Advances in Spatial Science

Titles in the Series

H. Eskelinen and *F. Snickars* (Eds.)
Competitive European Peripheries
VIII, 271 pages. 1995. ISBN 3-540-60211-9

A. Nagurney and *S. Siokos*
Financial Networks
XVI, 492 pages. 1997. ISBN 3-540-63116-X

M. M. Fischer and *A. Getis* (Eds.)
Recent Developments in Spatial Analysis
X, 434 pages. 1997. ISBN 3-540-63180-1

P. McCann
The Economics of Industrial Location
XII, 228 pages. 1998. ISBN 3-540-64586-1

R. Capello, P. Nijkamp and *G. Pepping* (Eds.)
Sustainable Cities and Energy Policies
XI, 282 pages. 1999. ISBN 3-540-64805-4

M. M. Fischer, L. Suarez-Villa and *M. Steiner* (Eds.)
Innovation, Networks and Localities
XI, 336 pages. 1999. ISBN 3-540-65853-X

J. Stillwell, S. Geertman and *S. Openshaw* (Eds.)
Geographical Information and Planning
X, 454 pages. 1999. ISBN 3-540-65902-1

G. Clarke and *M. Madden* (Eds.)
Regional Science in Business
VIII, 363 pages. 2001. ISBN 3-540-41780-X

M. M. Fischer and *Y. Leung* (Eds.)
GeoComputational Modelling
XII, 279 pages. 2001. ISBN 3-540-41968-3

M. M. Fischer and *J. Fröhlich* (Eds.)
Knowledge, Complexity and Innovation Systems
XII, 477 pages. 2001. ISBN 3-540-41969-1

M. M. Fischer, J. Revilla Diez and *F. Snickars*
Metropolitan Innovation Systems
VIII, 270 pages. 2001. ISBN 3-540-41967-5

L. Lundqvist and *L.-G. Mattsson* (Eds.)
National Transport Models
VIII, 202 pages. 2002. ISBN 3-540-42426-1

J. R. Cuadrado-Roura and *M. Parellada* (Eds.)
Regional Convergence in the European Union
VIII, 368 pages. 2002. ISBN 3-540-43242-6

G. J. D. Hewings, M. Sonis and *D. Boyce* (Eds.)
Trade, Networks and Hierarchies
XI, 467 pages. 2002. ISBN 3-540-43087-3

G. Atalik and *M. M. Fischer* (Eds.)
Regional Development Reconsidered
X, 220 pages. 2002. ISBN 3-540-43610-3

Z. J. Acs, H. L. F. de Groot and *P. Nijkamp* (Eds.)
The Emergence of the Knowledge Economy
VII, 388 pages. 2002. ISBN 3-540-43722-3

R. J. Stimson, R. R. Stough and *B. H. Roberts*
Regional Economic Development
X, 397 pages. 2002. ISBN 3-540-43731-2

S. Geertman and *J. Stillwell* (Eds.)
Planning Support Systems in Practice
XII, 578 pages. 2003. ISBN 3-540-43719-3

B. Fingleton (Ed.)
European Regional Growth
VIII, 435 pages. 2003. ISBN 3-540-00366-5

T. Puu
Mathematical Location and Land Use Theory,
2nd Edition
X, 362 pages. 2003. ISBN 3-540-00931-0

J. Bröcker, D. Dohse and *R. Soltwedel* (Eds.)
Innovation Clusters and Interregional Competition
VIII, 409 pages. 2003. ISBN 3-540-00999-X

D. A. Griffith
Spatial Autocorrelation and Spatial Filtering
XIV, 247 pages. 2003. ISBN 3-540-00932-9

J. R. Roy
Spatial Interaction Modelling
X, 239 pages. 2004. ISBN 3-540-20528-4

M. Beuthe, V. Himanen,
A. Reggiani and *L. Zamparini* (Eds.)
Transport Developments and Innovations
in an Evolving World
XIV, 346 pages. 2004. ISBN 3-540-00961-2

Y. Okuyama and *S. E. Chang* (Eds.)
Modeling Spatial and Economic Impacts
of Disasters
X, 323 pages. 2004. ISBN 3-540-21449-6

L. Anselin, R.J.G.M. Florax and *S. J. Rey*
Advances in Spatial Econometrics
XXII, 513 pages. 2004. ISBN 3-540-43729-0

R.J.G.M. Florax and *D. A. Plane* (Eds.)
Fifty Years of Regional Science
VIII, 400 pages. 2004. ISBN 3-540-22361-4

D. Felsenstein and *B.A. Portnov* (Eds.)
Regional Disparities in Small Countries
VI, 333 pages. 2005. ISBN 3-540-24303-8

A. Reggiani and *L.A. Schintler* (Eds.)
Methods and Models in Transport
and Telecommunications
XIII, 364 pages. 2005. ISBN 3-540-25859-0

Giuseppe Arbia

Spatial Econometrics

Statistical Foundations and Applications to Regional Convergence

With 19 Figures
and 11 Tables

Springer

Professor Giuseppe Arbia PH D Cantab.

Department of Business, Statistical,
Technological and Environmental Sciences,
University "G. d'Annunzio" of Chieti
Viale Pindaro, 42; I65127, Italy
also Professor at LUISS "Guido Carli" University of Rome (IT)
and at the USI University of Lugano (CH)
E-mail: arbia@unich.it
URL address: http://www.unich.it/~arbia/

Cataloging-in-Publication Data

Library of Congress Control Number: 2006921995

ISBN-10 3-540-32304-X Springer Berlin Heidelberg New York
ISBN-13 978-3-540-32304-4 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in Germany

Cover design: Erich Kirchner
Production: Helmut Petri
Printing: Strauss Offsetdruck

SPIN 11672227 Printed on acid-free paper – 88/3153 – 5 4 3 2 1 0

To Francesco and Enrica

Preface

In recent years the so-called *new economic geography* and the issue of regional economic convergence have increasingly drawn the interest of economists to the empirical analysis of regional and spatial data. However, even if the methodology for econometric treatment of spatial data is well developed, there does not exist a textbook theoretically grounded, well motivated and easily accessible to economists who are not specialists. Spatial econometric techniques receive little or no attention in the major econometric textbooks. Very occasionally the standard econometric textbooks devote a few paragraphs to the subject, but most of them simply ignore the subject. On the other hand spatial econometric books (such as Anselin, 1988 or Anselin, Florax and Rey, 2004) provide comprehensive and exhaustive treatments of the topic, but are not always easily accessible for people whose main degree is not in quantitative economics or statistics.

This book aims at bridging the gap between economic theory and spatial statistical methods. It starts by strongly motivating the reader towards the problem with examples based on real data, then provides a rigorous treatment, founded on stochastic fields theory, of the basic spatial linear model, and finally discusses the simpler cases of violation of the classical regression assumptions that occur when dealing with spatial data. I am convinced that, once the reader is introduced to the probabilistic and statistical arguments on which the basic linear model is grounded, he will be able to understand quite straightforwardly also the more sophisticated models and techniques that are present in the spatial econometric literature. A review of some more advanced topics excluded from the range of interest of this book, is confined to a final chapter which provides the references for further, more in-depth going, studies.

The project of a self-contained statistically based book on spatial econometrics dates back to 1996 when I first taught a course on advanced econometric topics in the faculty of Economics at the "G. D'Annunzio" University of Pescara. The series of lecture notes that I drafted in Italian were printed in a provisional working paper series and constitute the backbone of the present volume. This preliminary version underwent several changes and integrations when it was used in a series of post-graduate courses that I have been teaching at the faculty of Economics of Rome "Tor Vergata" since 1997. The typical student attending these courses was a Master or Doctorate student with a first degree in Economics, Political Sciences or one of the other Social Sciences.

Having in mind such a typology of student during the drafting of the book, the only pre-requisites for reading this monograph are a sound first course in Probabil-

ity at the level of books like Grimmett and Stirzaker (1994) and a course in Statistics at a level of books like Azzalini (1996) or Mood *et al.* (1974). A prior knowledge of the basic time series analysis methods at a level of the first five chapters of Hamilton (1994) helps, because there is a certain analogy between spatial and time methodology.

This volume could be used as a textbook for a post-graduate introductory course of around 20/30 hours, or as a reference book for post-graduate students engaged in Ph. D. work on quantitative economics (or other social sciences), involving spatial econometric estimation problems.

Having outlined the general aim of this book, described its genesis and its potential readers, I am now happy to fulfil the pleasant duty of thanking the many people who contributed to its preparation and final realization. This is the best moment when writing a book (as anyone who has a direct experience knows only too well!). There are at least two reasons for this. The first is that these are in fact the last words written by the author in this context, which means that the heavy work is finally done and that what in the last phase can only be described as an "obsession" will eventually give way to other (hopefully less compelling) tasks. The second is that this is the only place where one can talk about oneself and one's working environment and family and not merely about abstract concepts and formalism and this is an inevitable source of self-satisfaction for all authors.

It would have been almost impossible to realize this book without the help and support I received from many people. I would like to start by thanking Jean Paelinck for being the first author I read on the subject back in 1984 when I was a Ph. D. student at Cambridge University and, after knowing him in person, for having become a constant reference for my work on this subject. He must be thanked specifically for his constant help, the useful corrections and suggestions he provided on various drafts of the book and, most of all, for his friendship. Thanks are also due to Gianfranco Piras for his daily assistance over this last year with all the technical and non-technical problems arising during the volume's preparation, as well as for having co-authored Chapter 6 and the Appendix. My thanks also to Francesco Moscone of London School of Economics for his enthusiasm and encouragement from the beginning of this project, to Giovanni Lafratta and Paolo Postiglione (both from the "D'Annunzio" University of Pescara) for having co-authored Sections 2.4.2.10 and 4.4.2.3, respectively, and to Elisa Tosetti of Cambridge University for having read the manuscript at various stages and for having provided corrections and most useful suggestions and additions to the original text. Needless to say that I alone am wholly responsible for any errors remaining in the writing and calculations. I have also to thank Roberto Basile of Istituto di Studi ed Analisi Economica (ISAE) and Laura De Dominicis of "Tor Vergata" University and of the Free University of Amsterdam for their help in some of the computations reported in Section 1.3 and Chapter 5, as well as for revising parts of the book. I also wish to thank Robert Haining for providing hospitality at the Cambridge Department of Geography during the summer of 2004 when I was writing the final draft. Finally I want to thank all my PhD students for

their patience and support in the whole period when I could devote less time to their research and the actors in the theatre company "Palcoscenico '95" who had to do without me on many occasions during this period of intense writing.

A very special mention is due to my beloved parents who both died during the last year when I was working on the final version of this book. I have already dedicated my second book to them, but I feel I am particularly indebt to them. Apart from the immense gratitude I feel towards them for all the aspects of my life, I would like to state quite specifically that my parent's presence, love and constant encouragement to aim ever higher have made all the difference to my work and professional life as a whole. It will not be the same without them.

It is common for book prefaces to contain expressions of thanks to the author's family, wife and children for having allowed him to devote himself (almost) entirely to the burden of writing a book for a long period (during the working and non-working hours, during the weeks and the weekends) and for having provided him with the right environment to fulfil this task. This is doubtless desirable, but I do not believe it corresponds entirely to what happens in practice. Most of the time the various members of the family have no choice and, if they did they would rather have one book less on the shelf and more time to spend in a carefree fashion with their husband/father, more of his patience, more of his help in relation to their own tasks. Having said that, in this very moment when I am writing these that will (hopefully) be the final words to this book, my three kids are joyfully shouting in the neighbouring room and (even if I am tempted to join them) they are not providing exactly what one defines as the "right environment" for a writer. Nonetheless, this book is dedicated to my wonderful family and, in particular, to Francesco and Enrica since I dedicated my first book to Paola and Elisa before my two younger children were born. It is difficult to say why one deliberately decides to devote time to writing a book at the expense of his family. It might be for a pure love of knowledge, its spread or perhaps the fallacious illusion of leaving something to posterity, or plane madness, or again just self-satisfaction or any combination of the above. But, after all, the positiveness of our actions and the meaning of any single moment is not in our hands: these are things that just happen, as an undeserved gift. Like the one that we are celebrating this very day as we do every year.

Rome, Christmas Day 2005 Giuseppe Arbia

Proem

't is distance lends enchantment to the view …
William Wordsworth

While exerting the much valued privilege of reading the manuscript to this book, I was brought back in time – and in space…- to the early fifties when I started my econometric work. It was like rediscovering the path-breaking demand analyses by René Roy (Econometrica, 1947), then wading through Harald Cramer's *Mathematical Methods of Statistics*, Sixth Printing, 1954, to follow on with Johnston's *Econometric Methods*, 1963, and end up with some recent article on Almost Ideal Demand Systems.

Such is indeed the route of Giuseppe Arbia's book, which starts with an empirical spatial problem, showing the need for a rigorous analytical framework, working through a simplified but operational version of this, and then pointing out that for both the fundamental models, the assumptions are largely negated by spatial data and their appropriate representations, and this from all three of the probability model, the statistical generation process and the sampling points of view.

This synthetic presentation of the spatial model must be considered as particularly helpful to the reader wishing to initiate himself to spatial econometric models. Especially the fact that the analysis is conducted starting from the fundamental notion of random fields is important, as in this way the axiomatic basis is laid for any further regional analysis.

No wonder, thus, that the regional model, initially presented, had to be adapted from different angles, as a simple homogeneous approach could not cover the full complexity of spatial data.

These, indeed, show such a complex pattern, in part induced by the fact that they are fundamentally biased, as I have shown elsewhere, hence that apparent heterogeneity. In fact, from the advanced techniques proposed in the last chapter, perhaps the exploratory tools should be given special attention, and this not only for the crude variables – which already show more often than not hot spots, as spatial statisticians have repeatedly demonstrated – but also for the residuals obtained from a first – even duly spatialised – analysis (the "doggy-bag principle", as I like to call it: never throw away your leftovers…). I think indeed that spatial interdependencies and externalities are more involved than could be covered with spatialised multivariate Markov or even Yule schemes. How to model appropriately, i. e. to specify, such complex inter-linkages is, I think, one of the main challenges for future spatial econometric work.

Giuseppe Arbia's book is another testimony to the vitality of spatial econometric analysis through the last couple of decades. It goes hand in hand with the realisation by general economists that pre-geographic space should be an essential component of every economic analysis, which could not be ignorant of topological elements in its search for propositions of economic behaviour; as I use to put it friendly to my students: "space matters"! But on the other hand, fundamental contributions by general economists should be taken into account to allow spatial econometricians to improve on their theoretical specifications, as hinted at higher up. Indeed, the mathematical structures used in setting up a spatial model are uttermost germane to the use it will be put to; but I suspect that that validity is more general, as it refers to the economic processes the spatial econometrician wants to picture.

What I have tried to convey to the reader of this book, is a sense of its importance as a path-finder for young research workers, not necessarily in the field of spatial econometrics proper, but also in other fields of economic analysis, theoretical and applied: a price is not just a p, but a stochastic variable in a two-dimensional random field, in other words it has two-dimensional statistical distribution. Difficult at first to imagine but that is the way things are; our task, like the physicists' one, is to create beauty from garbage …

Lectori salutem.

January 2006 Jean H.P. Paelinck

Post-scriptum and Afterthought

The price random field mentioned above has been known to spatial observers for a long time; less well known is the distribution of spatial regimes, i. e. of behavioural parameters. In the last application chapter of this book it has been shown that Northern and Southern Italy could be approached as separate spatial (macro-) entities, and other instances have been studied: Northern and Southern Belgium, Canada and the United States (the so called "Continental Divide"). In Paelinck and Klaassen, *Spatial Econometrics*, 1979, it has been suggested that, given the nature of spatial patterns, a fuzzy subset approach might be appropriate; indeed, it is very probable that spatial regimes are not separate in any clear-cut fashion, but "overlap"' in a fuzzy sense, and that same idea might apply not only to parameters of a given behavioural pattern, but also to different "varieties" of those patterns, i. e. patterns based on different theoretical assumptions. In still other words, binary coefficients separating the regimes might themselves be fuzzy. The next step is to return to basics, and develop a theory of fuzzy random fields. *Ars longa, vita brevis …*

Table of Contents

“The hidden harmony is more powerful than the acclaimed one.”
Heraclitus, 500 B.C.

1 Motivation

1.1 Introduction

Until the end of the last century spatial econometric methods were like the six Pirandellian characters in search of an author (Pirandello, 1921). In the great Italian playwright's surreal drama, six characters wandered around in the scene desperately seeking an author who could explain to them what they had to do and give them a reason to live. The dramatic situation lay in the fact that they knew exactly what they had to do, but they did not know *why* they had to do it! Similarly, until a few years ago, spatial econometric methods were well developed in the literature, but the drama was that no one used them in the mainstream applied economic analysis!

Historically, spatial econometric methods stem directly from the twentieth-century developments in the statistical literature designed to give consideration to the problem of the violation of the classical sampling model (the urn paradigm) with a big emphasis on similarities caused by spatial proximity. These developments were necessary to provide the right environment for explaining spatial diffusion phenomena frequently encountered in many applied fields such as epidemiology, geography, agricultural studies, geology, image analysis, regional sciences, astronomy, archaeology and many others (see Haining, 2003 for a review).

The spatial statistical techniques providing the basis for spatial econometrics go back about half a century and can be conventionally dated to Peter Whittle's seminal 1954 paper (Whittle, 1954) followed by other important contributions by the same author (Whittle, 1962; 1963), by Bartlett (1963; 1975) and by Besag (1974) amongst the others. The main results obtained led to a first codification in the Seventies with important publications such as the celebrated books by Cliff and Ord (1973) and Bennett (1979). Other well-established textbooks followed in the Eighties (Ripley, 1981; 1988; Upton and Fingleton, 1985; 1989; Griffith, 1988; Arbia, 1989 amongst others), the Nineties (Haining, 1990; Cressie, 1991) and at the beginning of the new century (Haining, 2003).

The term "*spatial econometrics*" was coined by Jean Paelinck in the late Seventies during the general address he delivered to the annual meeting of the Dutch Statistical Association in May 1974 (see Paelinck and Klaassen, 1979), although the author himself suggests that the idea had already appeared at a previous conference (see Paelinck, 1967). In the foreword to the first book entirely devoted to the subject (Paelinck and Klaassen, 1979) it is suggested that the new branch of econometrics is a "blend of economic theory, mathematical formalisation and mathematical statistics" (page vii). There are five fundamental characteristics of

the new field that can be found in the Paelinck and Klaassen's seminal book (Paelinck and Klaassen, 1979), namely, (i) the role of spatial interdependence in spatial models, (ii) the asymmetry of spatial relations, (iii) the importance of explanatory factors located in other spaces, (iv) the differentiation between ex-post and ex-ante interaction, and (v) the explicit modelling of space. It is important to note that some of the significant contributions to regional economics (such as those well summarised in standard books like Isard (1960), Paelinck and Nijkamp (1975), Klaassen *et al.* (1979)), in this period deal mainly with the phase of specification placing a great emphasis on the underlying economic theory, rather than the phase of statistical estimation and hypothesis testing. When examples are reported of statistical estimation these are limited to the application of existing methods (OLS and ML essentially) to the case examined.

Apart from some remarkable examples of the possibility of using spatial statistical methods in economics (like that of the Nobel laureate Clive Granger, in the Sixties and Seventies; see Granger, 1969, 1974), the book published by Luc Anselin in the late Eighties (Anselin, 1988), certainly constitutes an important step forward in the historical development of the discipline. Collecting previous contributions, here for the first time the author presents a comprehensive treatment of topics, such as spatial dependence and spatial heterogeneity, that are fundamental in the analysis of spatial economic data as will be shown later in this book. He defines the subject as "the collection of techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models" (Anselin, 1988, p. 7).

Thus at the end of the Eighties the pioneering phase of the discipline seemed to have been completed. The methods were there, waiting for someone to use them, yet this did not happen! The Pirandellian drama!

Indeed, it was not until the years bridging the two millennia that mainstream econometrics expressed a growing interest in spatial statistical methods: an interest witnessed by the increasing number of papers referring to important spatial problems appearing in recent econometric and applied economic journals. Florax and De Vlist (2003) and Anselin *et al.* (2004) review these papers thoroughly and survey 11 articles in econometric journals and 30 in applied economic journals for the period after the year 2000 alone!

The integration of spatial methods with econometrics nevertheless remains in an early phase and still lags far behind that experienced by time series methods in the Seventies after the path-breaking book by Box and Jenkins (1970). In fact, while it is true that no standard econometric textbook can ignore the time series methodologies, the treatment of spatial methods is conversely still very rare and when it does occur no more than a few paragraphs are devoted to the subject. No mention is made, for instance, in some of the most recent introductory textbooks such as Baltagi (1999), Berndt (1991), Davidson (2000), Davidson and MacKinnon (1993), Dougherty (2002), Goldberg (1998), Gourieroux and Montfort (1995), Greene (2003), Hayashi (2000), Hendry and Morgan (1995), Holly and Weale

(2000), Johnston and Dinardo (1997), Judge *et al.* (1988), Ruud (2000), Seddighi *et al.* (2000), Spanos (1986), Stewart and Gill (1998), Stock and Watson (2003), Thomas (1997), Verbeek (2000) and Woolridge (2002b).

Remarkable exceptions in this sense are the books by Johnston (1991), Kmenta (1997), Maddala (2001), Baltagi (2001), Woolridge (2002a), Gujarati (2003) and Kennedy (2003).

Johnston (1991) devotes few lines to spatial problems pointing out that: "The above exposition has implicitly assumed a temporal, or time-series framework, but the same phenomenon may arise with cross-section data, where it is often referred to as *spatial autocorrelation*." (Johnston, 1991; p. 305).

Kmenta (1997) also acknowledges the problem of the non-independence of statistical observations in space, stating that: "In many circumstances the most questionable assumption of the preceding model is that the cross-sectional units are mutually independent. For instance, when the cross-sectional units are geographical regions with arbitrarily drawn boundaries – such as the states of the United States – we would not expect this assumption to be well satisfied." (page 512). No mention is made of possible solutions to this problem, however.

Maddala (2001) only briefly mentions to the problem of spatial dependence amongst contiguous residuals of a linear regression: "There are two situations under which error terms in the regression model can be correlated. In cross-section data it can arise among contiguous units. For instance, if we are studying the consumption pattern of households, the error terms for households in the same neighbourhood can be correlated. This is because the error terms pick up the effects of omitted variables and these variables tend to be correlated for households in the same neighbourhood (because of the "keeping up with the Jones" effect). Similarly, if our data are on states, the error terms for contiguous states tend to be correlated. All these examples fall in the category of *spatial correlation*." (see Maddala, 2001, p. 228).

The second edition of Baltagi's well known textbook on panel data includes a short discussion of the problems generated by treating spatial panels. The author states that "When one starts looking at a cross-section of countries, regions, states etc. these aggregate units are likely to exhibit cross-sectional correlation that has to be dealt with. There is an extensive literature using spatial statistics which deals with this type of correlation … Spatial dependence models may use a metric of economic distance which provides cross-sectional data with a structure similar to that provided by the time index in a time series". (Baltagi, 2001; pp. 195-197).

Gujarati presents the problem only briefly by remarking that: "if by chance such a correlation is observed in cross sectional units, it is called spatial autocorrelation, that is correlation in space rather than over time" (Gujarati, 2003; p. 441).

Kennedy (2003) warns the reader that: "not all instances of autocorrelated errors relate to time series data […]. Spatial econometrics refers to analysis of spatial data for which errors are correlated with errors associated with nearby regions. This type of non-sphericalness is referred to as spatial correlation".

Finally Woolridge (2002a) devotes a mention to the issue of spatial dependence in the very first pages of his book: "A situation that does requires special consideration occurs when cross section observations are not independent of one another. An example is *spatial correlation* models. This situation arises when dealing with large geographical units that cannot be assumed to be independent draws from a large population, such as the 50 states in the United States. It is reasonable to expect that the unemployment rate in one state is correlated with the unemployment in neighboring states. While statistical estimation methods – such as Ordinary Least Squares or Two Stages Least Squares – can usually be applied in these cases, the asymptotic theory needs to be altered. Key statistics often (although not always) need to be modified […]. For better or worse, spatial correlation is often ignored in applied work because correcting the problem can be difficult" (Woolridge, 2002; p. 6). This textbook also contains a short section that develops the idea a bit more thoroughly when dealing with the various forms of dependence amongst residuals "As the previous subsection suggests, cross-section data that are not the result of independent sampling can be difficult to handle. *Spatial correlation*, or, more generally, *spatial dependence*, typically occurs when cross-section units are large relative to the population, such as when data are collected at the county, state, province, or country level. Outcomes from adjacent units are likely to be correlated." (Woolridge, 2002a; p. 134).

