PySAL, 204
R, 20
SPSS, 27
STATA, 204
sparse weight matrices, 167–8, 172, 198
spatial
autocorrelation, 11–12, 15, 22–3, 26–7, 33, 36–9, 46–7, 49, 51, 73, 81–2, 84, 92–3, 104, 110, 120, 150–1
bootstrap, 195, 198
econometric models *see* models
expansion methods, 203
HAC procedures, 104–5, 153, 158, 160
interaction models, 111, 203
lag operator, 29
lag, 26, 29
quantile regression, 203
resampling, 195, 198
statistics, 27, 179, 203
weight matrix, 26, 132, 175
spatially lagged variable, 29
spherical disturbances, 2, 11, 13, 119–20
spillover, 133, 194
stable dynamics spatial panels, 142, 159
standardized,
logistic distribution, 113, 159
normal distribution, 86, 113
weight matrix, 29, 32, 34, 47–8, 51–2, 94, 148, 167, 169, 179, 181, 186
storage limitations, 168, 174, 189, 198
structural form of a model, 118, 159
subsample, 146, 149, 195, 197
sum of squares, 6
systems of non-linear equations, 4–6, 18, 180, 193, 199

T test, 5
Taylor series, 7
tests,
Breusch–Pagan, 12, 22
chi square, 12, 23, 33, 47

tests – *continued*
Durbin–Watson, 12, 23, 33, 37
F-Fisher, 6
Jarque–Bera, 8
likelihood ratio, 7, 22, 183
modified Moran, 84–6
Moran, 33–5, 46–7, 82–3
of isotropy, 179–83
Rao, 7
T-student, 5
Wald, 7, 115, 153
time series models, 12, 23, 29, 33, 99, 104, 142, 159, 170, 184, 186, 189, 202
time-demeaning, 134–5, 137, 139–40, 159–60
tri-cube kernel, 149
truncation, 172–4
Tukey–Hanning kernel, 106, 154
two-stage least square (2SLS), 17–18, 22, 66, 68, 73, 76, 78, 92, 100, 126, 141

unbiasedness, 4
uniform kernel, 105
unilateral approximation, 169, 178, 185, 196–8
unilateral spatial lag model, 185–6, 198–9
unobserved heterogeneity, 129, 132, 159
unobserved variable, 112
unstable dynamics spatial panels, 142, 159

variance-covariance matrix, 11–14, 26, 57–8, 62, 97, 100, 102–5, 118–19, 121, 124, 133, 137–8, 141, 153, 169–70, 174, 199

Wald test, 7, 115, 153
weak consistency, 2, 194
weight,
standardized *see* standardized weights
matrix *see* spatial weight matrix
White test for homoscedasticity, 12
within estimator, 130, 141
within residuals, 159

Zellner–Revankar production, 10, 18