The only problem mentioned in all these quotes is that of spatial correlation (a concept that will be treated at length in this book), but many other important concepts of spatial econometrics are ignored. More importantly, there is no discussion of the consequences of the problem on statistical estimation and hypothesis testing, of how to test hypotheses related to spatial relationships and of how to remove the emerging problems and derive correct inferential conclusions.

As a consequence of such lacunae in the standard econometric textbook literature the author feels obliged to motivate the reader towards these topics (an unusual task when dealing with other typologies of economic data, such as time series, cross-sections or panel data) by showing the emergence of the need for a proper modelling framework to analyse spatial data. Such is the purpose of this book's first, introductory, chapter.

An important event in the history of spatial econometrics is certainly the advent of the now-celebrated "new economic geography" theories, starting with Krugman's seminal "Gaston Eyskens lectures" in Louvain in 1990 (then published in Krugman, 1991; see also Fujta *et al.*, 1999). In fact these lectures provide a theoretical framework justifying a spatial analysis of economic data when approaching issues like regional convergence, regional concentration of economic activities and adjustment dynamics. Many of these theories have led to the specification of models that are susceptible to empirical validation and that have shown the importance of developing specific econometric tools for data distributed in space. In the remainder of this chapter we will review some of these models concentrating, in particular, on the regional convergence models (Barro and Sala-i-Martin, 1995) as these are a very helpful example for introducing the basic concepts of spatial econometrics.

In order to motivate the reader in his/her study of spatial econometric methods, we will first discuss the β-convergence model at some length (Section 1.2). We will then consider some empirical studies on regional convergence in the European Union and use them to show some of the incongruences and black holes left in this particular instance by the classical linear regression model (Section 1.3). Finally, we will conclude this introductory chapter by reviewing a list of topics that have been excluded from the volume and by presenting the latter's overall plan (Section 1.4).

1.2 Theoretical Economic Models Calling for Spatial Econometric Techniques

1.2.1 Introduction

Until a few years ago spatial statistics was a topic outside applied economists' range of interest. In recent years, however, there has been a flourishing of economic studies that, by developing theoretical models that involve relationships between variables observed across countries or regions, have greatly stimulated interest in the measurement and statistical modelling of spatial variables. Most of these models are linked to the developments in the so-called *new economic geography* (Krugman, 1991, Fujta *et al.*, 1999; Krugman and Venables, 1995; Ottaviano and Puga, 1998; Puga and Venables, 1997; 1999; Durlauf and Quah, 1999) and to economists' renewed interest in problems related to economic growth and the conditions under which the per-capita income levels of various regions tend to converge over time. This topic is currently of great interest especially in relation to the recent European debate. Indeed, in many respects, one of the main reasons for a European Union and for agreement on economic policies is the goal of reducing disparities in welfare distribution. One important aspect of such a goal concerns the reduction of disparities between growth rates of per-capita income levels, on the basis that this would, in the long run, guarantee a reduction of welfare disparities. The analysis of the dispersion of regional incomes is often considered as a proxy of the inequality in personal incomes distribution and is used as an indicator in economic policy debates (see EU, 2005). As a consequence, substantial empirical research efforts have been devoted to this topic by European scientists (as witnessed by books like Fingleton, 2003 for example).

In this section we will briefly review some theoretical models that have been introduced to formalize the idea of *regional convergence* and provide a rigorous basis for the empirical testing of such an hypothesis. In particular, in Section 1.2.2 we will concentrate on the popular *β-convergence* model (see Barro and Sala-i-Martin, 1995 for a review). This model will be used throughout the book to illustrate the various spatial econometric techniques. The reason for this choice is two-

fold. In the first place, such a model has been the most popular in the field of regional convergence and indubitably the one that has experienced the widest use of spatial econometric techniques in the literature (see e. g. Fingleton, 2003; Anselin *et al.,* 2004). Secondly, it represents a very simple and intuitively appealing example that provides a good basis both for illustrating the peculiarities of the various problems raised by a regression based on spatial data and for describing the various solutions offered by spatial econometric techniques.

1.2.2 The β-convergence Approach

The β-convergence approach was developed by Mankiw *et al.* (1992) and Barro and Sala-i-Martin (1995) starting from the earlier contributions of Solow (1956) and Swan (1956), and, from an economic-theoretic point of view it is considered one of the most convincing for explaining the economic convergence of per-capita income. The model was originally conceived to explain the differences between different national income levels, but indeed it discovered a stronger *raison d'être* when used to explain differentials between regional income within the same country or group of countries. In fact, the neoclassical growth model predicts conditional convergence to the one steady-state[1] of economies characterized by the same structural parameters in terms of tastes, endowments and technologies. However, the assumption of uniform structural conditions can be accepted more easily when referred to regions within a country rather than between countries, as is acknowledged, for instance, by Barro and Sala-i-Martin (1995) when they state: "although differences in technologies, preferences and institutions do exist across regions, these differences are likely to be smaller than those across countries (Barro and Sala-i-Martin, 1995; p. 382). They also explicitly remark: "Absolute convergence is more likely to apply across regions within countries than across countries".

The theoretical framework takes its moves from the neo-classical Solow-Swan model of long-run growth (Solow, 1956; Swan, 1956). The model assumes exogenous saving rates and a production function based on decreasing productivity of capital and constant returns to scale. Under these assumptions the model predicts that, in the long-run, an economy's per-capita income converges towards its steady-state.

The Solow-Swan model assumes constant (exogenously given) saving rates and considers the case of a closed economy constituted by isolated regions or countries. Let us formalize this assumption by defining the structural parameters of an economy as: (i) the saving rate (say, s), (ii) the capital depreciation rate (say δ), and (iii) the population growth defined as $n = \frac{\dot{L}}{L}$, L representing the labour force and $\dot{L}$ its derivative with respect to time.

---

[1] A steady-state is defined as the situation where all variables grow at constant rates.

The model also assumes that each country (or region) within the economy is characterized by the same production function. If we define the capital stock as K and the output as Y, then, as a consequence of the previous assumptions, each country (or region) would also show the same steady-state values for the capital-labour ratio, defined as $k = \frac{K}{L}$, and for the per-capita income defined as $y = \frac{Y}{L}$.

Let us now introduce an explicit expression for the production function and let us assume that the technical relationship between the inputs (labour and capital) and the output is well described by a Cobb-Douglas-type production function (Douglas, 1934). In its basic formulation, this function is characterized by constant returns to scale and is defined as:

$$y = f(k) = AK^a L^{1-a} \tag{1.1}$$

with A being the level of technology (initially assumed as a constant) and a ($0 \le a \le 1$) a parameter representing the elasticity of the output with respect to the capital stock.

A second fundamental equation in the model is the accumulation function for the physical capital given by:

$$\dot{K} = I - \delta K \tag{1.2}$$

with I the level of gross investment flows and $\dot{K}$ the derivative of the capital stock with respect to time. Thus in each period of time the growth of physical capital equals the amount of gross investment flows minus the capital depreciation. By assuming the investment to equal the total saving and the total saving to be proportional to income, we have:

$$I = sy = sAK^a L^{1-a} \tag{1.3}$$

and hence:

$$\dot{K} = sAK^a L^{1-a} - \delta K \tag{1.4}$$

In order to study the evolution of the economy's per-capita income y, let us consider the relationship:

$$\frac{\dot{k}}{k} = \frac{\dot{K}}{K} - \frac{\dot{L}}{L} = \frac{\dot{K}}{K} - n \tag{1.5}$$

By multiplying both sides by k we obtain

$$\dot{k} = \frac{\dot{K}}{K}k - nk = \frac{\dot{K}}{L} - nk \tag{1.6}$$

and, using Equation (1.2) one obtains:

$$\dot{k} = \frac{I - \delta K}{L} - nk \tag{1.7}$$

Let us now substitute Equation (1.3) into Equation (1.7). We have:

$$\dot{k} = \frac{sAK^{a}L^{1-a}}{L} - \frac{\delta K}{L} - nk = sAk^{a} - k(\delta + n) \tag{1.8}$$

Finally, by dividing both sides by *k,* we obtain a formal expression of the capital growth, say γ_k, as:

$$\gamma_k = \frac{\dot{k}}{k} = sAk^{a-1} - (\delta + n) \tag{1.9}$$

Since, from Equation (1.1), the growth of *y* is proportional to the growth of *k*, then Equation (1.9) also describes the per-capita income's growth rate.

If we assume that the structural parameters *s* and *n,* and the level of technology *A* are constant through time, Equation (1.9) implies that regions with a lower starting value of the capital-labour ratio have higher per-capita growth rates γ_k and, therefore, they tend to catch up or "*converge*" to those with higher capital-labour ratios.

In a wider sense, by introducing the idea of "*conditional convergence*", the neoclassical model predicts that each economy converges towards its own steady-state and that the speed with which this convergence occurs relates inversely to the distance from the steady-state itself.

This basic model has been augmented to account for technological progress, leading to the following expression:

$$\gamma_{\hat{k}} = sA(\hat{k})^{a-1} - (x + n + \delta) \tag{1.10}$$

where $\hat{k}$ is a transformed version of the capital-labour ratio defined as $\hat{k} = \frac{k}{A(t)} = \frac{K}{LA(t)}$, and *A(t)* is a technological term that grows at a constant rate *x* (see e. g. Barro and Sala-i-Martin, 1995, p. 32, for further details)

By considering a log-linear approximation of Equation (1.10) in the neighbourhood of the steady-state, we obtain (see Barro and Sala-i-Martin, 1995; p. 36, for the derivation):

$$\gamma_{\hat{k}} = d\left[\ln(\hat{k})\right]/\, dt \cong -b\left[\ln\left(\frac{\hat{k}}{\hat{k}^{*}}\right)\right] \tag{1.11}$$

where ln represents the natural logarithm, $\hat{k}^{*}$ the steady-state value of $\hat{k}$, and the coefficient b can be obtained from the expression $b = (1 - a)(x + n + \delta)$.

Let us now define, by analogy with the term $\hat{k}$, a transformed version of the per-capita output that accounts for technological changes and let us refer to this quantity with the symbol $\hat{y} = \frac{y}{A(t)} = \frac{Y}{LA(t)}$. If we substitute this expression into Equation (1.10) and use the Cobb-Douglas function (1.1), we obtain a differential equation, expressed in terms of $\ln[\hat{y}(t)]$, that admits a solution at point:

$$\ln[\hat{y}(t)] = (1 - e^{-bt}) \ln(\hat{y}^*) + e^{-bt} \ln[\hat{y}(0)] \tag{1.12}$$

where $\hat{y}^*$ represents the steady-state value of $\hat{y}$.

The coefficient *b* in Equation (1.12) comes from the log-linearization of Equation (1.10) around the steady-state and determines the speed with which $\hat{k}$ moves towards its steady-state value ($\hat{k}^*$). For this reason, it is referred to as *speed of convergence* (see Barro and Sala-i-Martin, 1995; Appendix to Chapter 1). By using the simple Cobb-Douglas production function (1.1) we find that the speed of convergence coefficient for the per-capita output *y* equals the speed of convergence for the capital-labour ratio *k*. Furthermore, its value does not depend on the level of technology *A(t)*.

A second crucial parameter for judging the convergence of the economy is represented by the so-called *half-life time* defined as the time that is necessary for $\ln[\hat{y}(t)]$ to be half way between the initial value $\ln[\hat{y}(0)]$ and the steady-state value $\ln[\hat{y}^*]$. In other words, it represents the time that it takes for half of the initial gap in the per-capita output to be eliminated.

This value should satisfy the condition that $e^{-bt} = 0.5$, as it is evident from Equation (1.12). By solving with respect to *t* we have:

$$t_{half-life} = \frac{\ln(2)}{b} = 0.69 b^{-1} \tag{1.13}$$

One of model (1.12)'s limitations is that it, quite unrealistically, assumes constant saving rates. In order to remove this limitation, some authors have tried to extend the previous framework by assuming the saving rate to be a function of the per-capita capital stock *k*. The model starts from a system of two differential equations that explains the behaviour of the technologically-augmented capital-labour ratio $\hat{k}$ and of the technologically augmented per-capita consumption $\hat{c}$, defined as $\hat{c} = \frac{C}{\hat{L}}$, *C* representing the consumption and $\hat{L} = LA(t)$.

This again leads to expressing the per-capita output as:

$$\ln[\hat{y}(t)] = (1 - e^{-bt}) \ln(\hat{y}^*) + e^{-bt} \ln[\hat{y}(0)] \tag{1.14}$$

For a proof based on Taylor's expansion around the steady-state, see Barro and Sala-i-Martin (1995; p. 87).

In any moment of time $t \geq 0$, the term $\ln[\hat{y}(t)]$ in Equation (1.14) thus appears to be a weighted average of the initial value of the per-capita output (that is, $\ln[\hat{y}(0)]$) and of the steady-state value (that is, $\ln(\hat{y}^*)$), with weights provided by e^{-bt} and $(1-e^{-bt})$. The weight on the initial value, therefore, decreases exponentially at the rate *b*. Equation (1.14) thus coincides with Equation (1.12) even if, in this new setting, the *speed of convergence* depends on the technological parameters and on the consumer's preferences and not only on the structural parameters of the economy.

If we consider the whole period between time 0 and time *T*, we obtain that the average growth rate of the per-capita output *y* over the interval is given by:

$$\frac{1}{T}\ln\left[\frac{y(T)}{\hat{y}(0)}\right] = x + \frac{(1-e^{-bt})}{T}\ln\left(\frac{\hat{y}^*}{\hat{y}(0)}\right) \tag{1.15}$$

with *x* the already defined growth rate of the technological term (see Barro and Sala-i-Martin, 1995; p. 80-81). If the term *x*, the convergence speed *b* and the length of the time interval *T* are constant, then the effect of the initial condition depends on the steady-state position $\hat{y}^*$ and the model predicts *conditional* convergence.

Let us now write Equation (1.14) in an operational form by re-expressing it in discrete time (e. g. years). If we further re-parametrize it in a convenient way, we obtain:

$$\ln\left[\frac{y_t}{y_t}\right] = \alpha' - \beta \ln[y_t] \tag{1.16}$$

with *t = 1, 2, ..., T*, and $\alpha' = x + (1-e^{-bt})[\ln(\hat{y}^*) + x_t]$ assumed to be constant in all regions. The parameter β is linked to the speed of convergence by the relationship $\beta = -(1-e^{-bt})$ and, inversely, $b = -\frac{\ln(1+\beta)}{t}$.

In particular, if we consider only two observations at the beginning and at the end of the time period, then Equation (1.6) implies that the average growth rate over the interval of length *T* is given by:

$$\frac{1}{T}\ln\left[\frac{y_T}{y_0}\right] = \alpha' - \beta \ln[y_0] \tag{1.17}$$

with $\alpha' = x + \frac{(1-e^{-bT})}{T}\ln(\hat{y}^*)$ and $\beta = -\frac{(1-e^{-bT})}{T}$ and hence $b = -\frac{\ln(1+\beta T)}{T}$.

Starting from this conceptual basis, authors like Mankiw *et al.* (1992) and Barro and Sala-i-Martin (1995) suggested augmenting Equation (1.17) in order to in-

clude a random disturbance reflecting unexpected changes in production conditions or preferences and to estimate it using cross-sectional observations. Thus they proposed the following statistical model:

$$\frac{1}{T}\ln\left[\frac{y_{T,i}}{y_{0,i}}\right] = \mu_{o,T,i} + \varepsilon_i \tag{1.18}$$

where $y_{t,i}$ *(t=0,..., T; i=1,...,n)* represents the per-capita income at time t in region i, $\mu_{0,T,i}$ represents the systematic component of the model given by:

$$\mu_{0,T,i} = \alpha' - \frac{(1-e^{-bT})}{T}\ln y_{0,i} \tag{1.19}$$

with b the speed of convergence, and ε_i the non systematic part of the model.

From Equations (1.18) and (1.19) we finally obtain:

$$\frac{1}{T}\ln\left[\frac{y_{T,i}}{y_{0,i}}\right] = \alpha' - \frac{(1-e^{-bT})}{T}\ln y_{0,i} + \varepsilon_i \tag{1.20}$$

From an inferential point of view, Equation (1.20) is usually estimated following two possible strategies. It is either directly estimated through Non-linear Least Squares or, alternatively, is re-parametrized by setting $\beta = -(1-e^{-bT})$ (and hence $b = -\frac{\ln(1+\beta)}{T}$) and $\alpha = T\alpha'$. In this second instance Equation (1.20) becomes:

$$\ln\left[\frac{y_{T,i}}{y_{0,i}}\right] = \alpha + \beta\ln y_{0,i} + \varepsilon_i \tag{1.21}$$

and the parameters are estimated via Ordinary Least Squares (Barro and Sala-i-Martin, 1995).

According to the formalization contained in Equation (1.21) absolute convergence is said to be favoured by the data if the estimate of β is negative and significantly different from 0. Therefore, we can use the usual statistical hypothesis testing procedures to validate the economic-theoretic hypothesis of convergence. Specifically, if the null hypothesis ($\beta = 0$) is rejected, in favour of the alternative hypothesis ($\beta < 0$), we can conclude on an empirical basis not only that poor regions grow faster than rich ones, but also that they all *converge* to the same level of per-capita income.

Let us now take a closer look at Equation (1.21) and let us consider the hypotheses made by Barro and Sala-i-Martin (1995) when using it for inferential purposes. In order to apply the OLS estimators correctly when dealing with the

probability model, the authors implicitly assume that the non-systematic component (ε_i) is normally distributed $(0, \sigma^2)$ independently of $\ln y_{0,i}$. Furthermore, dealing with the sampling model, they assume that $\{\varepsilon_1, \varepsilon_2, ..., \varepsilon_n\}$ are independent observations of the probability model. In the authors' words: "We assume that…[the error] … has zero mean, the same variance […] for all economies, and is independent over time and across economies" (Barro and Sala-i-Martin, 1995; p. 31).

The hypothesis formulated on the sampling model is particularly crucial and the authors themselves admit that it is often unrealistic in empirical cases (see Barro and Sala-i-Martin, 1995; p. 385). Many other authors engaged in validating the theoretical model based on observational data, pointed this out as a weakness in the approach. For instance, De Long and Summers (1991) state that: "Many comparative cross-country regressions have assumed there is no dependence across residuals, and that each country provides as informative and independent an observation as any other. Yet it is difficult to believe that Belgian and Dutch economic growth would ever significantly diverge, or that substantial productivity gaps would appear in Scandinavia. The omitted variables that are captured in the regression residuals seem *ex ante* likely to take on similar values in neighbouring countries. This suggests that residuals in nearby nations will be correlated" (p. 456).

Mankiw too, remarks that: "for the reported standard errors to be correct, the residuals for Canada must be uncorrelated with the residuals for the US. If country residuals are in fact correlated, as it is plausible, then data most likely contain less information than the reported standard error indicate". (see Mankiw, 1995; p. 304)

Finally Temple (1999; p. 130) acknowledges that: "without more evidence that the disturbances are independent, the standard errors in most growth regression should be treated with a certain degree of mistrust".

Due to the paramount importance of this issue, we will now illustrate it more thoroughly by referring to two sets of European regional data.

1.3 A β-convergence Analysis of European and Italian Regions

1.3.1 Introduction

In this section we introduce some regional datasets relating to economic growth that will be used throughout the book to illustrate the problems arising when quantifying relationships in space. More specifically, the aim is to illustrate some of the geographical features of GDP and growth data used in standard convergence analysis and to introduce the need for appropriate corrections in consideration of their spatial nature.

An important source of information regarding to the spatial distribution of income and wealth at a European level may be found in the REGIO database. REGIO is the Eurostat's harmonized regional statistical database covering the main aspects of economic and social life in the European Union. It was created in 1975

and is currently divided into ten statistical domains: demography, economic accounts, unemployment, labour force sample survey, energy statistics, transport, agriculture, living conditions, tourism and statistics concerning research and development. The database is based on the Nomenclature of Statistical Territorial Units (NUTS): a coherent system created to provide a geographical subdivision of the EU's territory. The NUTS system is a hierarchical classification. Each Member State of the EU is divided into a number of regions at the NUTS-1 level. Each of these is further divided into sub-regions at the NUTS-2 level and these, in turn, into smaller areas at the NUTS-3 level (Eurostat, 2002). There are 78 European Union regions at the NUTS-1 level, 211 basic administrative units at the NUTS-2 level, and 1093 subdivisions of basic administrative units at the finer NUTS-3 level of spatial disaggregation.

Two distinct datasets drawn from the REGIO database will be analysed. The first one refers to a time series of the per-capita GDP in the 92 Italian provinces, i.e. the NUTS-3 level regions of the official EU classification. A map of the Italian provinces is reported in Figure 1.1. The second dataset describes the per-capita GDP dynamics in 129 European NUTS-2 regions of 10 European countries whose map is displayed in Figure 1.2. We deliberately chose two different levels of spatial disaggregation because, as may easily be imagined, conclusions can be very different if we look at an economic phenomenon at a coarse or a fine level of geographical detail.

In the following two sections we will analyse GDP growth in the two datasets by applying the β-convergence framework described in the previous section.

Fig. 1.1. Map of the 92 Italian NUTS-3 level regions according to the official EU classification (provinces).

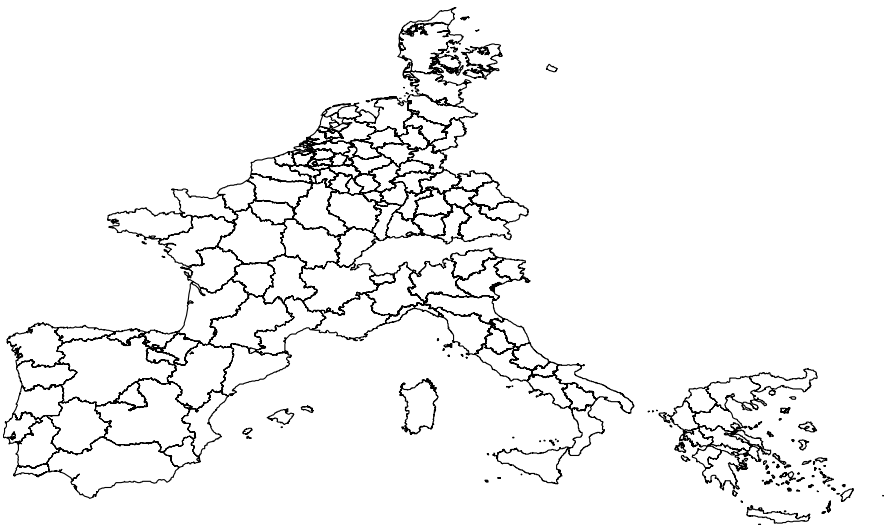


Fig. 1.2. Map of 129 European NUTS-2 level regions of the official EU classification of 10 European countries.

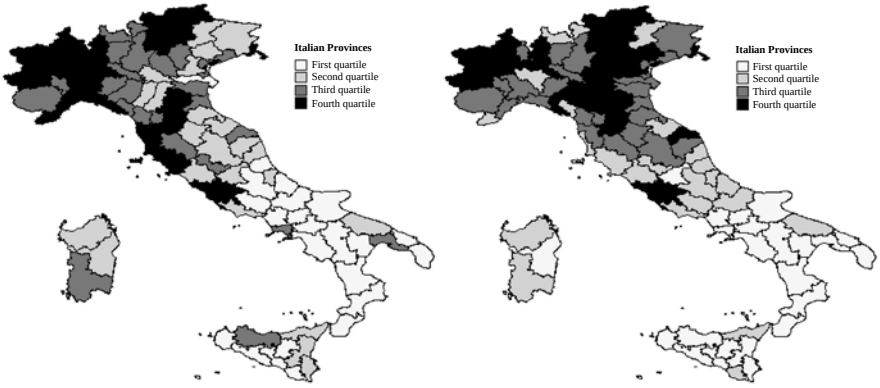
1.3.2 A β-convergence Analysis of Italian NUTS-3 Provinces (1951-1999)

Let us start with a descriptive analysis of the phenomenon of regional growth in the Italian provinces during the period 1951 – 1999. Table 1.1 sets out some descriptive statistics and Figures 1.3 and 1.4 display the geographical pattern of per-capita incomes at the beginning and end of the observed period and their relative growth rates, respectively.

Figure 1.3 is drawn as a quartile map and displays a marked core-periphery pattern in both years: The core is situated in the Northern part of the country, reflecting the well-known Italian dichotomy between the wealthy north and the poorer Southern provinces. This feature is evident in both years, with the only difference being a shift from the north-western to the north-eastern provinces during the period. One of the most remarkable features of this dataset is that the wealthy provinces tend to cluster in space and, as a consequence, the whole map displays a quite evident geographical trend decreasing from north to south. Indeed, when looking at the two maps, one has the visual impression that per-capita income is continuously distributed in space with only a few exceptions. This tendency of rich regions to be surrounded by other rich regions (and, conversely, poor regions to be surrounded by poor regions) is by no means peculiar to the case-study examined here and certainly represents one of the most common features displayed by spatial economic data. The continuity of economic phenomena in space represents the empirical evidence of the feature described in the various econometric textbooks

Table 1.1. Descriptive statistics of per-capita incomes and growth rates in the 92 Italian provinces (years 1951 and 1999).

Per-capita income									
Period	**Min**	**Max**	**Mean**	**First quartile**	**Second quartile**	**Third quartile**	**Coefficient of variation**	**Skewness**	**Kurtosis**
1951	3.28	14.00	6.97	4.72	5.90	7.17	0.38	1.14	3.89
1999	15.77	49.13	30.97	23.08	29.75	35.57	0.27	0.10	2.23
Growth rate									
Period	**Min**	**Max**	**Mean**	**First quartile**	**Second quartile**	**Third quartile**	**Coefficient of variation**	**Skewness**	**Kurtosis**
1951-99	1.82	4.43	3.16	2.94	3.33	3.63	0.16	-0.40	0.03



(a) (b)

Fig. 1.3. Distribution of the per-capita GDP (expressed in natural log) in the 92 Italian provinces; (a) year 1951 and (b) year 1999.

reviewed in Section 1.1 (Kmenta, 1997, Maddala, 2001, Johnston, 1991, Baltagi, 2001, Kennedy, 2003 and Woolridge, 2002a). Further on in the book this feature will be referred to as "*spatial dependence"*.

Figure 1.4 shows the growth rates recorded in the observed time interval. Here the pattern is quite different, with the higher rates scattered irregularly in the eastern part of the country. A closer look at both Figures 1.3 and 1.4 reveals evidence of β-convergence in that some of the poorer regions (namely, those distributed across the Adriatic sea) experienced a faster growth than most of the richer regions in the north during the observed time interval.

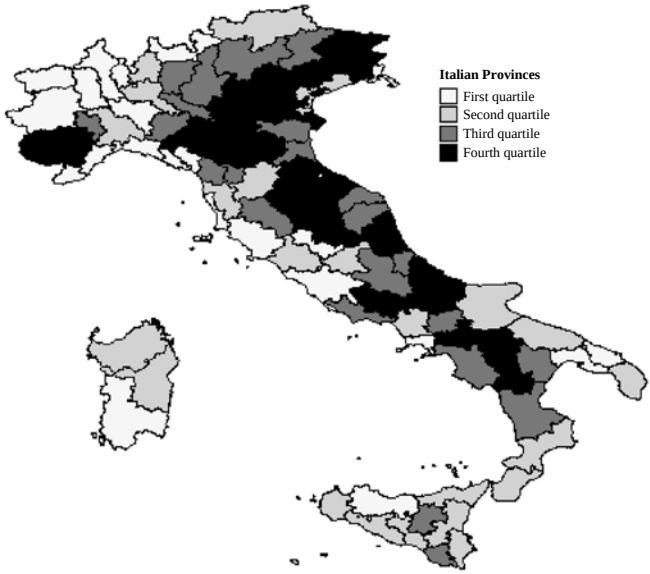


Fig. 1.4. Distribution of provinces' (log) per-capita GDP growth rates for the 92 Italian provinces during the period 1951-1999.

Obviously, a more formal analysis is required to corroborate this visual impression and to quantify the speed of convergence.

Figure 1.5 shows the dynamics of the real per-capita GDP dispersion (measured in log terms) over the period 1951-1999, synthetically measured by the coefficient of variation (the ratio between the standard deviation and the national average). The analysis of the geographical variability's time path is often referred in the literature as σ-*convergence* (Barro and Sala-i-Martin, 1995) an approach that, although heavily criticized e. g. by Arbia (2001b), is the one adopted in the official reports of EU (see for instance EU, 2005). Regional inequalities diminished by more than one half over the entire period, but the sharp trend towards convergence was confined to the period between 1951 and 1970. This is partly due to the significant effort to implement territorial development in the South (through the *Cassa del Mezzogiorno*) and partly to the development of the North-Eastern regions. The subsequent period was, instead, characterized by a substantial invariance of income inequalities.

Finally, Figure 1.6 shows the scatter diagram of the growth rates registered in the 92 Italian provinces during the period 1951-1999 plotted against their initial per-capita GDP. The graph displays a marked tendency towards a linear decrease and indicates how the poorer provinces are those that experienced the faster growth in the period considered, thus beginning to catch up with the richest.

All the elements considered in this first descriptive analysis therefore corroborate the idea of a long-run convergence of Italian provinces.

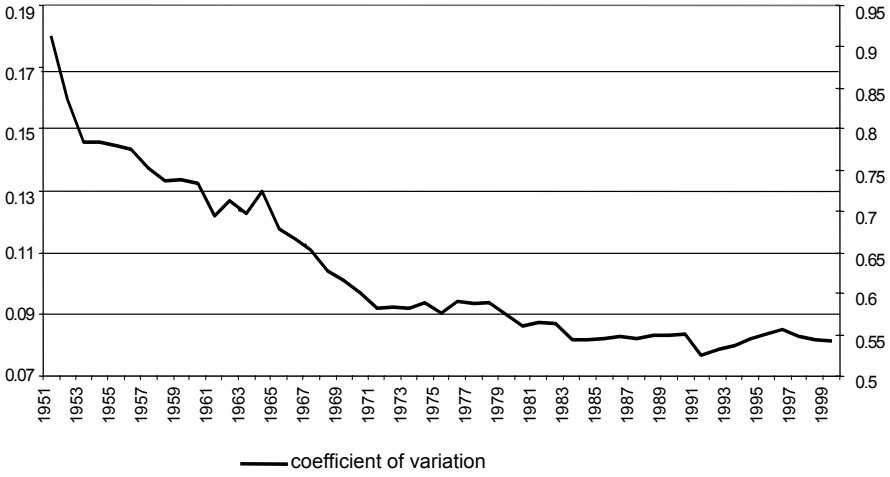


Fig. 1.5. Italian provinces σ-convergence of per-capita income in the period 1951-99 (Coefficient of variation).

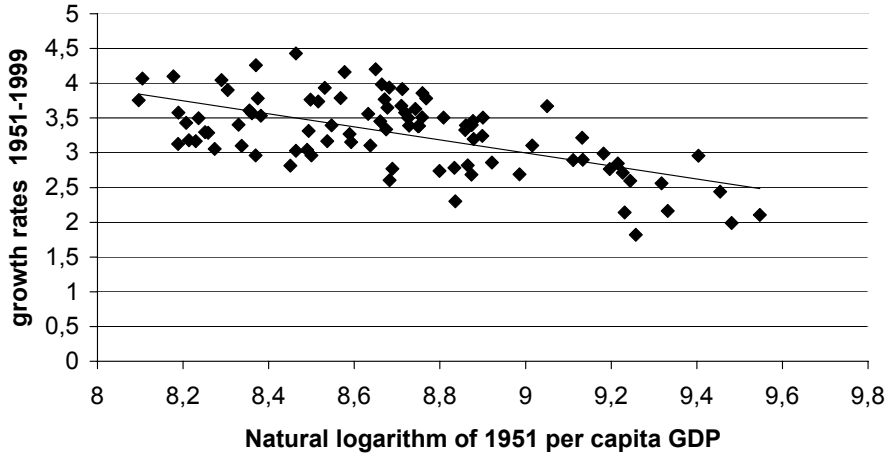


Fig. 1.6. β-convergence among the Italian provinces. Scatterplot of the growth rate during the period 1951-1999 versus the natural logarithm of per-capita GDP (1951).

Moving on to a more formal β-convergence analysis, we start by computing the Ordinary Least Squares estimates of the convergence parameters of Equation (1.21) using the data-set referring to the 92 Italian provinces. We will also judge the performance of the estimated regression by computing some of the standard misspecification tests.

Table 1.2 displays the results of the cross-sectional OLS estimation procedure. The dependent variable of the model is the growth rate of the province's per-

Table 1.2. OLS Estimates of the β-Convergence regression of per-capita income in the 92 Italian provinces (1950-1999). *(Numbers in brackets refer to the p-values).*

α (Constant)	-0.016 (0.735)
β	-0.909 (0.000)
Speed of convergence (*)	0.047
Half-life (**)	14,74
Goodness of fit	
Adjusted R^2	0.366
Schwartz Criterion	-105.890
Regression Diagnostics	
Jarque-Bera normality test	2.133 (0.344)
Breusch-Pagan heteroskedasticity test	0.050 (0.822)

(*) Speed of Convergence $b = -\frac{\ln(1+\beta)}{T}$*; (**) Half-life* $= t_{half-life} = \frac{\ln(2)}{b} = 0.69b^{-1}$

capita income, while the predictor is the initial level of per-capita income (expressed in natural logarithms). Both variables are scaled to the national average.

Our results appear very much in line with the previous findings on the development of Italian regions/provinces. The coefficient of β-convergence for the whole period is highly significant with the expected negative sign, confirming the presence of convergence over the years 1951-1999. Its value (-0.909) implies an annual rate of convergence of 4.7% and an half-life of 14.74 years.

Table 1.2 also reports some diagnostics for identifying misspecifications in the regression model. Firstly, the Jarque-Bera normality test (Jarque and Bera, 1980) is not significant. Consequently, we can safely interpret the results of the various misspecification tests that depend on the normality assumption. Secondly, since no problems were revealed with respect to a lack of normality, the Breusch-Pagan statistic is given (Breusch and Pagan, 1979). Its value is also far from significant, leading to the acceptance of the homoskedasticity assumption.

The regression output, therefore, reports a good fit and no misspecification warning. But, what about the hypothesis of independence assumed in the sampling model?

When dealing with a dynamic regression model based on time series of data, it is usual practice to contrast the assumption of independence of the non-systematic component (see Section 1.2) with that of a temporal dependence among residuals. In this case, the assumption of independence is formally tested via the Durbin-Watson test (Durbin and Watson, 1950, 1951) or via the various other alternatives

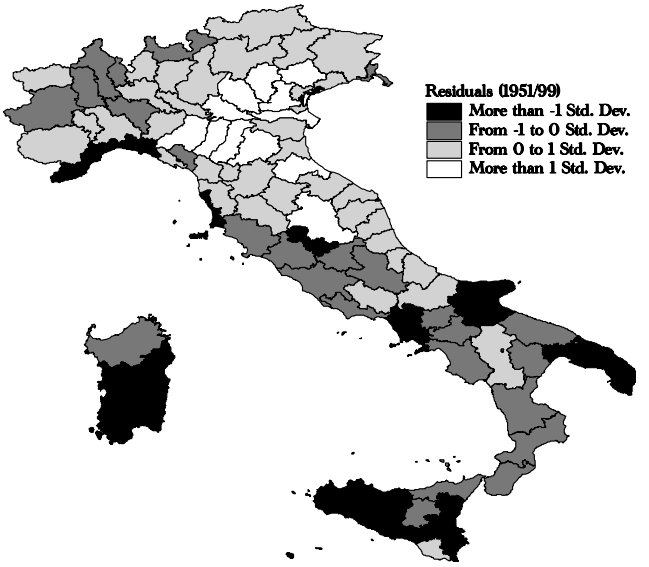


Fig. 1.7. Map of the empirical standardized residuals of Equation (1.21) estimated on the 92 Italian provinces. Growth rates have been measured for the period 1950-1999. Large residuals are identified as those that exceed plus or minus one standard deviation and are displayed in black and white respectively.

put forward in the literature (for example those proposed by Box and Pierce, 1970; Wallis, 1972; Ljung and Box, 1978; Kobayashi, 1991 amongst others).

In contrast, when dealing with regression estimated on spatial data, the econometric textbooks suggest no rigorous testing of the regression residuals' independence because there is no appropriate alternative. Given this lack, let us start, at least for the time being, with a simple visual inspection of the geographical map of the standardized residuals. This map is shown in Figure 1.7.

A visual inspection of Figure 1.7 reveals quite a distinct geographical pattern of residuals. In fact, large positive geographical residuals are concentrated in the extreme southern regions of the peninsula (Sardinia, Sicily and Apulia) whereas large negative residuals are concentrated in the central and northern areas corresponding to large parts of the Veneto, Emilia-Romagna and Umbria regions. Furthermore in only a few cases can we observe large positive residuals (those marked in black on the map) juxtaposed to large negative residuals (those marked in white). Residuals are rather regularly distributed in space with a smooth transition from a few "hot spots" of large positive values (located in the south) to lower values, giving the visual impression of a sort of spatial continuity.

From these considerations it appears clear enough that the estimated relationship is rather unsatisfactory notwithstanding the model's good performances measured by the standard model diagnostics. In fact, the model tends to overestimate (producing negative residuals) growth rates in some provinces located in

proximity to each other (namely, centre and northern areas). Conversely, the relationship tends to underestimate the growth rates (producing positive residuals) in some other provinces geographically located in the south. In these conditions, the relationship estimated in the above regression cannot be taken as a good representation of reality. Indeed, either the linear specification is not able to capture the relationship between the two variables or, alternatively, some relevant explicative variables related to the geography of the phenomenon have been omitted. Notice that we cannot change the model's specification in the specific case we are analysing because the economic-theoretic model that we want to test requires the linearity of the relationship as explained in Section 1.2. Thus the only possibility of improving the model's performances rests on the explicit consideration of space as a further variable that explains why some regions occupying a definite position in space, grow faster than they should according to the assumed economic theory.

Such considerations motivate us towards a more formal analysis of the peculiarity of space in the study of economic relationships.

1.3.3 A β-convergence Analysis of European NUTS-2 Regions (1980-1996)

Let us now consider the second data set referring to the European Union regions at a NUTS-2 level. In this analysis, we again consider the per-capita GDP (measured in purchasing power parities) of 129 European regions for the period 1980-1996. Like the Italian data considered in Section 1.3.2, this second set of observations also derive from the REGIO database provided by Eurostat, and refers to the territorial units of ten European countries (i. e. Belgium, Denmark, France, Greece, West Germany, Italy, Luxembourg, Portugal, Spain and The Netherlands) at the NUTS-2 level.

We have chosen the NUTS-2 subdivision in this second analysis in that it can be considered fine enough to observe spatial effects at a continental level.

Let us start again with a descriptive analysis based on the σ-convergence of the European regions during the period considered. Figure 1.8 shows the coefficient of variation's dynamics over the period 1980-1996. After a phase where regional inequalities seem to remain constant, after 1986 we observe a decreasing trend of the coefficient of variation and, hence, a decrease of economic disparities between regions.

Let us further consider a visual inspection of the geographical pattern of the GDP in the initial year (1980) and in the final year of the period (1996), and of the growth rate observed over the whole period. The maps of the GDP in the two years are reported in Figures 1.9a and 1.9b. Both maps display a spatial trend with a marked core-periphery pattern. In fact, the higher GDP values in both years are concentrated in the centre of the continent (located somewhere between southern Germany, eastern France and northern Italy) with a smooth decline towards the lower values that may be observed in the continent's peripheral regions.

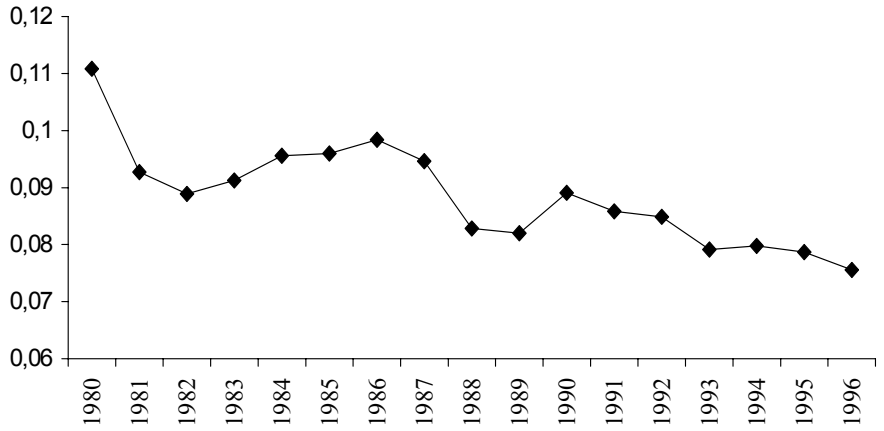


Fig. 1.8. European regions' σ-convergence of per-capita income during the period 1980-1996 (coefficient of variation).

Figure 1.10 reports the growth rates recorded in the period from 1980 and 1996. Similarly to Figure 1.4, regions characterized by lower initial levels of per-capita GDP (e. g. the Spanish regions) experienced higher growth's rates in the time interval, whereas regions with a high level of per-capita GDP in 1980 (remarkably French and northern Italian regions) display low levels of growth in the interval.

Since Figures 1.8 to 1.10 show evidence of convergence, we move to a formal β-convergence analysis by estimating Equation (1.21) on this new set of data.

To start with, Figure 1.11 displays the scatter diagram of the growth rate with respect to the initial level of per-capita income and further corroborates the hypothesis of convergence. Indeed, the poorer regions (corresponding to Portugal) display higher growth rates and the richer ones (corresponding to The Netherlands) grow relatively less: almost 80% more slowly than the poorer ones.

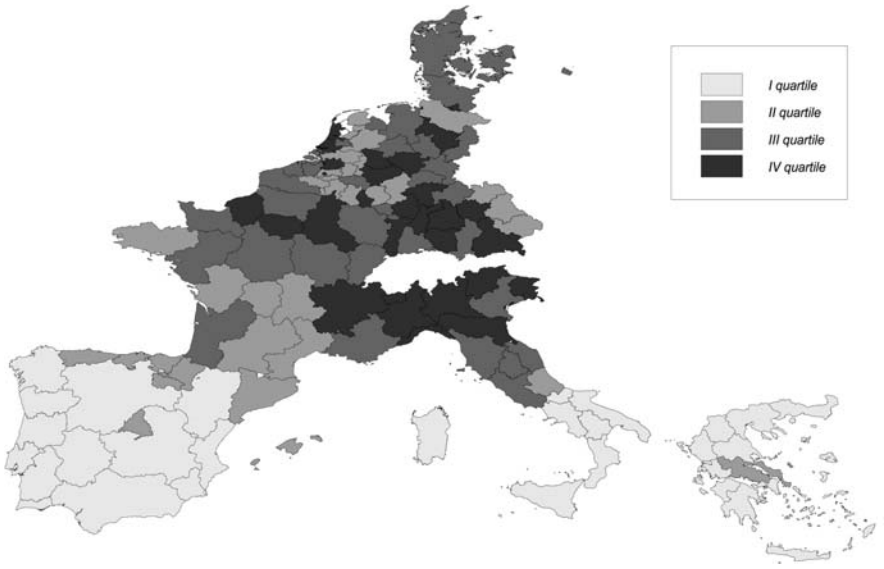
The main results of the OLS estimation procedures are summarized in Table 1.3.

The empirical results highlight convergence. The estimate of the β coefficient is negative and highly significant. The implied speed of convergence is 1.87% and the *half-life* is 36.96 years. This result is in agreement with the "*empirical law of 2%*" observed by Barro and Sala-i-Martin (1995) and Sala-i-Martin (1996)[2].

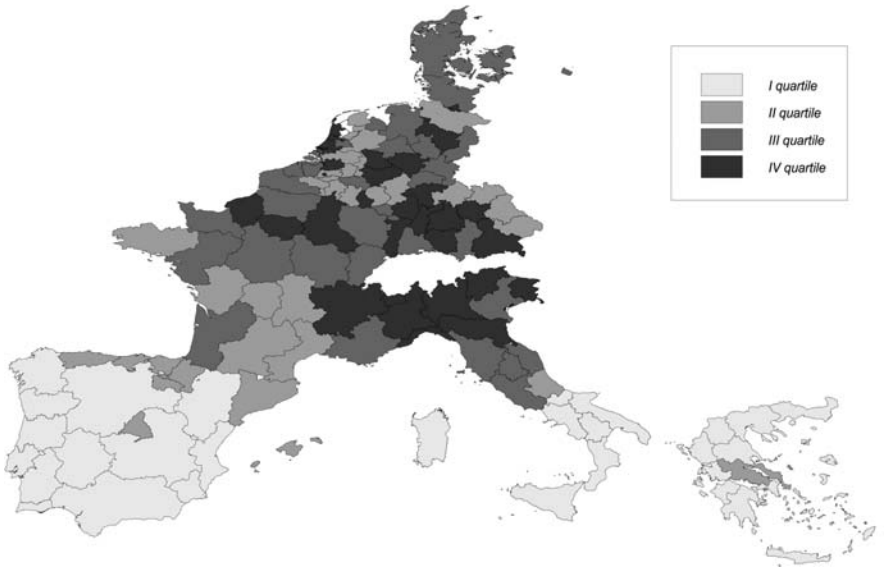
Table 1.3 also shows some diagnostics for evaluating the regression model's performance. Both the Jarque-Bera normality test and the Breusch-Pagan heteroskedasticity test are not significant.

---

[2] In Barro and Sala-i-Martin (1995) and Sala-i-Martin (1996), the authors found that, by analysing European, North American and Japanese regions separately, the speed of convergence was surprisingly similar across different regions and measured approximately 2%.



(a)



(b)

Fig. 1.9. Distribution of the per-capita GDP (expressed in natural log) in the 129 European NUTS-2 regions in (a) 1980 and (b) 1996.

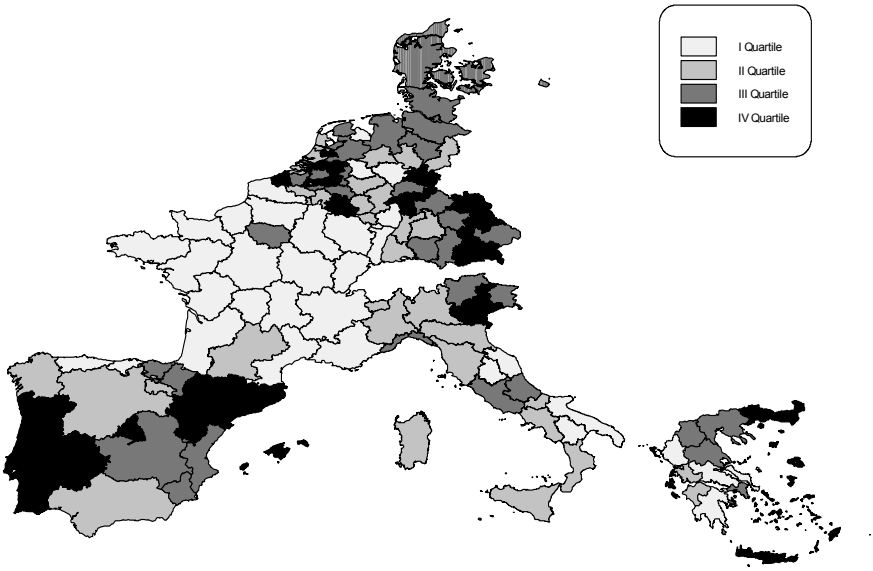


Fig. 1.10. Distribution of the per-capita GDP growth rates (expressed in natural log) in the 129 European NUTS-2 regions in the period 1980-1996.

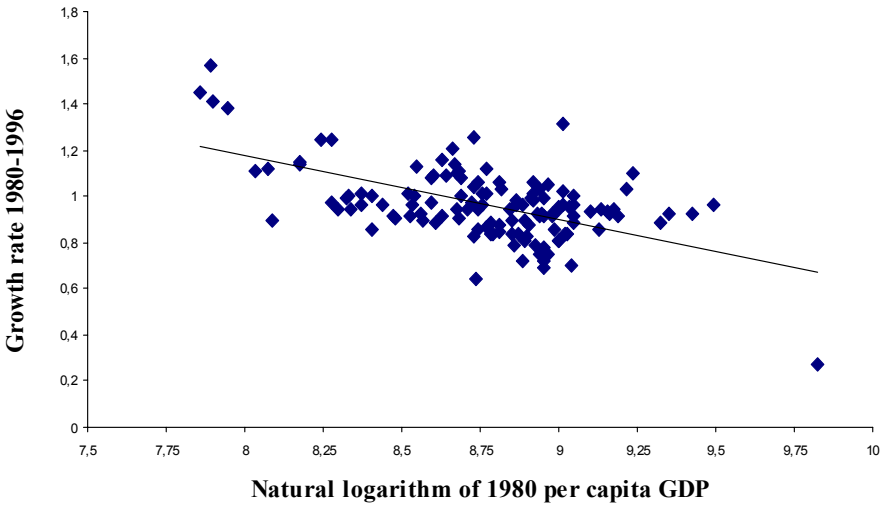


Fig. 1.11. β-convergence among the European regions. Scatterplot of the growth rate during the period 1980-1996 versus the natural logarithm of per-capita GDP (1980).

Table 1.3. OLS Estimates of the β-Convergence regression of per-capita income in the 129 European regions (1980-1996). *(Numbers in brackets refer to the p-values)*

α (Constant)	3.361 (0.000)
β	-0.273 (0.000)
Speed of convergence (*)	0.01875
Half-life (**)	36,96
Goodness of fit	
Adjusted R^2	0.322
Shwartz criterion	-154.545
Regression Diagnostics	
Jarque-Bera normality test	3.014 (0.222)
Breusch-Pagan heteroskedasticity test	0.067 (0.796)

(*) Speed of Convergence $b = -\frac{\ln(1+\beta)}{T}$ *; (**) Half-life =* $t_{half-life} = \frac{\ln(2)}{b} = 0.69b^{-1}$

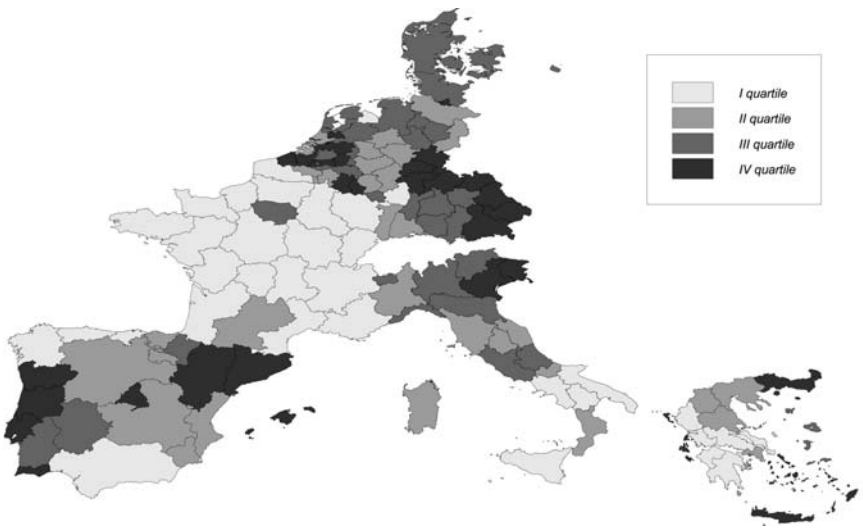


Fig. 1.12. Map of the empirical standardized residuals of Equation (1.10) estimated on the 129 regions at a NUTS-2 level over the period 1980-95. Residuals are classified in the 4 interquartile classes.

Again, the only possibility of investigating the violation of the assumption of independence among the regression model's non-systematic component is by visually inspecting the map of the standardized empirical residuals. This is set out in Figure 1.12. The map displays a marked geographical regularity with most of the regions in France and all southern Italian regions being systematically underestimated by the β-convergence regression. A further geographical feature, already observed in the analysis presented in Section 1.3.2, is constituted by evidently smooth variation from high positive to low positive residuals (the darker shades in Figure 1.11) and from high negative to low negative residuals (the lighter shades in Figure 1.11). Only occasionally are regions falling in the fourth interquartile class close to regions falling in the first interquartile class.

Like to the case discussed in Section 1.3.2, this second example shows that, notwithstanding its good performance with respect to the standard regression diagnostics, the model is not entirely satisfactory. Entire regions, that are not just randomly scattered in the geographical space, experience over/underestimation of growth rates on the basis of the postulated relationship and this evidence re-enphasizes the need to include space and spatial relationships formally in the theoretical formulation of a convergence model.

1.4 A list of Omitted Topics and an Outline of the Book

It is important to remark, at the end of this introductory chapter, that this volume does not aim at covering all the possible topics in spatial econometrics (a task that is better accomplished by other books such as Anselin, 1988; Anselin and Bera, 1998; Anselin and Florax, 1995 and Anselin *et al*., 2004). It is, rather, interested in introducing the basic spatial linear regression model by showing its relevance in one particular instance: the testing of regional convergence of per-capita income. We see this as a paradigmatic example that can help an understanding of more complicated models. The book is therefore limited in terms of the topics covered, even if it introduces in a rigorous way all the basic statistical concepts and tools so that the interested reader will be in a position to understand more advanced methods and techniques if required.

Thus many important topics and specific aspects of the discipline have been deliberately omitted and receive no consideration in this volume.

In particular, even if there are many ways in which spatial data can manifest themselves in economic analysis (e. g. points on a map, as in the case of plants, or lines on a map, as in the case of transport flows), this book will concentrate solely on the treatment of *geographical aggregates* i. e. data collected at a regional, census tract, county or national level. To date this particularly kind of spatial data has clearly been the typology most widely used both in economic analysis and in political debates (see Arbia and Espa, 1996a for a comprehensive review of the methods available for the other typologies of spatial economic data).

A second limitation of this book lies in the fact that it only considers here the inferential problems arising in the phases of estimation and hypothesis testing. No attention is given to the important problem of model identification or that of choosing the best specification among competing spatial models (see e. g. Paelinck and Klaassen, 1979).

A third important issue that is left outside the scope of the present book is that of spatial aggregation. It is clear that the result of any regression analysis based on spatial data depends essentially on the level of geographical aggregation chosen and cannot simply be extended from one level of aggregation to the other. For instance (considering once again the example chosen as the thread of this book), it is perfectly possible, when analysing the economic convergence process of a set of regions, to observe convergence at the European NUTS-2 level, and, conversely, divergence at another level (e. g. the NUTS-3 level). This indeterminacy of statistical results is one of the possible manifestations of the so-called *Modifiable Unit Problem* (Yule and Kendall, 1950), or, better, of its geographical counter-part known as *Modifiable Areal Unit Problem* (or MAUP; see Openshaw and Taylor, 1981; Arbia, 1989) also termed the "second law of geography" (Arbia *et al.,* 1996). The choice of the level of aggregation is thus of paramount importance in any spatial econometric analysis. Of course, this problem is by no mean typical of geographical studies and generally refers to the possible inconsistency between micro and macro relationships in economics, a problem that has received much attention in economics since Theil's seminal contribution (Theil, 1954). Yet, as we have said, it has been left outside the scope of the present book.

A fourth limitation of this monograph is represented by the fact that it will only consider the case of synchronic spatial series of data. Indeed, especially in recent years, there is an increasing availability of spatially distributed data observed through time and this fact, in turn, has much stimulated the development of techniques and models that seek to capture simultaneously the spatial and the temporal dynamics of economic phenomena. A short account of space-time statistical models and of spatial panel data models is reported in Chapter 6 at the end of the book, but the topic is not considered in the rest of the monograph.

Having (I hope) motivated the reader to the study of economic relationships using spatial data, and having clarified the limits within which we have planned to confine ourselves in the present monograph, we can now present the book's structure in more detail.

Chapters 2 and 3 provide the necessary background for introducing the theory of spatial linear modelling. More specifically, Chapter 2 is devoted to a rigorous treatment of the stochastic fields theory as the natural framework for analysing spatial data. Here we will introduce the definition of a random field and the restrictions that we need to consider in order to perform statistical inference. We will also discuss in detail the characteristics of some basic random fields that are particularly useful in econometric analysis. Finally, we will discuss some important limiting theorems that are useful for evaluating the results of asymptotic theory of random fields. In Chapter 3 we will introduce the concept of likelihood and

its related testing procedures and we will discuss some of the adaptations that are necessary in order to define a spatial likelihood.

Chapter 4 represents the core of the whole book and contains the main topics encountered when analysing a spatial linear model. Here we will discuss the major violations of the classical linear regression hypotheses that occur when the model is estimated with spatial data. In doing so we will analyse the violations related to the sampling model (SM) separately from those connected with the underlying probability model (PM).

In Chapter 5 we will return to the introductory examples discussed in the present chapter and we will re-analyse the same data set in order to show how the lesson learnt about the treatment of spatial data can modify the conclusions concerning to the convergence of Italian and European regions.

Finally Chapter 6 is devoted to all those who want to learn more about spatial econometrics and consider this book only as a first reading before attacking the deeper questions and the more recent developments in the topics. Here we will briefly review the recent literature and provide a guided tour of the various more advanced techniques proposed in the last decades.

The book's Appendix contains some references to the statistical software that is currently available for the actual application of spatial econometric procedures to empirical datasets.

2 Random Fields and Spatial Models

2.1 Introduction

An econometric model amenable to statistical estimation is built starting from a set of *n* (possibly multivariate) empirical observations, say $\mathbf{x}=(x_1, x_2,...., x_n)$, that are conceived as the result of a single experiment related to a set of random variables $(X_1, X_2,...., X_n)$. In order to progress along this path it is necessary to introduce a series of hypotheses on the nature of such random variables and the way in which they are related to the observed values. It is useful to classify these hypotheses within two distinct categories, namely:

1. Hypotheses related to a **probability model** (PM) that postulate a plausible form for the joint distribution of the random variables involved. It is convenient to express such a model in the form of a parametric family of density functions $\Phi = \left\{ f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n;\boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}$, with $f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n;\boldsymbol{\theta})$ the joint density function of the random variables associated with the *n* observations, $\boldsymbol{\theta}$ the vector of unknown parameters to be estimated, $\boldsymbol{\Theta}$ the parametric space and Φ the parametric family of density functions; and

2. Hypotheses concerning a **sampling model** (SM) for the *n* observations available, which incorporates information about the criterion of selection from the PM and provides a link between such a PM and the observational data.

The hypotheses underlying the PM and the SM are expressed by taking the nature of the specific economic data considered into account.

Broadly speaking, we can distinguish between four major typologies of economic data arising in empirical research.

(i) A first typology is represented by ***cross-sectional data*** referring either to a single economic agent (household or firm) or to groups of them (such as economic sectors). When building a statistical model for such economic data, the more adequate form of sampling model is the one based on the notion of a *random sample* in which we assume that each random variable is independently distributed with a density function $f_{X_i}(x_i;\boldsymbol{\theta})$. In this case, the joint probability density function considered in the PM can be simplified as the product of the marginal densities, that is:

$$f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n;\boldsymbol{\theta}) = \prod_{i=1}^{n} f_{X_i}(x_i;\boldsymbol{\theta})$$

(ii) A second typology of data relates to a **time series,** referring once again to either single economic agents, or to groups of them, observed in different moments of time. In this case the probability model can be expressed as $\Phi = \left\{ f_{X_{t_1}, X_{t_2}, ..., X_{t_n}}(x_{t_1}, x_{t_2}, ..., x_{t_n}; \boldsymbol{\theta}), t_i \in T, \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}$ in which $X_{t1}, X_{t2}, ..., X_{tn}$ represent a collection of random variables ordered with respect to time (a *random process*), and $f_{X_{t_1}, X_{t_2}, ..., X_{t_n}}(x_{t_1}, x_{t_2}, ..., x_{t_n}; \boldsymbol{\theta})$ their joint density function. In building a statistical model based on time series observations it is not possible to break the joint probability density function down into the same simple form as in the previous case. However, due to the intrinsic order of temporal data, it can be written as the product of the conditional densities, that is:

$$f_{X_{t_1}, X_{t_2}, ..., X_{t_n}}(x_{t_1}, x_{t_2}, ..., x_{t_n}; \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{X_{t_i} | X_{t_2}, ..., X_{t_n}}(x_{t_i} | x_{t_1}, x_{t_2}, ..., x_{t_{i-1}}; \boldsymbol{\theta})$$

(iii) A third typology of economic data relates to *spatial series*, i. e. the typology to which the present book is devoted. This kind of data arises when observing individual economic agents (or groups of them), with additional information about their position in space. This is the first time that we use the word "space" in this monograph and, given that the whole book is devoted to this subject, it is important to provide a formal definition of it. Indeed various definitions are possible according to whether we refer to a geographical space, an economic space, a technical space or other forms of it. Paelinck (1983) defines a "*space*" in the most general terms as the pair $S \overset{\Delta}{=} \{O, R\}$ with *O* representing the objects of the study and *R* the relationships existing between them. The set of objects *O* can represent any geometrical figure such as points, lines, or polygons in $\Re^k$. As for the relationship between objects Paelinck (1983) considers different definitions and lists the three remarkable cases of topological structures, economic structure, and technical structures leading to different definitions of space. We will go back to these concepts in Section 2.2.1.

When data provide extra information about their position in space, the observed sample can be though of as drawn from a Probability Model specified as $\Phi = \left\{ f_{X_{s_1}, X_{s_2}, ..., X_{s_n}}(x_{s_1}, x_{s_2}, ..., x_{s_n}; \boldsymbol{\theta}), s_i \in S, \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}$ in which $X_{s1}, X_{s2}, ..., X_{sn}$ represent a collection of random variables ordered with respect to their geographical location, $f_{X_{s_1}, X_{s_2}, ..., X_{s_n}}(x_{s_1}, x_{s_2}, ..., x_{s_n}; \boldsymbol{\theta})$ represents their joint probability density function, and *S* is an index referring to the spatial location whose nature will be clarified in the following pages. In this third instance none of the previous simplifications of the joint probability density function are feasible in the probability model.

The problem therefore arises as to how to fully redefine the probability model in order to be able to take account of the geographical dependence structure among the random variables involved.

(iv) Finally a fourth typology of economic information relates to panel data referring to either single economic agent or spatial units such as regions or states. In this second case we refer, more specifically, to *spatial panel data*. This typology is originated by the combination of typologies (i) and (ii) or, alternatively (ii) and (iii) above. When we avail a spatial panel data set, the observed sample is generated by the PM specified as:

$$\Phi = \left\{ f_{X_{s_1t_1}, X_{s_2t_1}, \ldots, X_{s_nt_1}, X_{s_1t_2}, \ldots, X_{s_nt_m}} (x_{s_1t_1}, x_{s_2t_1}, \ldots, x_{s_nt_1}, x_{s_1t_2}, \ldots, x_{s_nt_m}; \boldsymbol{\theta}), s_i \in S, t_i \in T; \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}$$

and the random variables are simultaneously ordered with respect to a temporal index *T* and a spatial index *S*.

The most natural way of keeping into account the problems of geographical dependence emerging in the probability model for spatial data is represented by the extension of the familiar concepts relating to a temporal random process (see e. g. Hamilton, 1994) to the idea of a two-dimensional random process (also referred to in the literature as a *random field*).

The idea of a random field was first introduced by Yaglom (1957; 1961; 1962) and then studied by Matern (1960) (more recently reprinted in Matern, 1986) and Whittle (1954; 1963). This chapter provides an introduction to this topic by concentrating only on those aspects that are necessary to specify a statistical linear regression model based on spatial data. For a more thorough review see Guyon (1995).

2.2 The Concept of a Random Field

Let us start by introducing the following definition.

Definition 1. Let $(\Omega, B, P(.))$ be the triplet defining a probability space, with Ω representing the sample space, *B* the associated Borel set and P(.) a probability measure. Moreover, let S be a non-empty set in R^2, and let us define the function X(.,.), with X(.,.) such that: Ω x $\mathsf{S} \rightarrow \mathsf{R}$. The ordered sequence of random variables $\{\mathbf{X}(.,\mathbf{s}), \mathbf{s} \in \mathsf{S}\} = \{\mathbf{X}(\mathbf{s}), \mathbf{s} \in \mathsf{S}\}$, indexed with respect to **s,** is called a *spatial random process* or *random field***.**

There are two fundamental characteristics of a random field $\{X(\mathbf{s}), \mathbf{s} \in \mathsf{S}\}$. The first relates to the nature of the indices $\mathbf{s} \in \mathsf{S}$, i. e. to the topology of the observations. The second concerns the spatial dependency structure displayed by the set of random variables constituting the field. These two features will now be discussed in turn in Sections 2.2.1 and 2.2.2.

2.2.1 The Nature of the Index S

2.2.1.1 Generalities

The indices belonging to the set S can be either continuous or discrete. In the first instance, the index **s** represents the coordinates of n points in R^2 (see Figure 2.1a); in the second case it represents the coordinates on a regular (usually squared) grid (see Figure 2.1b) or an ordered series of values relating to a finite set of polygons (or "regions" in economic geography) (see Figure 2.1c). In this last instance, the vector **s** actually represents a scalar $s \in I^+$ relating to an arbitrary number assigned to each polygon. Just to give the reader an idea, some real cases of point and polygon spatial data are given in Figures 2.2 and 2.3.

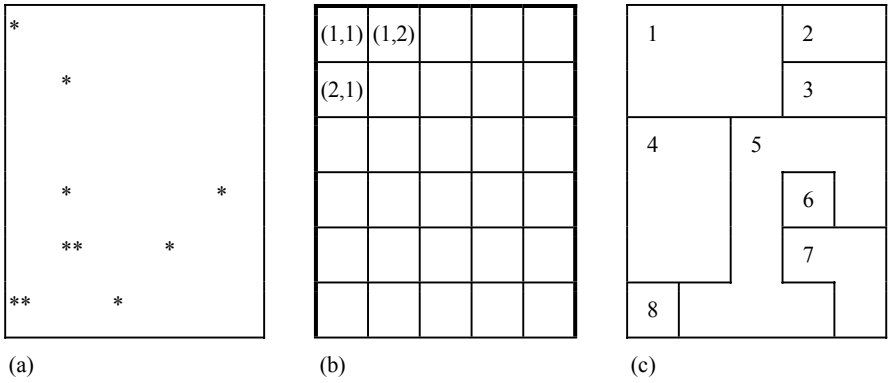


Fig. 2.1. Three possible typologies of discrete spatial data: (a) Points, (b) Regular lattice grid and (c) Irregular lattice.

The statistical information in economic data analysis usually refers either to points in the space economy relating to the position of a single economic agent, or to aggregates observed within sub-national territorial units such as municipalities or regions. The case of data distributed on a regular grid (a case that is extremely interesting in other applied fields such as remote sensing and image analysis, see e. g. Arbia, 1993) is, conversely, still very limited in economic analysis. For some examples of this typology of economic spatial data the reader is referred to (e. g.) Arbia (1996a).

The remainder of this book will concentrate on the study of random fields relating only to point data and regional data with a greater emphasis on regional data. We will refer to the first case as to a continuous-parameter random field and we will define it as $\{X(\mathbf{s}), \mathbf{s}=(r,s), r,s \in \mathsf{R}\}$. In the second instance we will refer to a discrete-parameter random field that will be defined as $\{X(\mathbf{s}), \mathbf{s}=s, s \in I^+\}$.

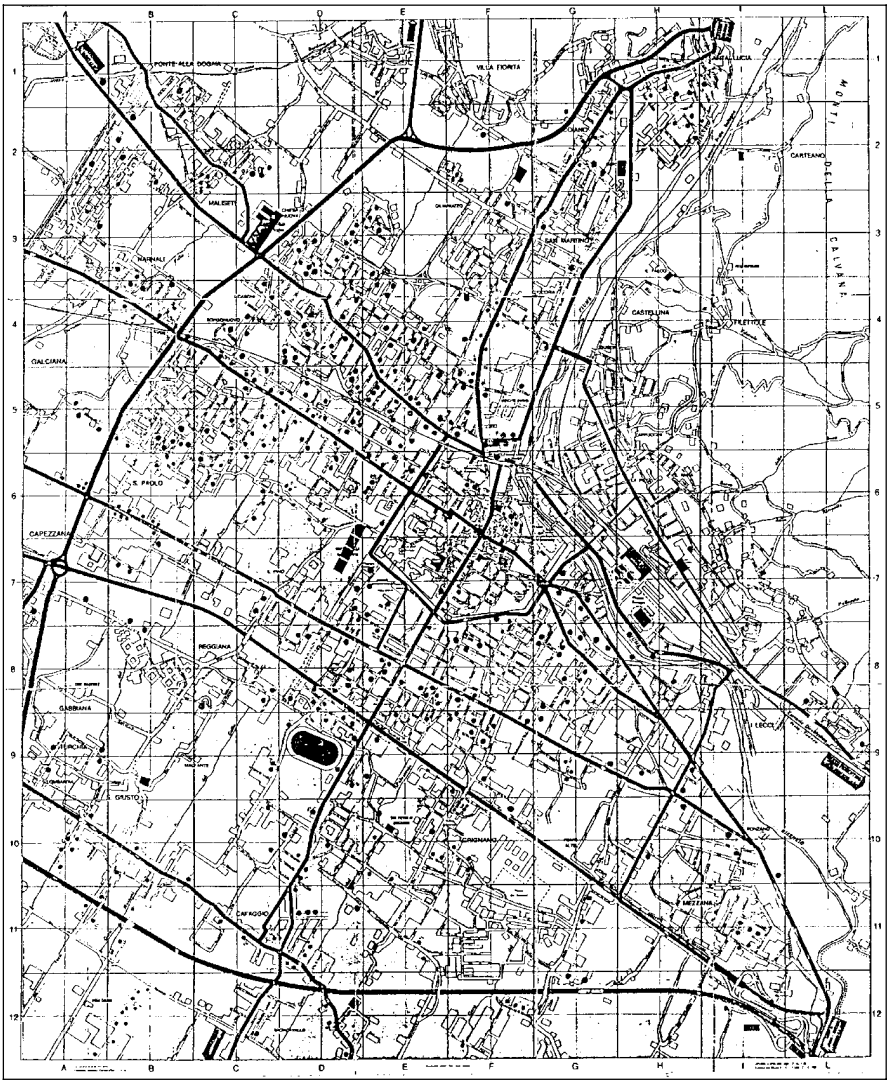


Fig. 2.2. Example of point data: the location of textile companies within the Prato council area. Each point represents a firm. Reported from Arbia and Espa (1996a).

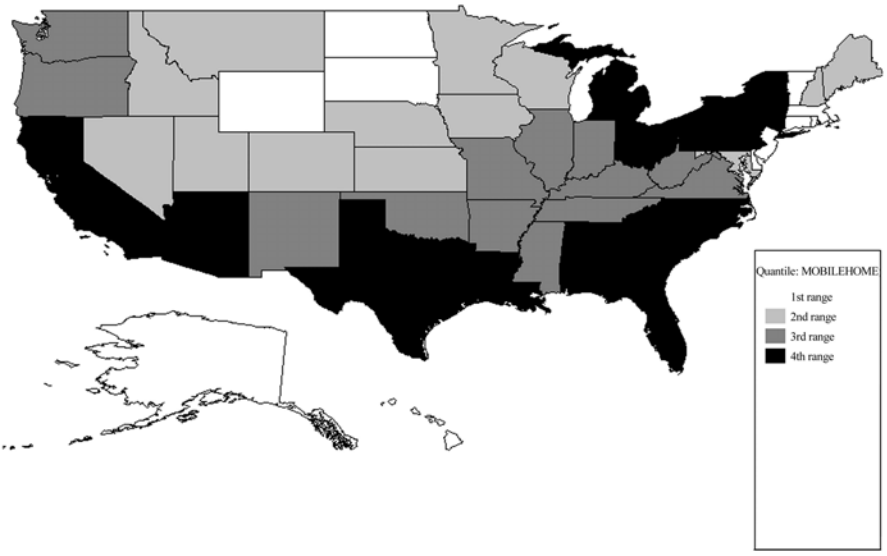


Fig. 2.3. Example of polygon data: Spatial distribution of mobile phones in the US states in 2004.

2.2.1.2 The Topology of a Random Field

In the case of a continuous-parameter random field, the topology of the reference space is fully specified through the concept of metric or distance. In the case of an Euclidian distance in n-space, for instance, one has:

$$d_{ij} = d(\mathbf{s}_i, \mathbf{s}_j) = \sqrt{(\mathbf{s}_i - \mathbf{s}_j)^T (\mathbf{s}_i - \mathbf{s}_j)}$$

with d_{ij} the distance between point $\mathbf{s}_i$ and point $\mathbf{s}_j$. Conversely, in the case of discrete-parameter random fields, the topology needs to be specified exogenously by the researcher, thus inevitably introducing a certain degree of subjectivity into the analysis.

A common way of proceeding is to consider the distance between the centroids of the polygons as representative of the distance between them. This solution is not entirely satisfactory, however, especially in the presence of very irregularly shaped polygons. As an alternative, one may use the Hausdorff concept of inter-polygon distance named after Felix Hausdorff and defined as the maximum distance of a polygon to the nearest point in the other polygon. (See Hausdorff, 1914 and Edgard, 1995). More formally, Hausdorff distance from the polygon A and the polygon B is a *maxmin* function, defined as $H(A,B) = \max_{a \in A} \left\{ \min_{b \in B} \left[d(a,b) \right] \right\}$, where *a* and *b* are points of polygons *A* and *B* respectively, and *d(a, b)* is any

measure of distance between these points. Many other alternative definitions have been suggested in the literature. For a review see, e. g., Paelinck (1983).

In an economic context it is questionable whether a simple Euclidean distance is the most appropriate to capture the geographical links between economic agents and regions. Alternative definitions include a distance based on the time required to reach site $\mathbf{s}_j$ from site $\mathbf{s}_i$ or on the economic cost involved in the trip. Furthermore, some authors suggest the use of social distances (Doreian, 1980) or a wider definition of economic distance (Case *et al.*, 1993; Conley and Topa, 2002). Distances measured in terms of the empirically observed flows (Murdoch *et al*, 1997) or on trade-based interaction measures (Aten, 1996, 1997) could also be taken into account in empirical research.

An important definition in the analysis of the topological links within continuous and discrete parameters random fields is that of the *neighbourhood* of a site. Various definitions are possible. Here we review some of the most commonly used in the spatial econometric literature.

Definition 2. *Critical cut-off neighbourhood.* Two sites $\mathbf{s}_i$ and $\mathbf{s}_j$ are said to be neighbours if $0 \le d_{ij} < d^*$, with d_{ij} the appropriate distance adopted, and d^* representing the critical cut-off.

Definition 3. *Nearest neighbour*. Two sites $\mathbf{s}_i$ and $\mathbf{s}_j$ are said to be neighbours if $d_{ij} = Min\left(d_{ik}\right) \ \forall i,k$.

Definition 4. *Contiguity-based neighbourhood*. In the case of discrete-parameter random fields, a simple definition of neighbourhood could be based on the mere adjacency between two polygons. In this case two polygons indexed by $\mathbf{s}_i$ and $\mathbf{s}_j$ are said to be neighbours if they share a common boundary.

Let us define $N(i)$ as the set of all neighbours (however defined) of site $\mathbf{s}_i$ and η_i as the cardinality of this set. By definition we have that $\mathbf{s}_i \notin N(i)$.

It needs to be stressed here that the result of any econometric analysis will be dependent on the specific topology (and, hence, of the neighbouring structure) chosen for the random field. Consequently it is always wise to test the robustness of the results obtained by adopting several definitions of neighbourhood.

Having clarified the idea of a distance between sites both in the continuous-parameter and in the discrete-parameter case, and also the concept of a neighbourhood, let us now introduce a fundamental tool that will be used extensively in the remainder of the book to express these concepts analytically. This is the so-called *connectivity matrix* (or weights' matrix).

The general form of a binary connectivity matrix $\mathbf{W}$ of generic element w_{ij} is provided by:

$$w_{ij} = \begin{cases} 1 & if \quad j \in N(i) \\ 0 & otherwise \end{cases} \tag{2.1}$$

with $N(i)$ specified according to any of the previous definitions. Notice that we can now express the cardinality of $N(i)$ as $\eta_i = \sum_j w_{ij}$. A connectivity matrix thus formally expresses the proximity links existing between all pairs of sites that constitute a random field according to a pre-specified concept of neighbourhood.

Instead of a simple binary weights' matrix, one can introduce a generalized set of weights that allows the researcher to incorporate his prior knowledge on the geography of the phenomenon under study. This allows a greater flexibility and the possibility of introducing items such as natural barriers, polygon dimensions and shapes. On the other hand, however, the more complex the structure of the connectivity matrix, the more difficult it will be to distinguish between what is a genuine spatial effect and an effect that is forced in by the investigator.

The general expression for a generalized weights' matrix is provided by:

$$w_{ij} = g[d_{ij}] \tag{2.2}$$

with *g* an inverse function of the distance. An example is provided by the gravitational-type weighting:

$$w_{ij} = d_{ij}^{-\alpha} \qquad \alpha > 0 \tag{2.3}$$

It is sometimes useful to consider weights' matrices that are row-standardized, in the sense that each row sums up to 1. This is achieved by defining a set of weights $w^*_{ij} \in \mathbf{W}^*$ such that

$$w^*_{ij} = \frac{w_{ij}}{\sum_j w_{ij}} = \frac{w_{ij}}{\eta_i} \tag{2.4}$$

so that $\sum_j w^*_{ij} = 1$. This last definition is particularly useful for introducing the notion of a *spatially lagged variable*. In time series analysis it is well known that, given a sequence of random variables constituting a random process, say $X_{t1}, X_{t2}, \ldots, X_{tn}$, the lag operator is defined as:

$$L(X_{t_i}) = X_{t_{i-1}}$$

so that the lagged variable is the variable adjacent to X_{t_i} in the temporal scale.

In a spatial context, however, the concept is of difficult extension. In fact, due to the multilaterality of proximity in space, the lagged value of the variable $X(\mathbf{s}_i)$ can be any of the neighbours of $\mathbf{s}_i$ according to the neighbourhood definition chosen. The solution commonly adopted in the literature is that of defining the spatial lag of a random variable $X(\mathbf{s}_i)$ as the mean of the random variables observed in the neighbourhood of site $\mathbf{s}_i$. Consequently we have the following definition.

Definition 5. Given a random field $\{\mathrm{X}(\mathbf{s}),\ \mathbf{s}\in\mathsf{S}\}=\{X(\mathbf{s}_1);X(\mathbf{s}_2);...,X(\mathbf{s}_n)\}$, and given the topology induced by the neighbourhood definition $N(i)$, the *spatially lagged value* of the random variable $X(\mathbf{s}_i)$ is defined as:

$$L[X(\mathbf{s}_i)]=\frac{1}{\eta_i}\sum_{\mathbf{s}_j\in N(i)}X(\mathbf{s}_j) \tag{2.5}$$

or, equivalently, by

$$L[X(\mathbf{s}_i)]=\frac{1}{\eta_i}\sum_{j=1}^{n}w_{ij}X(\mathbf{s}_j)=\sum_{j=1}^{n}w_{ij}^{*}X(\mathbf{s}_j) \tag{2.6}$$

using Equation (2.4), or, in matrix notation, for the whole field

$$L[\mathbf{X}(\mathbf{s})]=\mathbf{W}^{*}\mathbf{X}(\mathbf{s}) \tag{2.7}$$

where $\mathbf{X}(\mathbf{s})$ is the column vector $((\mathbf{X}(\mathrm{s}_1),\mathbf{X}(\mathrm{s}_2),...,\mathbf{X}(\mathrm{s}_n))^{T}$.

2.2.2 The Dependence Structure of a Random Field

Let us now move to the second fundamental characteristic of a random field that, as we anticipated, refers to the dependence structure of the random variables $\{\mathrm{X}(\mathbf{s}),\ \mathbf{s}\in\mathsf{S}\}$. Such a structure can be determined by the cumulative joint probability distribution function of the field, say:

$$F[x(\mathbf{s}_1),x(\mathbf{s}_2),...,x(\mathbf{s}_n)]\ =\ \Pr\{X(\mathbf{s}_1)\le x(\mathbf{s}_1),X(\mathbf{s}_2)\le x(\mathbf{s}_2),...,X(\mathbf{s}_n)\le x(\mathbf{s}_n)\} \tag{2.8}$$

or, in the case of random variables that are absolutely continuous, by the joint probability density function $f_{X(\mathbf{s}_1),X(\mathbf{s}_2)...,X(\mathbf{s}_n),}[x(\mathbf{s}_1),x(\mathbf{s}_2),...,x(\mathbf{s}_n)]$ defined in such a way that:

$$F[x(\mathbf{s}_1),x(\mathbf{s}_2),...,x(\mathbf{s}_n)]=\int_{-\infty}^{x(\mathbf{s}_1)}...\int_{-\infty}^{x(\mathbf{s}_n)}f_{u_1,u_2,...,u_n,}[u_1,u_2,...,u_n]du_1du_2...du_n \tag{2.9}$$

Note that the continuity of the random variables should not be confused with the continuity of the index **s**. We can have continuous random variables distributed over a continuous or a discrete space and, similarly, we can have discrete random variables distributed over a continuous or a discrete space. An example of a discrete random variable distributed on a continuous space is represented by the number of employees in a set of industrial locations in $\Re^2$ space (for an example see, e. g., Arbia, 2001a) ; an example of a discrete random variable on a discrete space is represented by the number of employees in a set of regions; an example

of a continuous random variable on a discrete space is represented by the regional GDP; and, finally, an example of a continuous random variable on a continuous space is represented by the value of the production realized in different industrial locations.

The advantage of operating with random variables distributed in the space is represented by the fact that the geographical position of the random variable can suggest restrictions onto the probability density function, as we will see in the future.

Definition 6. For each sequence of random variables $X(\mathbf{s}_1),...,X(\mathbf{s}_n), n \geq 1,$ belonging to the field {X(**s**), **s**∈S}, the marginal density function of $X(\mathbf{s}_1)$ can be defined as:

$$f_{X(\mathbf{s}_1)}[x(\mathbf{s}_1)] = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f_{X(\mathbf{s}_1),X(\mathbf{s}_2),...,X(\mathbf{s}_n)}[x(\mathbf{s}_1),x(\mathbf{s}_2),...,x(\mathbf{s}_n)]dX(\mathbf{s}_2),...,dX(\mathbf{s}_n) \quad (2.10)$$

Definition 7. For each random variable $X(\mathbf{s}_i)$ belonging to the random field $\{X(\mathbf{s}), \mathbf{s} \in \mathbf{S}\}$, the mean, the variance and the higher moments of $X(\mathbf{s}_i)$ can be defined as:

$$E[X(\mathbf{s}_i)] = \mu(\mathbf{s}_i) \qquad \forall i$$

$$Var[X(\mathbf{s}_i)] = \sigma^2(\mathbf{s}_i) \qquad \forall i$$

$$E[X(\mathbf{s}_i)^r] = \mu_r(\mathbf{s}_i) \qquad \forall i,\ \forall r \geq 1, \mathbf{s} \in \mathbf{S}$$

As we can see, such characteristics of $X(\mathbf{s})$, can generally be expressed as functions of the index **s** since, at least in principle, any random variable $X(\mathbf{s})$ has a different probability density function $f_{X(\mathbf{s})}x(\mathbf{s})$ at any point **s** of the geographical space.

Definition 8. For each pair of random variables $X(\mathbf{s}_1)$ and $X(\mathbf{s}_2)$ belonging to the random field $\{X(\mathbf{s}), \mathbf{s} \in \mathbf{S}\}$, the joint bivariate probability density function is defined as:

$$f_{X(s_1),X(s_2)}[x(s_1),x(s_2)] =$$

$$= \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f_{X(\mathbf{s}_1),X(\mathbf{s}_2),...,X(\mathbf{s}_n)}[x(\mathbf{s}_1),x(\mathbf{s}_2),...,x(\mathbf{s}_n)]dX(\mathbf{s}_3),...,dX(\mathbf{s}_n) \quad (2.12)$$

The various joint moments relating to the joint distributions described under Definition 8 assume a particular meaning, given the importance of the index $\mathbf{s} \in \mathbf{S}$ in a geographical context. In particular, let us consider the following definition:

Definition 9. Given the field $\{X(\mathbf{s}), \mathbf{s}\in S\}$, the quantity:

$$\gamma(\mathbf{s}_i,\mathbf{s}_j) = E\left[\left(X(\mathbf{s}_i)-\mu(\mathbf{s}_i)\right)\left(X(\mathbf{s}_j)-\mu(\mathbf{s}_j)\right)\right] \qquad \mathbf{s}_i,\mathbf{s}_j \in R^2 \tag{2.13}$$

is called the spatial auto-covariance function of the field. Furthermore, its standardised version, given by:

$$\rho(\mathbf{s_1}\mathbf{s}_2) = \frac{\gamma(\mathbf{s_1}\mathbf{s}_2)}{\sqrt{\gamma(\mathbf{s_1})\gamma(\mathbf{s}_2)}} \qquad \mathbf{s}_1, \mathbf{s}_2 \in R^2 \tag{2.14}$$

is said to be the spatial autocorrelation function on the field.

One of the most important examples of random fields in econometrics is the Gaussian (or Normal) field. In fact, members of this family have convenient mathematical properties that enormously simplify calculations. Furthermore, it has been found that the distribution of many empirical processes can be satisfactorily approximated by it. We therefore introduce the following definition:

Definition 10. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ is called **Gaussian** if, for each finite subset in R^2, say $(\mathbf{s}_1, \mathbf{s}_2,\ldots,\mathbf{s}_n)$, $(X(\mathbf{s}_1),\ldots,X(\mathbf{s}_n)) = \mathbf{X(s)}$ is such that:

$$\mathbf{X(s)} \sim \text{MVN}\ (\boldsymbol{\mu}, \mathbf{V})$$

or:

$$f_{X(\mathbf{s}_1),X(\mathbf{s}_2),\ldots,X(\mathbf{s}_n)}[x(\mathbf{s}_1),x(\mathbf{s}_2),\ldots,x(\mathbf{s}_n)] =$$

$$= \frac{\det(\mathbf{V})^{-1/2}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(\mathbf{X(s)}-\boldsymbol{\mu}(\mathbf{s}))^T \mathbf{V}^{-1}(\mathbf{X(s)}-\boldsymbol{\mu}(\mathbf{s}))\right\} \tag{2.15}$$

with $\gamma(\mathbf{s}_i\mathbf{s}_j)\in \mathbf{V}; i=1,\ldots,n; j=1,\ldots,n$; $\mathbf{V}$ an n-by-n auto-covariance matrix and $\boldsymbol{\mu} \equiv [\mu(\mathbf{s}_1),\mu(\mathbf{s}_2),\ldots,\mu(\mathbf{s}_n)]^T$ a n-by-*1* vector of expected values.

A spatial Gaussian field is fully specified in terms of its first two moments which, in turn, are functions of the index **s.**

In econometric analysis, we often jointly analyse various phenomena that are spatially distributed, for instance regional growth rate and regional GDP, or prices and quantities sold in various locations in space. To tackle this important aspect it is necessary to introduce the concept of a vector random field of dimension *k*.

Definition 11. A vector random field $\{\mathbf{X(s)}, \mathbf{s}\in S\}$ is defined as the field $\mathbf{X(s)}=(\mathbf{X}_1(\mathbf{s}), \mathbf{X}_2(\mathbf{s}),\ldots,\mathbf{X}_k(\mathbf{s}))^T$ in which each component $\mathbf{X}_i(\mathbf{s})$ represents a random field $\{X_i(\mathbf{s}), \mathbf{s}\in S\}$.

Such a definition introduces an extra dimension to the analysis because a random vector $\mathbf{X}(\mathbf{s})$ of dimension k is associated with each site considered, $\mathbf{s}_i$, $i=1,...,n$. The cumulative joint probability distribution function of the random vector is defined as:

$$F[\mathbf{x}(\mathbf{s}_1),\mathbf{x}(\mathbf{s}_2),...,\mathbf{x}(\mathbf{s}_n)] = P\{\mathbf{X}(\mathbf{s}_1) \le \mathbf{x}_1, \mathbf{X}(\mathbf{s}_2) \le \mathbf{x}_2,\ldots, \mathbf{X}(\mathbf{s}_n) \le \mathbf{x}_n\} \tag{2.16}$$

and, in the case of continuous random variables, the joint probability density function is:

$$f_{\mathbf{X}(\mathbf{s}_1),\mathbf{X}(\mathbf{s}_2),...,\mathbf{X}(\mathbf{s}_n)}[\mathbf{x}(\mathbf{s}_1),\mathbf{x}(\mathbf{s}_2),...,\mathbf{x}(\mathbf{s}_n)] \tag{2.17}$$

Many of the definitions introduced for random fields can be extended to vector random fields (e. g. mean, variance and r-th moment) with a simple change of notation. However, it is necessary to emphasize the importance of concepts that describe the complex relationships between pairs of random variables. For this reason, we introduce the following further definitions:

Definition 12. Given the vector random field $\{\mathbf{X}(\mathbf{s}), \mathbf{s}\in\mathsf{S}\}$ the quantities:

$$\gamma_{lm}(\mathbf{s}_i,\mathbf{s}_j) = E\left[\left(X_l(\mathbf{s}_i)-\mu_l(\mathbf{s}_i)\right)\left(X_m(\mathbf{s}_j)-\mu_m(\mathbf{s}_j)\right)\right] \qquad \forall l,m;\ \forall \mathbf{s}_i,\mathbf{s}_j \in \mathsf{R}^2 \tag{2.18}$$

and

$$\rho_{lm}(\mathbf{s}_i\mathbf{s}_j) = \frac{\gamma_{lm}(s_i s_j)}{\sqrt{\gamma_{ll}(s_i)\gamma_{mm}(s_j)}} \qquad \forall l,m;\ \forall \mathbf{s}_i,\mathbf{s}_j \in \mathsf{R}^2 \tag{2.19}$$

are called, respectively, the spatial cross-covariance and the spatial cross-correlation functions of the field.

By using the notation employed in Definition 11, we can now introduce the concept of a vector Gaussian field.

Definition 13. A vector random field $\{\mathbf{X}(\mathbf{s}), \mathbf{s}\in\mathsf{S}\}$ is said to be Gaussian if:

$$\begin{pmatrix} \mathbf{X}(\mathbf{s}_1) \\ \mathbf{X}(\mathbf{s}_2) \\ \\ \\ \mathbf{X}(\mathbf{s}_n) \end{pmatrix} \sim MVN\left(\begin{pmatrix} \boldsymbol{\mu}(\mathbf{s}_1) \\ \boldsymbol{\mu}(\mathbf{s}_2) \\ \\ \\ \boldsymbol{\mu}(\mathbf{s}_n) \end{pmatrix}\right)\begin{pmatrix} \mathbf{V}(\mathbf{s}_1) & \mathbf{C}(\mathbf{s}_1,\mathbf{s}_2) & ... & \mathbf{C}(\mathbf{s}_1,\mathbf{s}_n) \\ & \mathbf{V}(\mathbf{s}_2) & & \\ ... & & ... & ... \\ \mathbf{C}(\mathbf{s}_n,\mathbf{s}_1) & & ... & \mathbf{V}(\mathbf{s}_n) \end{pmatrix} \tag{2.20}$$

with $\boldsymbol{\mu}(\mathbf{s}_i)$ k-by-1 vectors of expected values, and $\mathrm{V}(\mathbf{s}_i)$ and $\mathrm{C}(\mathbf{s}_i, \mathbf{s}_j)$, respectively, represent the k-by-k matrices of cross-covariance at site $\mathbf{s}_i$, and the k-by-k matrices of spatial auto-covariance and spatial cross-covariance between pairs of sites, defined by:

$$\mathbf{V}(\mathbf{s}_i) = \begin{pmatrix} \gamma_{11}(\mathbf{s}_i, \mathbf{s}_i) & \gamma_{12}(\mathbf{s}_i, \mathbf{s}_i) & \gamma_{1k}(\mathbf{s}_i, \mathbf{s}_i) \\ & & \\ \gamma_{k1}(\mathbf{s}_i, \mathbf{s}_i) & & \gamma_{kk}(\mathbf{s}_i, \mathbf{s}_i) \end{pmatrix}$$

$$\text{and } \mathbf{C}(\mathbf{s}_i, \mathbf{s}_j) = \begin{pmatrix} \gamma_{11}(\mathbf{s}_i, \mathbf{s}_j) & \gamma_{12}(\mathbf{s}_i, \mathbf{s}_j) & \gamma_{1k}(\mathbf{s}_i, \mathbf{s}_j) \\ & & \\ \gamma_{k1}(\mathbf{s}_i, \mathbf{s}_j) & & \gamma_{kk}(\mathbf{s}_i, \mathbf{s}_j) \end{pmatrix} \tag{2.21}$$

The definition of random fields given so far is too general to allow the construction of estimable statistical models. In fact, in econometric analysis, the researcher almost invariably has only one single realization of the field, i. e. one single observation for each site. Hence the number of parameters to be estimated is much higher than the size of the available information.

The purpose of the next sections is to consider some particular forms of random fields on which a number of restrictions are imposed, so as to obtain a situation where the number of parameters is reduced to a point where their value can be inferred from the single realization available. These restrictions refer to:

1. the spatial heterogeneity of the field (Section 2.2.1), and
2. the spatial dependence of the field (Section 2.2.2).

2.3 Restrictions on Random Fields

2.3.1 Restrictions on the Spatial Heterogeneity of a Random Field

Generally speaking, for a random field $\{X(\mathbf{s}), \mathbf{s} \in S\}$ the joint probability distribution function $F(X(\mathbf{s}), \boldsymbol{\theta}_s)$ depends on the site $\mathbf{s}$ and is characterized by a set of parameters $\boldsymbol{\theta}_s$ which are also dependent on the site. Such a general definition, however, is not operational. Indeed, due to the non-experimental nature of economic data, the standard situation is that of dealing with a single replication available for each random variable of the field. Thus the number of parameters to be estimated is far larger than the available observations in empirical circumstances. For this reason, one is forced to restrict one's attention to a sub-class of random fields which, whilst preserving an acceptable degree of generality, presents a certain level of spatial homogeneity and can be used to model real phenomena. One example of these fields is represented by the class of stationary random fields.

Definition 14: A random field $\{X(\mathbf{s}), \mathbf{s} \in S\}$ is said to be **stationary** (in a strict sense) if, for each subset of sites $(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n)$ belonging to the space S, the joint probability density function $f(X(\mathbf{s}), \mathbf{s} \in S)$ does not change when the subset is shifted in the space.

When dealing with stationarity in time, the notion of (unidimensional) shift does not present any problem. In contrast, when dealing with spatial data (since space is -at least- two-dimensional) a subset of random variables can have two different kinds of shifts. A set of random variables can be rotated at a certain angle, or translated with a rigid motion. If a random field remains unchanged in terms of its joint probability density function after a translation, it is said to be stationary under translations, or *homogeneous* (a concept that is the opposite of *heterogeneous*), which implies that the kind of dependence structure within the random field does not change systematically from one place to the other. If a random field remains unchanged in terms of its joint probability density function after a rotation, it is said to be stationary under rotations around a fixed point or *isotropic* (as opposed to *anisotropic*), which implies that the dependence structure does not change systematically along different directions. The two cases are illustrated in Figure 2.4.

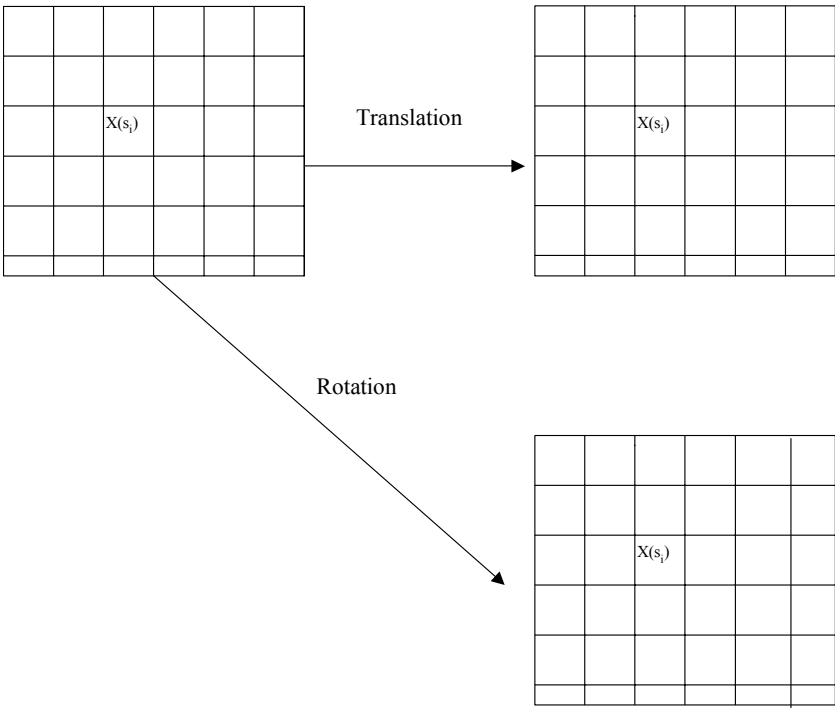


Fig. 2.4. If the process characteristics are unchanged under translation, the process is said to be homogeneous. If they are unchanged under rotation, the process is said to be isotropic.

In the case of continuous-parameter random fields, having defined a vector of arbitrary constants $\sigma \equiv (l,m)$; l, m $\in$ R, stationarity implies that the joint probability density function is :

$$f_{X(\mathbf{s}_1),X(\mathbf{s}_2),\ldots,X(\mathbf{s}_n)}[x(\mathbf{s}_1),x(\mathbf{s}_2),\ldots,x(\mathbf{s}_n)] =$$
$$= f_{X(\mathbf{s}_1+\sigma),X(\mathbf{s}_2+\sigma),\ldots,X(\mathbf{s}_n+\sigma)}[x(\mathbf{s}_1+\sigma),x(\mathbf{s}_2+\sigma),\ldots,x(\mathbf{s}_n+\sigma)] \qquad \sigma \in \mathsf{R} \tag{2.22}$$

In the case of discrete-parameter random fields, on the other hand, the extension of the concept of stationarity from one to two dimensions is much more complicated, because the polygons that constitute the georeference of the random field are characterized, not only by their position in the plane, but also by their size and (irregular) shape.

If a random field is homogeneous and isotropic, it is said to be stationary in the strict sense.

A consequence of strict sense stationarity is that all univariate moments and all mixed moments of any order do not vary when the reference space is modified.

Such a concept of stationarity, however, is very rarely realized in empirical circumstances and one therefore has to introduce the weaker concept of stationarity of order *k* which can be defined on the basis of the first *k* moments of the random field.

Definition 15. A random field $\{X(\mathbf{s}), \mathbf{s} \in S\}$ is said to be stationary of order *k* if, for each subset ($\mathbf{s}_1$, $\mathbf{s}_2$, ..., $\mathbf{s}_n$) of S, its moments up to order *k* do not change when the subset is subject to translations or rotations. We will consider two particular instances:

(i) First-order stationarity:

A random field $\{X(\mathbf{s}), \mathbf{s} \in S\}$ is said to be stationary of order 1 if $E(|X(\mathbf{s})|) < +\infty$, $\forall\, \mathbf{s} \in S$ and, furthermore,

$$E(X(\mathbf{s})) = E(X(\mathbf{s}+\boldsymbol{\delta}) = \mu \qquad \forall \mathbf{s},\boldsymbol{\delta} \tag{2.23}$$

(ii) Second-order stationarity (or *weak sense*):

The random field $\{X(\mathbf{s}), \mathbf{s} \in S\}$ is said to be stationary of order 2 if $E(|X(\mathbf{s})|)^2 < \infty$, $\forall\, \mathbf{s} \in S$ and, furthermore

a. $E(X(\mathbf{s})) = E(X(\mathbf{s}+\boldsymbol{\delta})) = \mu \qquad \forall \mathbf{s},\boldsymbol{\delta}$

b. $E[X(\mathbf{s})]^2 = E(X(\mathbf{s}+\boldsymbol{\delta}))^2 = \sigma^2 \qquad \forall \mathbf{s},\boldsymbol{\delta}$ (2.24)

c. $E[X(\mathbf{s_i})\, X(\mathbf{s_j})] = \gamma\,(d_{ij}) \qquad \forall \mathbf{s}_i, \mathbf{s}_j$

where the symbol d_{ij} denotes the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$.

Stationarity of order 2, therefore, implies that the mean and the variance of a random field do not depend on **s** and that the spatial covariances of the random field $\gamma(\mathbf{s}_i,\mathbf{s}_j)$ depend only on the distance between $\mathbf{s}_i$ and $\mathbf{s}_j$ and not on their absolute position in the plane.

In the case of Gaussian random fields, stationarity in the weak sense implies stationarity in the strict sense and this explains the emphasis in the literature devoted to a weak sense concept.

2.3.2 Restrictions on The Spatial Dependence of a Random Field

The restrictions on spatial dependence play a fundamental role in the study of random fields. Their role in this field is definitely much more crucial than the analogous restrictions in the study of temporal random processes.

In the case of series of spatially distributed economic data that are conceived as a single realization of a random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$, one usually expects the following proposition to hold.

First Law of Geography: "Everything is correlated with everything else, but close things are more correlated than things that are far away" (Tobler, 1970).

For instance, if $X(\mathbf{s})$ refers to the price of a good produced by a firm located at the site with coordinates **s**, we expect the dependence between $X(\mathbf{s}_1)$ and $X(\mathbf{s}_2)$ to be stronger when the two sites are close to each other, and tends to decrease when the distance increases. Similarly, the income of a region will depend much more on the income of the neighbouring regions rather than on regions that are distant in the geographical space. Furthermore, it is reasonable to assume that blocks of spatial units which are sufficiently distant in space (e. g. groups of individuals living in regions very far apart) tend to assume an independent behaviour.

A formal description of such a spatial friction can be derived in terms of the joint probability density function of a random field through the following definition.

Definition 16. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ is said to be asymptotically independent if, by taking two subsets of sites $\{\mathbf{s}_1, \mathbf{s}_2,..., \mathbf{s}_n\}\subset S$ and $\{\mathbf{t}_1, \mathbf{t}_2,..., \mathbf{t}_n\}\subset S$ in such a way that $d(\mathbf{s}_i, \mathbf{t}_i) = \delta$ (with $d(\mathbf{s}_i, \mathbf{t}_i)$ a distance measure) one has:

$$Abs\{F[X(\mathbf{s}_1),...,X(\mathbf{s}_n),X(\mathbf{t}_1),...,X(\mathbf{t}_n)]-F[X(\mathbf{s}_1),...,X(\mathbf{s}_n)]F[X(\mathbf{t}_1),...,X(\mathbf{t}_n)]\}\le\beta(\delta)$$

and

$$\lim_{\delta\to+\infty}\beta(\delta)=0 \tag{2.25}$$

A weaker form of the restriction can be obtained through the concept of asymptotic uncorrelation summarised in the following definition:

Definition 17. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ is said to be asyptotically uncorrelated if there exists a sequence of constants $\{\kappa(d_{ij})\}$, with d_{ij} the distance between the site located at $\mathbf{s}_i$ and the site located at $\mathbf{s}_j$, defined by:

$$|\rho(\mathbf{s}_i, \mathbf{s}_j)| \leq \kappa(d_{ij}) \tag{2.26}$$

such that $0 \leq \kappa(d_{ij}) \leq 1$ and $\sum_{d_{ij}} \kappa(d_{ij}) < +\infty$, with the summation extended over all possible distances within the spatial system.

In other words, asymptotical uncorrelation can be expressed by an upper limit, depending on the distance d_{ij}, that is imposed on the autocorrelation coefficients. Obviously, in the case of Gaussian random fields, the concept of asymptotical uncorrelation and the concept of asymptotical independence coincide.

A further form of restriction on the dependence structure of a random field which will prove useful for developing asymptotical theories for spatial fields, can be derived from the following definition.

Definition 18. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ is said to be strongly mixing if, by taking two subsets of sites $(\mathbf{s}_1, \mathbf{s}_2,...,\mathbf{s}_n) \in S$ and $(\mathbf{t}_1, \mathbf{t}_2,..., \mathbf{t}_n) \in S$, in such a manner that $d(\mathbf{s}_i, \mathbf{t}_i) = \delta$, and defining A as the σ-algebra generated by the random variables $X(\mathbf{s}_1),...,X(\mathbf{s}_n)$, and B as the σ-algebra generated by the random variables $X(\mathbf{t}_1)$, ..., $X(\mathbf{t}_n)$, one has for each event $A\in A$ and $B\in B$:

$$\lim_{\delta\to+\infty} \alpha(\delta) = 0$$

$$\text{with } \alpha(\delta) = \sup_{\substack{A \\ B}} [P(A\cap B)-P(A)P(B)] \tag{2.27}$$

The quantity $\alpha(\delta)$ intuitively represents a measure of the dependence between two groups of variables which constitute the two subsets of sites at a given distance δ. In fact, in the case of independence, one has $P(A\cap B)=P(A)P(B)$ and, therefore, $\alpha(\delta) = 0$ (see Joe, 1997).

An alternative measure of dependence between groups of random variables is supplied by the quantity:

$$\varphi(\delta) = \sup_{\substack{A \\ B}} [P(A|B)-P(A)] \qquad P(B)>0 \tag{2.28}$$

In fact, even in this case one has $P(A|B)=P(A)$, or, alternatively, $\varphi(\delta) = 0$ in the case of independence (see again Joe, 1997). Based on this a further measure of dependence it is possible to introduce the concept of uniformly mixing fields through the following definition.

Definition 19. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ is said to be uniformly mixing if, by taking two subsets of sites $(\mathbf{s}_1, \mathbf{s}_2,..., \mathbf{s}_n) \in S$ and $(\mathbf{t}_1, \mathbf{t}_2,..., \mathbf{t}_n) \in S$ in such a way that $d(\mathbf{s}_i, \mathbf{t}_i) = \delta$, and calling A the σ-algebra associated to the random variables $X(\mathbf{s}_1),...,X(\mathbf{s}_n)$, and B the σ-algebra associated with the random variables $X(\mathbf{t}_1)$, ..., $X(\mathbf{t}_n)$, one has, for each event $A\in A$ e $B\in B$:

$$\lim_{\delta\to+\infty} \varphi(\delta) = 0 \tag{2.29}$$

The weakest form of restriction on the dependence between random variables belonging to a random field, derives from the so-called ergodicity property. Ergodicity is a term borrowed from statistical mechanics (Khinchin, 1949) used extensively in time series literature and extended to random fields (Christakos, 1992). In a temporal random process it is a condition ensuring that the memory of the process, measured by the pairwise correlations, "weakens by averaging over time" (Spanos, 1986; Hamilton, 1994). This implies that the mean and covariance of $X(\mathbf{s})$ coincide with those calculated by means of the single available realizations. Applying a similar idea to a random field leads to the following definition.

Definition 20. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ which is stationary up to the second order is said to be ergodic if:

$$\lim_{d_{ij}\to+\infty} \frac{1}{\Delta} \sum_{d_{ij}} \rho(\mathbf{s}_i, \mathbf{s}_j) = 0 \tag{2.30}$$

with the summation extended to all possible distances in the spatial system, and Δ the number of such distances.

As a consequence of the ergodicity property we have that the spatial averages converge in probability to the set averages (on the concept of convergence in probability, see Definition 35 in Section 3.2 below). Therefore:

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{p} \mu \\ &\frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \xrightarrow{p} \sigma^2 \\ &\frac{1}{n(n-1)}\sum_{j=1}^{n}\sum_{i=1}^{n} (x_j - \mu)(x_i - \mu) \xrightarrow{p} \gamma(d_{ij}) \end{aligned} \tag{2.31}$$

All the restrictions introduced, both on the heterogeneity and on the dependence of a random field, enable a reduction in the number of parameters needed to characterize the field and allows their statistical estimation. They will be reconsidered further on in this chapter when dealing with asymptotical results (see Section 2.5 below).

2.4 Some Special Random Fields

The aim of this section is to introduce several typologies of random field models that can be used in spatial modelling. Some of these models have already been extensively exploited in the literature; others represent the basis for setting up models that are more sophisticated than those currently used in applied spatial econometrics literature.

2.4.1 Spatial White Noise

The simplest of all random fields is the extension to two dimensions of the concept of a white noise time process. Let us introduce the following definition:

Definition 21. A random field $\{u(\mathbf{s}), \mathbf{s} \in S\}$ is said to be a spatial white noise if:

$$\text{(i) } E(u(\mathbf{s})) = 0 \qquad \forall\, \mathbf{s} \tag{2.32}$$

$$\text{(ii) } E(u(\mathbf{s}_i), u(\mathbf{s}_j)) = \begin{cases} \sigma^2 & \text{for} \quad \mathbf{s}_i = \mathbf{s}_j \\ 0 & \text{otherwise} \end{cases}$$

that is if the random variables constituting it have zero mean and are pairwise uncorrelated, no matter what their position is on the plane. The spatial white noise is obviously of no interest *per se* in a spatial context where we need to model dependent events. It does, however, represent the basis for building up more meaningful fields.

An important class of parametric random fields models is that based on the two-dimensional Markov property. This class of models will be introduced in the next section.

2.4.2 Markov Random Fields

2.4.2.1 Generalities

An important class of random fields much studied in the literature is that of the Markov random fields. The concept of a Markov random field is essentially due to Dobrushin (1968) and represents a way of extending the Markov property, originally defined for time processes, to more than one dimension. As is well known, for random processes in time, the Markov property can be expressed by the fact that the "future" of the process, given the "present", is not affected by the "past". For spatial random fields the extension has to take into account the multilaterality of dependence in space. Generally speaking, it is possible to introduce the following definition.

Definition 22. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ is called a Markov field if, for each $\mathbf{s}_i \in S$ one has:

$$E\left[\left|X(\mathbf{s}_i)\right|\right] < \infty$$

and (2.33)

$$E\left[X(\mathbf{s}_i)\middle|B_{-\mathbf{s}_i}\right] = E\left[X(\mathbf{s}_i)\middle|\sigma\left(X(\mathbf{s}_j): \mathbf{s}_j \in N(i)\right)\right]$$

where $B_{-\mathbf{s}_i}$ represents the σ-algebra associated with all the $X(\mathbf{s}_j)$ excluding $X(\mathbf{s}_i)$ and $\sigma\left(X(\mathbf{s}_j): \mathbf{s}_j \in N(i)\right)$ represents the σ-algebra associated only with the random variables $X(\mathbf{s}_j)$ belonging to the neighbourhood of $X(\mathbf{s}_i)$.

An important result useful for particularizing a family of parametric Markov random fields is represented by the Hammersley and Clifford theorem. Due to the paramount importance of this result we will devote the next section to its discussion before introducing some special parametric Markov random fields.

2.4.2.2 The Hammersley and Clifford Theorem

A way in which Markov random fields can be specified in practice consists of assuming a particular functional form for the conditional density function of each random variable and, subsequently, deriving the resulting joint density function. Nevertheless, except in trivial cases, there is no obvious way of deducing the joint probability structure of a process deriving from a conditional distribution set. Moreover, the conditional probability structure is subject to a series of internal coherence conditions and, as a consequence, in some cases it is impossible to define well-conditioned variance-covariance matrices departing from hypotheses about the conditional distributions.

In order to guarantee that the joint density function of the random field exists, is unique and well-conditioned, it is necessary to impose a set of restrictions on the conditional densities that are related to their functional form. These restrictions are identified by the Hammersey-Clifford theorem, a celebrated result that circulated for many years in a private form (Hammersey and Clifford, 1971) before being published by other authors (see e. g. Besag, 1974; Preston, 1973; and Grimmett, 1973 for a simplified version).

Let us start by defining the quantity

$$Q(\mathbf{x}) = \ln \frac{f_X(\mathbf{x})}{f_X(\mathbf{0})} \tag{2.34}$$

with $f_X(\mathbf{x})$ the joint density function of a single realization $\mathbf{x} = (x_1, x_2, ...x_n)$ of a field $X(\mathbf{s})$, and $f_X(\mathbf{0})$ the value assumed by joint density function at point $\mathbf{x}_0 = (0,0,.....0)$. The function $Q(\mathbf{x})$ is sometimes referred to as the "negpotential

function" of a random field (Kaiser and Cressie, 1997). Let us further assume that $f_X(\mathbf{0}) > 0$, which means that such a realisation has a non-zero probability of occurring. The Hammersley and Clifford theorem identifies the most general form of $Q(\mathbf{x})$ that ensures a joint probability structure that is valid and consistent for the spatial system under study.

Such a general form is provided by the expression:

$$Q(x) =$$

$$= \sum_{i=1}^{n} x_i G(x_i) + \sum_{i=1}^{n}\sum_{j=1}^{n}{\sum_{i<j}} x_i x_j G_{ij}(x_i, x_j) + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}{\sum_{i<j<k}} x_i x_j x_k G_{ijk}(x_i, x_j, x_k) + \ldots + \tag{2.35}$$

$$+ x_1 x_2 .. x_n G_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n)$$

where the interaction terms $G(.)$ are non-zero only if the set of random variables included are neighbours. In such a way, one can define valid conditional probabilities. In fact from Equation (2.35) one has:

$$\frac{f_{X_i|X_j}(x_i | x_j; j \in N(i))}{f_{X_i|X_j}(0 | x_j; j \in N(i))} = Exp\left[Q(\mathbf{x}) - Q(\mathbf{x}_i)\right] \tag{2.36}$$

with $\mathbf{x}_i = (x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)$.

The importance of the previous result lies not only in the fact that it guarantees the existence of a joint density function given some restrictions on the conditional densities of spatially interacting random variables, but also in that it suggests the constraints to be imposed on such conditional densities in order to obtain valid representations. Through the Hammersley-Clifford theorem, the joint densities can always be broken down into the interaction between pairs of variables, triplets of variables and so on. In practice, however, Markovian fields have been usually studied by restricting to only pairwise interaction (Ripley, 1990). This is an obvious simplification, but with some empirical foundations. In physics, for example, gravitational and electrostatic forces provide two examples in which the interaction between the various elements in space is purely pairwise (Künsch, 1981). In natural sciences, moreover, many animal species have only pairwise struggles when defending their territories (Ripley, 1990). There nevertheless exist many other instances, including economic situations, where it would be interesting to deepen our knowledge of higher order interaction. The case of spatial price competition (Haining, 1990) constitute one such instance. The pairwise spatial interaction models of random fields are also known as *auto-models*, after Besag (1974). For these models, Equation (2.35) reduces to:

$$Q(\mathbf{x}) = \sum_{i=1}^{n} x_i G(x_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i<j}}^{n} x_i x_j G_{ij}(x_i x_j) \tag{2.37}$$

and

$$Q(\mathbf{x}) - Q(\mathbf{x}_i) = \exp\left[x_i \left(G_i(x_i) + \sum_{\substack{j=1 \\ i<j}}^{n} \rho_{ij} x_j \right) \right] \tag{2.38}$$

where the parameters ρ_{ij} represent pairwise spatial interaction terms and are such that $\rho_{ij} = 0$ unless $j \in N(i)$ and $\rho_{ij} = \rho_{ji}$. In such a way, the restrictions to be imposed on the joint probability density function reduce to restrictions on the possible range of values assumed by the parameters ρ_{ij}.

Starting from the results of the Hammersley-Clifford theorem, Besag (1974) examined what kind of distribution functions can possibly be defined, leading to valid models for random fields. Some of these processes are potentially very useful in economic analyses and will be reviewed in the following sections.

2.4.2.3 Ising's Law

A simple case which occurs in practical instances is the situation where the observed variable is dichotomous. This occurs when we detect the presence of a certain economic agent in one location, or the presence of a certain characteristic in one region (for instance a technological innovation or a particular type of industrial settlement). In this case it is necessary to model the conditional probability:

$$P\left\{X(\mathbf{s}_i) \middle| X(\mathbf{s}_j), \mathbf{s}_i \neq \mathbf{s}_j\right\} = P\left\{X(\mathbf{s}_i) \middle| X(\mathbf{s}_j); \mathbf{s}_j \in N(i)\right\} \tag{2.39}$$

where the random variables $X(\mathbf{s}_i)$ can only assume the values 0 and 1. On this basis one can therefore define the probabilities:

$$p_i = P\left[X(\mathbf{s}_i) = 1 \middle| X(\mathbf{s}_j) = x(\mathbf{s}_j); \mathbf{s}_j \in N(i)\right] \tag{2.40}$$

and

$$q_i = P\left[X(\mathbf{s}_i) = 0 \middle| X(\mathbf{s}_j) = x(\mathbf{s}_j); \mathbf{s}_j \in N(i)\right] \tag{2.41}$$

(with $p_i + q_i = 1$), representing the presence and absence probabilities at site $\mathbf{s}_i$, conditional upon the presence or absence of the phenomenon in the neighbouring sites $\mathbf{s}_j, j \in N(i)$.

If, in particular, the non-zero parameters are those associated with a neighbouring structure that consists of only a single site or pairs of sites, the result is a pairwise interaction model such that:

$$Q(\mathbf{x}) = \sum_{i=1}^{n} \alpha_i x_i + \sum_{\substack{i=1 \\ i<j}}^{n} \sum_{j=1}^{n} \rho_{ij} x_i x_j \tag{2.42}$$

From Equations (2.35) and (2.36) one has:

$$\frac{f_{X(\mathbf{s}_i)|X(\mathbf{s}_j)}(x(\mathbf{s}_i)|x(\mathbf{s}_j); \mathbf{s}_j \in N(i))}{f_{X(\mathbf{s}_i)|X(\mathbf{s}_j)}(0|x(\mathbf{s}_j); \mathbf{s}_j \in N(i))} = \exp\left[x_i (\alpha_i + \sum_{j=1}^{n} \rho_{ij} x_j) \right] \tag{2.43}$$

with $\rho_{ij} = 0$ unless $\mathbf{s}_j \in N(i)$ and $\rho_{ij} = \rho_{ji}$.

On the basis of the previous notions we can now introduce the following definition.

Definition 23. A random field $\{\mathbf{X}(\mathbf{s}), \mathbf{s} \in \mathsf{S}\}$ is said to obey to the auto-logistic (or Ising's) law, if the conditional density function of each random variable with respect to the others can be written as:

$$f_{X(\mathbf{s}_i)|X(\mathbf{s}_j)}\left[x(\mathbf{s}_i)|x(\mathbf{s}_j)\right] = f_{X(\mathbf{s}_i)|X(\mathbf{s}_j)}\left[x(\mathbf{s}_i)|x(\mathbf{s}_j); \mathbf{s}_j \in N(i)\right] =$$

$$= \frac{\exp[(X(\mathbf{s}_i)(\alpha_i + \rho_{ij} X(\mathbf{s}_j))]}{1 + \exp[(\alpha_i + \rho_{ij} X(\mathbf{s}_j))]} \tag{2.44}$$

with $X(\mathbf{s}_i) = (0, 1)$, and α_i and ρ_{ij} parameters. In particular one has:

$$p_i = P\left[X(\mathbf{s}_i) = 1 | x(\mathbf{s}_j), j \in N(i)\right] = \frac{\exp[\sum_{\mathbf{s}_j \in N(i)} (\alpha_i + \rho_{ij} X(\mathbf{s}_j))]}{1 + \exp[\sum_{\mathbf{s}_j \in N(i)} (\alpha_i + \rho_{ij} X(\mathbf{s}_j))]} \tag{2.45}$$

and

$$q_i = P\left[X(\mathbf{s}_i) = 0 | x(\mathbf{s}_j), \mathbf{s}_j \in N(i)\right] = \frac{1}{1 + \exp[\sum_{\mathbf{s}_j \in N(i)} (\alpha_i + \rho_{ij} X(\mathbf{s}_j))]} \tag{2.46}$$

Equations (2.45) and (2.46) show how our expectation of the "presence" of the phenomenon at site $\mathbf{s}_i$ is modified as a function of the number of presences in the neighbouring sites. The set of parameters ρ_{ij} represent the pairwise interaction between neighbouring sites. The parameters α_i are, in contrast, local parameters, and can be used to model the smooth variation of a variable across space. The meaning of the parameters α_i can also be clarified, considering that, in the case of no spatial interaction, one obviously has $\rho_{ij} = 0$ and, hence:

$$p_i = \Pr\left[X(\mathbf{s}_i) = 1 \middle| x(\mathbf{s}_j), j \in N(i)\right] = \frac{\exp[\alpha_i]}{1+\exp[\alpha_i]} \tag{2.47}$$

In those cases where $\rho_{ij} \neq 0$, these parameters regulate the intensity with which the information on the local context modifies the probability of presence.

In order to illustrate the meaning of the parameters, let us consider a simple example and let us assume that a random field obeys Ising's law, with $\alpha_i = \alpha = 1$ for each *i* and, furthermore:

$$\rho_{ij} = \begin{cases} \rho \text{ if } & \mathbf{s}_j \in N(i) \\ 0 & \text{otherwise} \end{cases}$$

or, equivalently, that $\rho_{ij} = \rho w_{ij}$, with w_{ij} the generic element of a simple contiguity matrix **W** such that $w_{ij} = \begin{cases} 1 \text{ if } & \mathbf{s}_j \in N(i) \\ 0 & \text{otherwise} \end{cases}$.

Let us also consider the following spatial configuration:

The probability that location $\mathbf{s}_i$ presents a value equal to 1 will be given by:

$$p_i = P\left[X(\mathbf{s}_i) = 1 \middle| x(\mathbf{s}_j), \mathbf{s}_j \in N(i)\right] = \frac{\exp[(1 + \sum_{\mathbf{s}_j \in N(i)} \rho_{ij} X(\mathbf{s}_j))]}{1+\exp[(1 + \sum_{\mathbf{s}_j \in N(i)} \rho_{ij} X(\mathbf{s}_j))]}$$

In the case of no spatial interaction between the sites, we simply have:

$$p_i = P\left[X(\mathbf{s}_i) = 1 \middle| x(\mathbf{s}_j), \mathbf{s}_j \in N(i)\right] = \frac{\exp[1]}{1+\exp[1]} = 0.73$$

for each location. This value represents the probability that *a priori* (i. e. before observing what happens in the neighbourhood) we attribute to the event $X(\mathbf{s}_i) = 1$, and it depends on the parameter α alone. In our example, three neighbours out of four present a value equal to 1.

In the more general case we therefore have:

$$p_i = P\left[X(\mathbf{s}_i) = 1 \middle| x(\mathbf{s}_j), \mathbf{s}_j \in N(i)\right] = \frac{\exp[1+3\rho]}{1+\exp[1+3\rho]}$$

In the case of moderate spatial dependence (e. g. with ρ = 0.33), we have that p_i =0.88, a value that is greater than the unconditional value previously obtained (0.73). In the case of a high level of spatial interaction (e. g. ρ = 1) we have, instead, that p_i = 0.98 which is even greater. Finally, in the case of a negative interaction (e. g. ρ = -0.33), the conditional probability will assume the value p_i = 0.5, which shows how the unconditional probability (0.73) is reduced by conditioning upon the values assumed in the neighbourhood.

The auto-logistic model corresponds in physics to *Ising's law of ferromagnetism* (Ising, 1925) and has been exploited in economics, e. g., by Haining (1990) amongst the others. It can be considered an extension to the spatial version of the classical logistic model (Cox, 1970), apart from the fact that the explanatory variables are represented, in this instance, by the spatial neighbouring variables. Some interesting simulated realisations of an auto-logistic field, that are useful for understanding the mechanism underlying this process, can be seen, for instance, in Cross and Jain (1983).

2.4.2.4 The Strauss Auto-model

The immediate extension of the auto-logistic field is represented by the case in which the observed variable is categorical, with a discrete number of, say, *c* unordered outcomes. This instance is quite common in economics and occurs, for example when modelling an individual agent's discrete choices. In this case, from Equation (2.37) one has:

$$Q(x) = \sum_{k=1}^{c} \alpha_k n_k - \sum_{k=1}^{c} \sum_{\substack{l=1 \\ k<l}}^{c} \rho_{kl} n_{kl} \tag{2.48}$$

where n_k represents the number of sites belonging to class *k* and, similarly n_{kl} the number of pairs of sites belonging to classes *k* and *l* respectively. Finally, α_k and ρ_{kl} are parameters. From (2.48) and remembering (2.38), one can derive the following definition.

Definition 24. A random field {X(s), s∈S} is said to obey the Strauss law if the conditional density function of each random variable with respect to the others can be written as:

$$P\left[X(\mathbf{s}_i) = k \middle| x(\mathbf{s}_j), \mathbf{s}_j \in N(i)\right] = \frac{\exp\left[\alpha_k + \sum_{\substack{l=1 \\ l \neq k}}^{c} \rho_{kl} u_{\mathbf{s}_i}(l)\right]}{\sum_{l=1}^{c} \exp\left[\alpha_k + \sum_{\substack{l=1 \\ l \neq k}}^{c} \rho_{kl} u_{\mathbf{s}_i}(l)\right]} \tag{2.49}$$

with $u_{\mathbf{s}_i}(l)$ representing the number of neighbours of sites $\mathbf{s}_i$ belonging to class *l*.

In the hypothesis that the classes are interchangeable, and, furthermore that $\rho_{kl} = \rho$, the expression simplifies to:

$$P\left[X(\mathbf{s}_i) = k \middle| X(\mathbf{s}_j), \mathbf{s}_j \in N(i)\right] = \frac{\exp\left[\rho u_{\mathbf{s}_i}(k)\right]}{\sum_{l=1}^{c} \exp\left[\sum_{\substack{l=1 \\ l \neq k}}^{c} \rho u_{\mathbf{s}_i}(l)\right]} \tag{2.50}$$

Such a Markovian model was proposed for the first time by Strauss (1977) and obviously reduces to the auto-logistic when $c = 2$.

2.4.2.5 The Auto-binomial Field

A second extension of the auto-logistic field is represented by the case in which the observed variable can assume a discrete set of numerical outcomes (e. g. the number of industrial plants in one region, or the number of people employed by one firm).

In such an instance, the value in each site can be represented as the outcome of, say, c trials where the probability of success in each trial is p. In other words, we can assume that each random variable constituting the field is conditionally distributed with a binomial density. This leads to the following definition.

Definition 25. A random field {X(s), s∈S} is said to obey an auto-binomial law (Besag, 1974), if the conditional density function of each random variable with respect to the others can be written as:

$$P\left[X(\mathbf{s}_i) = k \middle| x(\mathbf{s}_j), \mathbf{s}_j \in N(i)\right] = \binom{c}{k} p_i^k (1 - p_i)^{c-k} \tag{2.51}$$

with $k = 0, 1, \ldots, c$, and p_i being a spatially varying parameter. In this case, following the Hammersley and Clifford theorem, we have:

$$\log\left[\frac{p_i}{1 - p_i}\right] = \alpha_i + \sum_{j=1}^{n} \rho_{ij} x_j \tag{2.52}$$

and, therefore,

$$p_i = \frac{\exp\left[\alpha_i + \sum_{j=1}^{n} \rho_{ij} x_j\right]}{1 + \exp\left[\alpha_i + \sum_{j=1}^{n} \rho_{ij} x_j\right]} \tag{2.53}$$

with α_i and ρ_{ij} parameters and, for example $\rho_{ij} = \rho w_{ij}$, with $w_{ij} \in \mathbf{W}$ an appropriate contiguity matrix. For $c = 1$ the model obviously reduces to the simpler auto-logistic field.

2.4.2.6 The Auto-Poisson Model

Another interesting instance occurs when the random field is constituted by variables which represent counts of events (*tallies*), but where these events present a very low probability of occurrence. In this case, it is reasonable to assume that the conditional distribution of each random variable in each site follows a Poisson probability law with an expected value, say λ_i, at site *i*, that depends on the occurrences in the neighbouring sites. In this case, the following definition holds.

Definition 26. A random field {X(s), s∈S} is said to obey the auto-Poisson law (Besag, 1974), if the conditional density function of each random variable with respect to the others can be written as:

$$P\left[X(\mathbf{s}_i) = k \middle| x(\mathbf{s}_j), \mathbf{s}_j \in N(i)\right] = \frac{\lambda_i^{\,x} e^{-\lambda_i}}{x!} \tag{2.54}$$

and

$$\lambda_i = \exp\left[\alpha_i + \sum_{j=1}^{n} \rho_{ij} \lambda_j\right] \tag{2.55}$$

Besag (1974) demonstrated that the consistency conditions deriving from the Hammersley-Clifford theorem impose the constraint $\rho_{ij} < 0$ on the parameter ρ_{ij} for each pair of sites *i* and *j*, and that therefore within this framework only negative spatial dependence can be modelled. This fact has strongly limited the use of such a model in economic analysis where the phenomena of positive spatial dependence are of paramount importance. More recently, however, Kaiser and Cressie (1997) developed a model that also allows positive spatial dependence by specifying conditional distributions based on a truncated Poisson (often termed Winsorized Poisson after Galambos, 1988). This Winsorized Poisson random field can be used to model either positive or negative dependencies among spatially distributed random variables and has been used in the literature to describe the mortality and morbidity pattern in epidemiological studies (Clayton and Kaldor, 1987).

2.4.2.7 The Auto-normal (or CAR) Field

If the variable observed is of the continuous type then it is often possible to formulate a Markov process based on a normal distribution. In this case the following definition holds:

Definition 28. A random field {X(**s**), **s**∈S} is called auto-normal, or *Conditional AutoRegressive* (CAR) (Besag, 1974), if the conditional density function of each random variable with respect to the others can be expressed as:

$$f_{X(\mathbf{s}_i)|X(\mathbf{s}_j)}\left[x(\mathbf{s}_i)\middle|x(\mathbf{s}_j)\right] = f_{X(\mathbf{s}_i)|X(\mathbf{s}_j)}\left[x(\mathbf{s}_i)\middle|x(\mathbf{s}_j);\mathbf{s}_i \neq \mathbf{s}_j;\mathbf{s}_j \in N(i)\right] =$$

$$(2\pi\sigma_i^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_i^2}\left[X(\mathbf{s}_i) - \mu_i - \sum_{i\neq j}\rho_{ij}(X(\mathbf{s}_j) - \mu_j)\right]\right\} \tag{2.56}$$

where $\mu_i = E(X(\mathbf{s}_i))$, $\sigma_i^2 = \mathrm{Var}(X(\mathbf{s}_i))$, and ρ_{ij} is a set of constants such that, for example, $\rho_{ij} = \rho w_{ij}$, with $w_{ij} \in \mathbf{W}$ and **W** an appropriate connectivity matrix.

From these definitions it follows that:

$$f_{Y_i|\mathbf{X}_i}(Y_i = y_i|\mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_i) \tag{2.57}$$

and

$$Var\left[X(\mathbf{s}_i) = x\middle|X(\mathbf{s}_j), \mathbf{s}_j \in N(i)\right] = \sigma_i^2 \tag{2.58}$$

From the previous definitions it also follows that the joint probability density function of the random field is MVN(**μ**,**V**), with a vector of means $\boldsymbol{\mu} \equiv (\mu_1, \mu_2, ..., \mu_n)$ and a variance-auto-covariance matrix **V** such that:

$$\mathbf{V} = (\mathbf{I} - \mathbf{B})^{-1}\,\boldsymbol{\Sigma} \tag{2.59}$$

with $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & & \\ & & & \\ 0 & & & \sigma_n^2 \end{pmatrix}$ and $\mathbf{B} = \{\rho_{ij}\}$, with **B** symmetrical, in order to guarantee the symmetry of **V**.

The Markov field described is by far the most used in practice in the applications (see Besag 1986; Mardia, 1990; and Ripley, 1988 for a review).

2.4.2.8 The Intrinsic Gaussian Field

Besag *et al*. (1991) introduced a different specification for an auto-normal model based on the direct specification of the joint probability density function of the field. This is expressed as

$$f_{X_{s_1}, X_{s_2}, ..., X_{s_n}}(x_{s_1}, x_{s_2}, ..., x_{s_n}; \boldsymbol{\theta}) = \frac{c(x)}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j<i}}^{n} w_{ij}(x_i - x_j)^2\right\} \tag{2.60}$$

with $w_{ij} \in \mathbf{W}$ and $\mathbf{W}$ a connectivity matrix. Such a joint probability density function is defined by the conditional expected values:

$$E(X(s_i)|rest) = E(X(s_i)|X(s_j); j \in N(i)) = \frac{1}{\eta_i} \sum_{j \in N(i)} x_j \tag{2.61}$$

with η_i already defined in Section 2.2.1.2, and

$$Var(X(s_i)|rest) = Var(X(s_i)|X(s_j); j \in N(i)) = \frac{\sigma^2}{\sum_{j=1}^{n} w_{ij}} \tag{2.62}$$

and possesses zero unconditional expected values and a variance-covariance matrix defined as:

$$V = \begin{cases} \dfrac{\sum_{j=1}^{n} w_{ij}}{\sigma^2} if & i = j \\ \dfrac{w_{ij}}{\sigma^2} if & i \neq j \end{cases} \tag{2.63}$$

For further details on the intrinsic Gaussian random field see also Molliè (1996).

2.4.2.9 The Bivariate Auto-normal Field

Given its importance in spatial econometrics, (as will become patent in Chapter 4) in this and the following section we will extend the idea of an autonormal field to encompass the case of vector random fields. To start with, in the present section we will define the bivariate auto-normal random field. In the next section we will extend this concept further and introduce the multivariate auto-normal field.

Definition 29. A random field $\{\mathbf{Z}(\mathbf{s}), \mathbf{s} \in \mathsf{S}\}$, $\mathbf{Z}(\mathbf{s})^{\mathrm{T}} = (\mathbf{Y}(\mathbf{s})^{\mathrm{T}}, \mathbf{X}(\mathbf{s})^{\mathrm{T}})$ is said to be bivariate auto-normal (or *twofold* CAR; see Kim *et al.*, 2001) if the conditional density function of each random variable with respect to the others is given by:

$$\begin{aligned} & f_{Y(\mathbf{s}_i)|Y(\mathbf{s}_i),X(\mathbf{s}_j),X(\mathbf{s}_i)}\left[y(\mathbf{s}_i)\middle|y(\mathbf{s}_j),x(\mathbf{s}_i),x(\mathbf{s}_j)\right] = \\ & f_{Y(\mathbf{s}_i)|Y(\mathbf{s}_i),X(\mathbf{s}_j),X(\mathbf{s}_i)}\left[y(\mathbf{s}_i)\middle|y(\mathbf{s}_j),x(\mathbf{s}_i),x(\mathbf{s}_j);\mathbf{s}_j \in N(i)\right] = \\ & = (2\pi\sigma_i^2)^{-\frac{1}{2}}\left\{-\frac{1}{{}_Y\sigma_i^2}\left[Y(s_i)-{}_Y\mu_i - \sum_{i \neq j}\alpha_{ij}(X(s_j)-{}_X\mu_j) - \beta(X(s_i)-{}_X\mu_i) - \sum_{i \neq j}\rho_{ij}(Y(s_j)-{}_Y\mu_j)\right]\right\} \end{aligned} \tag{2.64}$$

where ${}_X\mu_i = E(X(\mathbf{s}_i))$, ${}_Y\mu_i = E(Y(\mathbf{s}_i))$, $f_{\mathbf{x}_i}(\mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_i) = \mathrm{Var}(X(\mathbf{s}_i))$ and ${}_Y\sigma_i^2 = \mathrm{Var}(Y(\mathbf{s}_i))$, with α, β, γ a set of parameters such that $\alpha_{ij} = \alpha w_{ij}$, $\rho_{ij} = \rho w_{ij}$, $w_{ij} \in \mathbf{W}$ and $\mathbf{W}$ a properly defined connectivity matrix.

Note that, from Equation (2.64), the conditional expectation of $Y(s_i)$ given all other random variables can be expressed as:

$$\begin{aligned} &E\left[Y(s_i)\middle|X(s_j);Y(s_j);(X(s_i)\right]= \\ &={}_Y\mu_i + \sum_{i\neq j}\alpha_{ij}\left[X(s_j)-{}_X\mu_j\right]+\beta\left[X(s_i)-{}_X\mu_i\right]+\sum_{i\neq j}\rho_{ij}\left[Y(s_j)-{}_Y\mu_j\right] \end{aligned} \tag{2.65}$$

that is as a linear function both of $X(s_i)$ and of the spatially lagged values of the variables $X(s_j)$ and $Y(s_j)$.

The joint probability density function of the resulting random field is provided by:

$$\mathbf{Z} \sim \mathbf{MVN}\left[\begin{pmatrix} {}_Y\boldsymbol{\mu} \\ {}_X\boldsymbol{\mu} \end{pmatrix}; \mathbf{Q}\right] \tag{2.66}$$

with the variance-covariance matrix $\mathbf{Q}$ given by

$$\mathbf{Q} = \left\{\mathbf{I}^* - \begin{bmatrix} \rho\mathbf{W} & \alpha\mathbf{W}+(\beta-\alpha)\mathbf{I} \\ \alpha\mathbf{W}+(\beta-\alpha)\mathbf{I} & \rho\mathbf{W} \end{bmatrix}\right\}^{-1}\boldsymbol{\Sigma}$$

where $\mathbf{I}^*$ is the identity matrix of dimension $2n$, $\mathbf{I}$ the identity matrix of dimension n, $w_{ij} \in \mathbf{W}$, and $\boldsymbol{\Sigma} = diag({}_Y\sigma_1^2, {}_Y\sigma_2^2, \ldots, {}_Y\sigma_n^2, {}_X\sigma_1^2, {}_X\sigma_2^2, \ldots, {}_X\sigma_n^2)$. Assuming the variances to be constant in the n observations, this simplifies as $\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{I}\sigma_x^2 & 0 \\ 0 & \mathbf{I}\sigma_y^2 \end{pmatrix}$.

2.4.2.10 The Multivariate Auto-normal (or MCAR) Field

Definition 30. A k-dimensional random field $\mathbf{X} = \{\mathbf{X}(\mathbf{s}_i), i \in 1,\ldots,n\}$, $\mathbf{X}(\mathbf{s}_i) \in \Re^k$, is said to be a multivariate CAR (or MCAR, or a multivariate auto-normal random field; see Mardia, 1988) if it admits the following conditional distributions:

$$f_{\mathbf{X}(\mathbf{s}_i)|\mathbf{X}(\mathbf{s}_j),\, j\neq i} \sim N\left(\boldsymbol{\mu}_i + \sum_{j\,:\,j\neq i}\mathbf{C}_{ij}\left(\mathbf{X}(\mathbf{s}_j)-\boldsymbol{\mu}_j\right), \mathbf{V}_i\right) \tag{2.67}$$

where $\boldsymbol{\mu}_i$ is a column vector of expected values each of dimension k, $\mathbf{C}_{ij} \in \mathbf{C}$ are the elements of a k-by-k matrix of parameters (such that, for example, $\mathbf{C}_{ij} = \rho w_{ij}$), and $\mathbf{V}_i$ is the variance-covariance matrix of $\mathbf{X}(\mathbf{s}_i)$ conditional upon $\mathbf{X}(\mathbf{s}_j), j \neq i$.

Let us now consider the k-by-1 vector $\mathbf{Z} = \left(\mathbf{X}(\mathbf{s}_1) \;\cdots\; \mathbf{X}(\mathbf{s}_n)\right)^T$. Under specific conditions, $\mathbf{Z}$ can be considered distributed as the Gaussian field:

$$\mathbf{Z} \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega}), \tag{2.68}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 \quad \cdots \quad \boldsymbol{\mu}_n)^T$ and, by setting $\mathbf{C}_{ii} = -\mathbf{I}_p$,

$$\boldsymbol{\Omega}^{-1} = \begin{bmatrix} -\mathbf{V}_1^{-1}\mathbf{C}_{11} & \cdots & -\mathbf{V}_1^{-1}\mathbf{C}_{1n} \\ \vdots & \ddots & \vdots \\ -\mathbf{V}_n^{-1}\mathbf{C}_{n1} & \cdots & -\mathbf{V}_n^{-1}\mathbf{C}_{nn} \end{bmatrix}. \tag{2.69}$$

In particular, Equation (2.68) holds true if we assume the following hypotheses:

1. $\mathbf{C}_{ij}\mathbf{V}_j = \mathbf{V}_i\mathbf{C}_{ji}^T$, $\forall\, i, j = 1, \ldots, n$;
2. at least one among matrices

$$\begin{bmatrix} -\mathbf{V}_1^{-1} & \cdots & -\mathbf{V}_1^{-1} \\ \vdots & \ddots & \vdots \\ -\mathbf{V}_n^{-1} & \cdots & -\mathbf{V}_n^{-1} \end{bmatrix}$$

and

$$\begin{bmatrix} -\mathbf{C}_{11} & \cdots & -\mathbf{C}_{1n} \\ \vdots & \ddots & \vdots \\ -\mathbf{C}_{n1} & \cdots & -\mathbf{C}_{nn} \end{bmatrix}$$

is positive definite.

The previous result can be proved as follows. Observe that the joint density $f_{\mathbf{Z}}$ is related to the conditional densities $f_{\mathbf{X}(\mathbf{s}_i)|\mathbf{X}(\mathbf{s}_j),\, j \neq i}$ by the property (see, for example, Brook, 1964):

$$\frac{f_{\mathbf{Z}}(\mathbf{z})}{f_{\mathbf{Z}}(\boldsymbol{\mu})} = \prod_{i=1}^{n} \frac{f_{\mathbf{X}(\mathbf{s}_i)|\mathbf{X}(\mathbf{s}_j),\, j \neq i}(\mathbf{x}_i; \mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \boldsymbol{\mu}_{i+1}, \ldots, \boldsymbol{\mu}_n)}{f_{\mathbf{X}(\mathbf{s}_i)|\mathbf{X}(\mathbf{s}_j),\, j \neq i}(\boldsymbol{\mu}_i; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)}, \tag{2.70}$$

where $\mathbf{x}_i$ stands as a short form for $\mathbf{x}(\mathbf{s}_i)$ and $\mathbf{Z} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$. Furthermore, from Equation (2.67) we can write

$$f_{\mathbf{X}(\mathbf{s}_i)|\mathbf{X}(\mathbf{s}_j),\, j \neq i}(\mathbf{x}_i; \mathbf{x}_1, \ldots, \mathbf{x}_{i-1}, \boldsymbol{\mu}_{i+1}, \ldots, \boldsymbol{\mu}_n) = (2\pi)^{-k/2} |\mathbf{V}_i|^{-1/2} \exp\left(-\frac{1}{2}\mathbf{d}_i^T \mathbf{V}_i^{-1} \mathbf{d}_i\right), \tag{2.71}$$

where

$$\mathbf{d}_i = \mathbf{x}_i - \boldsymbol{\mu}_i - \sum_{j=1}^{i-1} \mathbf{C}_{ij}(\mathbf{x}_j - \boldsymbol{\mu}_j), \tag{2.72}$$

and

$$\frac{\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)^T \mathbf{W}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)}{\sqrt{J}\sigma^2} \tag{2.73}$$

where

$$\boldsymbol{\delta}_i = -\sum_{j=i+1}^{n} \mathbf{C}_{ij}\left(\mathbf{x}_j - \boldsymbol{\mu}_j\right). \tag{2.74}$$

As a consequence, from Equation (2.70) we can write

$$\begin{aligned}
&-2\ln\left(\frac{f_{\mathbf{Z}}(\mathbf{y})}{f_{\mathbf{Z}}(\boldsymbol{\mu})}\right) = -2\ln\left(\prod_{i=1}^{n}\exp\left(-\frac{1}{2}\mathbf{d}_i^T\mathbf{V}_i^{-1}\mathbf{d}_i + \frac{1}{2}\boldsymbol{\delta}_i^T\mathbf{V}_i^{-1}\boldsymbol{\delta}_i\right)\right) = \\
&= \sum_{i=1}^{n}\mathbf{d}_i^T\mathbf{V}_i^{-1}\mathbf{d}_i - \sum_{i=1}^{n}\boldsymbol{\delta}_i^T\mathbf{V}_i^{-1}\boldsymbol{\delta}_i = \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\mathbf{x}_i - \boldsymbol{\mu}_i\right)^T\mathbf{V}_i^{-1}\mathbf{C}_{ij}\left(\mathbf{x}_j - \boldsymbol{\mu}_j\right) = \\
&= \left(\mathbf{z}-\boldsymbol{\mu}\right)^T\boldsymbol{\Omega}^{-1}\left(\mathbf{z}-\boldsymbol{\mu}\right),
\end{aligned} \tag{2.75}$$

so that

$$-2\ln\left(f_{\mathbf{Z}}(\mathbf{z})\right) = -2\ln\left(\left(2\pi\right)^{-nk/2}\left|\boldsymbol{\Omega}\right|^{-1/2}\right) + \left(\mathbf{y}-\boldsymbol{\mu}\right)^T\boldsymbol{\Omega}^{-1}\left(\mathbf{y}-\boldsymbol{\mu}\right), \tag{2.76}$$

which implies

$$f_{\mathbf{Z}}(\mathbf{z}) = \left(2\pi\right)^{-nk/2}\left|\boldsymbol{\Omega}\right|^{-1/2}\exp\left(-\frac{1}{2}\left(\mathbf{z}-\boldsymbol{\mu}\right)^T\boldsymbol{\Omega}^{-1}\left(\mathbf{z}-\boldsymbol{\mu}\right)\right), \tag{2.77}$$

as previously stated.

The simple univariate CAR field presented in Section 2.4.2.7 can thus be obtained from the previous formulae by setting $k = 1$. In fact, in this case, $\mathbf{X}$ collapses to the univariate random field, the variance-covariance matrices $\mathbf{C}_{ij}$ become scalars and can be referred to as dependency parameters c_{ij}, with $c_{ii} = -1$, and, finally, $\mathbf{V}_i$ represents the (scalar) conditional variance of $X(\mathbf{s}_i)$, say σ_i^2. In addition, we have $\mathbf{Y}^T = \left(X(\mathbf{s}_1) \quad \cdots \quad X(\mathbf{s}_n)\right)$, $\boldsymbol{\mu}^T = \left(\mu_1 \quad \cdots \quad \mu_n\right)$, and

$$\mathbf{\Omega}^{-1} = \begin{bmatrix} 1/\sigma_1^2 & -c_{12}/\sigma_1^2 & \cdots & -c_{1n}/\sigma_1^2 \\ -c_{21}/\sigma_2^2 & 1/\sigma_2^2 & \cdots & -c_{2n}/\sigma_2^2 \\ \vdots & \vdots & \ddots & \vdots \\ -c_{n1}/\sigma_n^2 & -c_{n2}/\sigma_n^2 & \cdots & 1/\sigma_n^2 \end{bmatrix}, \tag{2.78}$$

or, in a compact matrix notation, as:

$$\mathbf{\Omega}^{-1} = \mathbf{M}^{-1}(\mathbf{I} - \mathbf{B}), \tag{2.79}$$

which is formally equivalent to Equation (2.59). Further details on MCAR fields may be found in Banerjee *et al*. (2004) .

2.4.3 Non-Markovian Fields

Even if the class of Markovian random fields is sufficiently wide to encompass most empirical situations, other random field models deriving from different hypotheses have been introduced in the literature. Some of those that have been exploited in the spatial econometric literature will be reviewed below.

2.4.3.1 The Simultaneous Autoregressive Random Field (SAR)

A first class of non-Markov random fields derives from the extension of the temporal autoregressive model, making use of the concept of spatial lag presented in Definition 5. We have the following definition.

Definition 31. A random field $\{\mathbf{X}(\mathbf{s}),\ \mathbf{s}\in\mathsf{S}\}$ is said to be simultaneous autoregressive (or SAR; Whittle, 1954), if it satisfies the following stochastic difference equation:

$$X(\mathbf{s}_i) - \mu_i = \sum_{i\neq j} \rho_{ij}\left[X(\mathbf{s}_j) - \mu_j\right] + u(\mathbf{s}_i) \tag{2.80}$$

with $\rho_{ij} = \rho w_{ij}$, $w_{ij} \in W$ and $\mathbf{W}$ an appropriate connectivity matrix, and $\{u(\mathbf{s_i}),\ \mathbf{s}\in\mathsf{S}\}$ a spatial white noise process. In this way each of the random variables that constitute the field is seen as a weighted sum of the neighbouring random variables through the weighting matrix $\mathbf{W}$. In this sense, the term $\sum_{i\neq j} \rho_{ij}\left[X(\mathbf{s}_j) - \mu_j\right]$ assumes the meaning of a spatial lag of the centred variable $X(\mathbf{s}_i) - \mu_i$, and Equation (2.80) could also be written as:

$$X(\mathbf{s}_i) - \mu_i = \rho L\left[X(\mathbf{s}_i) - \mu_i\right] + u(\mathbf{s}_i) \tag{2.81}$$

with $L[.]$ the spatial lag operator introduced in Section 2.2.1.2.

From this equation we can obtain some properties of the joint distribution of $X(\mathbf{s}_i)$. Specifically, if we set $\mathrm{Var}(X(\mathbf{s}_i)) = \sigma_i^2$ and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & & \\ & & & \\ 0 & & & \sigma_n^2 \end{pmatrix}$, the matrix of variances and spatial auto-covariances deriving from this model is given by:

$$\mathbf{V} = (\mathbf{I} - \mathbf{B})^{-1}\Sigma(\mathbf{I} - \mathbf{B})^{-T} \tag{2.82}$$

with $\mathbf{B} \equiv \{\rho_{ij}\}$, and $\mathbf{V}$ subject to the restriction that $(\mathbf{I} - \mathbf{B})$ be non-singular to ensure the existence of the inverse $(\mathbf{I} - \mathbf{B})^{-1}$. If, in particular, $\sigma_i^2 = \sigma^2 \; \forall i$, Equation (2.82) reduces to:

$$\mathbf{V} = \sigma^2\left[(\mathbf{I} - \mathbf{B})^T(\mathbf{I} - \mathbf{B})\right]^{-1} \tag{2.83}$$

The SAR field is certainly the most widely used in spatial econometrics (Anselin, 1988; Anselin *et al.*, 2004).

Note that (as stressed by Anselin, 2001a), due to the fact that the diagonal elements in Equation (2.83) are not constant even with i.i.d. error terms, the model generally defines fields that are not covariance-stationary. One exception to this occurs when each site has an identical number of neighbours, but this is a case of limited practical use in econometrics. This lack of stationarity has an important bearing when one is interested in the application of central limit theorems and laws of large numbers for the purpose of deriving asymptotic properties for the estimators and test statistics. This point has recently been raised by Anselin (2001a), but, so far, has not received in the literature the attention it deserves.

The relationship between a CAR model (Equation 2.56) and the SAR specification is the same as that existing between a fixed start-up and a random start-up autoregressive model in the time series literature (Hamilton, 1994). Indeed, in a time series autoregression theory too, we can specify a model in two alternative ways: either as the expectations of the dependent variable conditional upon the independent variables, or, alternatively, as a linear relationship imposed *a-priori* between the dependent variable on one side and the independent variables on the other with an additional stochastic term. There is a remarkable difference with the spatial case, however. In the case of a time series autoregression under the hypothesis of normality, both specifications lead to the same inferential conclusions in terms both of point estimates and of their second-order properties. In the spatial case, conversely, the second-order properties of the two specifications are different, as is clear by comparing the variance-covariance matrix obtained in Equation (2.59) with that obtained in Equation (2.83). Furthermore, the CAR specification leads to a stationary model, while the SAR specification generally does not. Note

also that, in the Gaussian case, any simultaneous model defined by the matrix **B** (call it $\mathbf{B}_s$) can be expressed as a conditional autoregressive model defined by the matrix $\mathbf{B}_c = \mathbf{B}_s + \mathbf{B}_s^T - \mathbf{B}_s\mathbf{B}_s^T$. On the relationships between the CAR and the SAR models the interested reader is referred to Brook (1964). Note, finally, that in this context we use the acronym SAR as it is used in the spatial statistics literature (see e. g. Cressie, 1991), with a meaning different to that used in the spatial econometric literature where it indicates the *spatial mixed regression- autoregression model* (LeSage, 1999; see Chapter 4.3.6 below).

2.4.3.2 The Moving Average Random Field

A second class of non-Markov random field models can be constructed by extending the concept of a moving average (or filtering) to more than one dimension. This leads to the following definition.

Definition 32. A random field $\{\mathbf{X}(\mathbf{s}), \mathbf{s}\in\mathsf{S}\}$ is said to be moving average (SMA; Haining, 1978), if the random variable $X(\mathbf{s_i})$ can be expressed as:

$$X(\mathbf{s}_i) = \mu_i + \sum_{i\neq j} m_{ij}u(\mathbf{s}_j) + u(\mathbf{s}_i) \tag{2.84}$$

with m_{ij} a set of constants such that $m_{ij} = mw_{ij}$, $w_{ij} \in \mathbf{W}$, $\mathbf{W}$ an appropriate connectivity matrix and $\{u(\mathbf{s_i}), \mathbf{s}\in\mathsf{S}\}$ a spatial white noise process.

If we assume that $\mathrm{Var}(u(\mathbf{s_i})) = \sigma^2_{\mathbf{i}}$, and $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & & \\ & & & \\ 0 & & & \sigma_n^2 \end{pmatrix}$, then the matrix of variances and spatial auto-covariances resulting for the model $\{X(\mathbf{s}), \mathbf{s}\in\mathsf{S}\}$ is given by:

$$\mathbf{V} = \left[(\mathbf{I}+\mathbf{M})\Sigma(\mathbf{I}+\mathbf{M})\right]^T \tag{2.85}$$

with $\mathbf{M} \equiv \{ m_{ij} \}$. If, in addition, $\sigma^2_i=\sigma^2\ \forall i$, the expression simplifies to:

$$\mathbf{V} = \sigma^2\left[(\mathbf{I}+\mathbf{M})(\mathbf{I}+\mathbf{M})^T\right] \tag{2.86}$$

This field represents the spatial analogue to the temporal moving average models. Here, dependence is introduced by filtering a white noise field in a fashion that closely resembles the Slutsky-Yule effect in time series analysis. Again, as in the case of the spatial autoregressive processes, the diagonal elements in Equation (2.85) are not constant even if the error term is a white noise. The model, therefore, defines fields that are not covariance-stationary.

2.4.3.3 The Autoregressive Moving Average Random Field

Definition 33. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ is said to be autoregressive moving average (or SARMA; see Huang, 1984), if the random variable $X(\mathbf{s_i})$ can be expressed as:

$$X(\mathbf{s}_i) = \mu_i + \sum_{i\neq j} \rho_{ij}\left[X(\mathbf{s}_j) - \mu_j\right] + \sum_{i\neq j} m_{ij} u(\mathbf{s}_j) + u(\mathbf{s}_i) \tag{2.87}$$

where m_{ij} is a set of constants so that $m_{ij} = m w_{ij}$; $\rho_{ij} = \rho w_{ij}$, $w_{ij} \in \mathbf{W}$, $\mathbf{W}$ is an appropriate connectivity matrix and $\{u(\mathbf{s_i}), \mathbf{s}\in S\}$ a spatial white noise field.

If we assume that $\mathrm{Var}(X(\mathbf{s_i})) = \sigma^2_i$, and $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & & \\ & & & \\ 0 & & & \sigma_n^2 \end{pmatrix}$, the matrix of variances and spatial auto-covariances resulting for the model $\{\mathbf{X(s)}, \mathbf{s}\in S\}$ is given by:

$$\mathbf{V} = (\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} + \mathbf{M})\Sigma(\mathbf{I} + \mathbf{M})(\mathbf{I} - \mathbf{B})^{-T} \tag{2.88}$$

with $\mathbf{M} \equiv \{ m_{ij} \}$ and $\mathbf{B} \equiv \{ \rho_{ij}\}$. This process represents the spatial analogue of the ARMA models in a time series context. In the simpler case where $\sigma^2_{\mathbf{i}} = \sigma^2 \ \forall$ $\mathbf{s_i}$, the variance covariance matrix reduces to:

$$\mathbf{V} = \sigma^2(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} + \mathbf{M})(\mathbf{I} + \mathbf{M})(\mathbf{I} - \mathbf{B})^{-T} \tag{2.89}$$

2.4.3.4 The Spatial Error Component Random Field

Related to the previous specification is the so-called *spatial error component* model introduced by Keleijan and Robinson (1993, 1995, 1997). This spatial field is very similar to an SMA field, but, rather than being specified in terms of a single white noise term, it contains two independent white noise components.

Definition 34. A random field $\{\mathbf{X(s)}, \mathbf{s}\in S\}$ is called spatial error component (SEC), if the random variable $X(\mathbf{s_i})$ can be expressed as:

$$X(\mathbf{s}_i) = \mu_i + \sum_{i\neq j} m_{ij} v(\mathbf{s}_j) + u(\mathbf{s}_i) \tag{2.90}$$

with $m_{ij} = mw_{ij}$ a set of constants, $w_{ij} \in \mathbf{W}$, $\mathbf{W}$ an appropriate connectivity matrix and $\{u(\mathbf{s_i}), \mathbf{s} \in S\}$ and $\{v(\mathbf{s_i}), \mathbf{s} \in S\}$ two independent spatial white noise fields. The term $\sum_{i \neq j} m_{ij} v(\mathbf{s}_j)$ implies a smoothing of neighbouring values referred to as "regional effect" (Anselin, 2001a), whereas the term $u(\mathbf{s_i})$ is a location-specific innovation effect.

The resulting matrix of variances and spatial auto-covariances is given by (Keleijan and Robinson, 1993; Anselin, 2001a):

$$\mathbf{V} = \sigma^2{}_u \mathbf{I} + \sigma_v^2 \mathbf{M}\mathbf{M}^T \tag{2.91}$$

where $\mathbf{M} \equiv \{ m_{ij} \}$ and $\sigma^2{}_u$ and σ_v^2 are the variances components relative to the innovation and regional effect, respectively.

2.4.3.5 The Direct Representation of a Random Field

An alternative way of specifying a random field model consists of expressing the elements of the variance covariance matrix in a parsimonious way as some function of the distance between pairs of sites. If we adopt this strategy, we can introduce the following definition.

Definition 35. A random field $\{X(\mathbf{s}), \mathbf{s} \in S\}$ is directly represented (DR) (Mardia, 1990; Anselin, 2001a; Anselin *et al.* 2004), if each element of the matrix of variances and spatial auto-covariances between the random variables $X(\mathbf{s_i})$ can be expressed as:

$$\gamma(\mathbf{s}_1, \mathbf{s}_2) = \sigma^2 f(d_{ij}, \boldsymbol{\varphi}) \tag{2.92}$$

with d_{ij} the distance between site *i* and site *j*, $f(.)$ a distance decay function such that $\frac{\partial f}{\partial d_{ij}} < 0$, $\left| f(d_{ij}, \boldsymbol{\varphi}) \right| \leq 1$ and $\boldsymbol{\varphi}$ represents an appropriate vector of parameters. For instance, Mardia and Watkins (1989) proposed the function:

$$\gamma(\mathbf{s}_1, \mathbf{s}_2) = \sigma^2 \left(1 - \frac{d_{ij}}{\alpha} \right)^4 \tag{2.93}$$

Using Equation (2.93) to define each individual element of covariance, the full matrix of variance and spatial auto-covariances is given by:

$$\mathbf{V} = \sigma^2 \boldsymbol{\Omega}(d_{ij}, \boldsymbol{\varphi}) \tag{2.94}$$

which must be positive definite with elements $\omega_{ij} \in \mathbf{\Omega}$ such that $\omega_{ii} = 1$ and $|\omega_{ii}| \leq 1$ for each i and j. Since each element of the variance-covariance matrix is modelled directly, we can use the "direct representation" to overcome the problem of non-stationarity in variance observed with SAR, SMA and SARMA models. In the econometric literature the direct representation has been employed to analyse the urban housing market by, e. g., Dubin (1988, 1992), Olmo (1995) and Basu and Thibodeau (1998) amongst the others. A series of problems emerging with the direct representation have been analysed by Anselin (2001a). Further extensions of this approach are considered within a semi-parametric approach to spatial modelling: a topic that is not treated in details here. See Section 6.2.8 below both for a more thorough discussion and for references.

2.5 Limiting Theorems for Random Fields

2.5.1 Introduction

At the heart of the various problems of estimation, inference and hypothesis testing in econometrics there lies the problem of determining the distribution of a set of parameters' estimators relating to the probability model $\Phi = \{f(X, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$.

Generally speaking, these estimators can be written as functions of the random vector $\mathbf{X}$. The classical results of statistical inference are almost exclusively based on the case of normally and independently distributed random variables, or variables that can be reduced to this paradigm using asymptotic arguments.

The non-independence nature of observations in space, however, precludes the application of most of the traditional inferential theory's results.

On the other hand, even the asymptotic theory (on which much of the inference for dynamic economic models is based), seems to find obstacles when translated into the spatial domain. Indeed, in this context, it is not always clear how spatial data can approach infinity considering that, unlike time, geographical space is inherently limited.

Cressie (1991) provides some clarification in this respect by distinguishing between two types of asymptotic theories applicable to spatial fields. The number of spatial units on which a random process is observed can tend to infinity, either within a domain that also tends to infinity, or within a domain that is still bounded, but with an increasing number of (more densely distributed) units within it. Cressie calls the first type *increasing-domain asymptotic*, and the second *infill asymptotic*. While in the time series context the increasing-domain asymptotic type seems the most appropriate, in spatial econometrics the choice depends on the nature of the problem being addressed. In the case of a discrete-parameter random field (e. g. one observed in geographical partitions of economic space) one can apply the increasing-domain approach by expanding the number of observed

random variables outside the border of the observed area. Alternatively, one can apply an infill asymptotic theory by considering a limited area and levels of increasingly detailed subregional partitions by subsequently disaggregating the existing units. It is important to notice, however, that even in this case it is not always possible to apply a criterion which ensures that the observations tend to infinity in a regular manner: This is due both to the different shape and size of the various spatial units and to the arbitrariness with which a subsequent disaggregation may take place (see Arbia, 1989, for a thorough discussion of this topic).

In the instance of continuous parameter random fields, the infill asymptotic theory is probably conceptually more appropriate. In fact, in this second instance, it is possible to imagine a mechanism which leads to an increasing number of observations within the same domain, given that in this case the economic phenomenon may be observed in any of the infinite points of space, but the domain should often be thought of as delimited.

According to Anselin (2001a) "to date some formal results for the spatial dependence case are still lacking", but "the intuition behind the asymptotics is fairly straightforward in that regularity conditions are needed to limit the extent of spatial dependence (memory) and heterogeneity of the spatial series in order to obtain appropriate (uniform) laws of large numbers and central limit theorems to establish consistency and asymptotic normality" (Anselin, 2001a).

Even with the precautions stated above it is possible to identify conditions within which the asymptotic properties of a random field are valid for inferential purposes. In the remainder of this chapter we will thus consider some extensions of traditional asymptotic theory to spatial random fields.

2.5.2 Some Limit Theorems for Random Fields

Given the importance of the subject, we will now cite some convergence concepts useful for understanding the limit theorems in spatial analysis.

Definition 36. A sequence of random variables $\{X_n(\mathbf{s}), n \in N\}$ is said to converge *almost surely* (or with probability 1) to the random variable $X(\mathbf{s})$, and is indicated as $X_n(\mathbf{s}) \overset{AS}{\to} X(\mathbf{s})$, if:

$$P\{s : \lim_{n\to+\infty} X_n(\mathbf{s}) = X(\mathbf{s})\} = 1 \tag{2.95}$$

Definition 37. We say that a sequence of random variables $\{X_n(\mathbf{s}), n \in N\}$ converges *in probability* to the random variable X(s), and is indicated as $X_n(\mathbf{s}) \overset{P}{\to} X(\mathbf{s})$, if:

$$\lim_{n\to+\infty} P\{s: |X_n(\mathbf{s}) - X(\mathbf{s})| < \varepsilon\} = 1 \tag{2.96}$$

Definition 38. We say that a sequence of random variables $\{X_n(\mathbf{s}),\ n\in N\}$, characterised by distribution function $\{F_n(X),\ n\in N\}$, converges *in distribution* to the random variable $X(\mathbf{s})$, and we indicate this $X_n(\mathbf{s}) \xrightarrow{D} X(\mathbf{s})$, if:

$$\lim_{n\to+\infty} F_n(X) = F(X) \tag{2.97}$$

in all points where $F(X)$ is continuous.

We can now introduce the two fundamental probabilistic laws on which the asymptotic inferential theory is based.

The almost certain convergence criterion is associated with the so-called:

Strong Law of Large Numbers (Kolmogorov Theorem 2): Let $\{X_n,\ n\geq 1\}$ be a sequence of independent random variables so that $E(X_i) = \mu_i$, and $Var(X_i) = \sigma^2_i$ exist for all $i = 1,2,\ldots$ and that $\sum_{k=1}^{\infty} \frac{1}{k^2} Var(X_k) < +\infty$, then:

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} \mu_i\right) \xrightarrow{AS} 0 \tag{2.98}$$

The SLLN states the conditions under which random variables' averages converge to their expected values, and therefore it constitutes the basis for obtaining the distribution of estimators and tests statistics that are built as linear combinations of the observed values.

Once again, the independence hypothesis among random variables considered in the formulation of the law precludes its use in the case of spatial data. However, McLeish (1975) and White (1984) demonstrated that the SLLN can be extended to the case of dependent random variables as long as they obey particular forms of restriction. More specifically, the following result is valid:

Strong Law of Large Numbers for Mixing Fields: Let$\{X_n,\ n\geq 1\}$ be a mixing random field (strongly or uniformly: see Definitions 18 and 19) where we have:

$$\alpha(m) = O(m^{-\delta}) \qquad \delta > \frac{r}{2r-1} \tag{2.99}$$

For strongly mixing fields or, alternatively,

$$\varphi(m) = O(m^{-\delta}) \qquad \delta > \frac{r}{r-1}$$

for uniformly mixing fields, and such that it is dominated by the process $\{Z_n,\ n\geq 1\}$in the sense that $|X_n| \leq Z_n$. Then:

$$(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i) \xrightarrow{AS} 0 \tag{2.100}$$

provided that $\sum_{k=1}^{\infty}\frac{1}{k^2}Var(X_k) < +\infty$, and in addition, $E(|Z^{r+d}|) < D < +\infty$, $n \geq 1$, $\forall d > 0$. In this formulation, the value of r is connected with $\alpha(m)$ and $\varphi(m)$, i. e. with the degree of dependence among the random variables. The higher the degree of dependence, the more the constraints are extended to the higher-order moments.

The criterion of convergence in probability is associated with the so-called:

Central Limit Theorem (CLT): Let $\{X_n, n \geq 1\}$ be a sequence of random variables independently and identically distributed so that $E(X_i) = \mu_i$, and $Var(X_i) = \sigma_i^2 < +\infty \ \forall i = 1,2...,n$ Furthermore, let us define the random variable $Y_n = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$, with $S_n = \sum_{i=1}^{n} X_i$, and characterised by a distribution function $F_n(Y)$. Then:

$$Y_n \xrightarrow{D} Z \sim N(0,1) \tag{2.101}$$

The CLT is the basis for computing the limiting distributions of many estimators and test statistics. Nevertheless, like the SLLN, it cannot be applied as such to sequences of random variables ordered in space, due to the conditions of independence that is imposed. It has, however, been extended to the case of mixed random variables through the following result

Central Limit Theorem for Mixing Fields. Let $\{X_n, n \geq 1\}$ be a (strongly or uniformly) mixing random field where we have:

$$\alpha(m) = O(m^{-\delta}) \ \delta > \frac{r}{2r-1}$$

or (2.102)

$$\varphi(m) = O(m^{-\delta}) \ \delta > \frac{r}{r-1}$$

so that it is dominated by the process $\{Z_n, n \geq 1\}$ in the sense that $|X_n| \leq Z_n$, and so that $E(X_n) = 0$, $n \geq 1$, and $E(|X_n|^{2r}) \leq k < +\infty$. Furthermore, let $S_n(\delta) = \frac{1}{\sqrt{n}}\sum_{i=\delta+1}^{n+\delta} X_i$ be defined so that $\forall V \neq 0$ and $\forall \delta$, it happens that $\lim_{\delta \to +\infty} [E(S_n(\delta))^2 - V] = 0$. Then:

$$\frac{S_n(0)}{\sqrt{nV}} \xrightarrow{D} Z \sim N(0,1) \tag{2.103}$$

(see White and Domowitz, 1984).

A second way of extending the CLT to spatial processes is obtained by introducing the concepts of *regular* random fields and *locally covariant* random fields.

Definition 39. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ with $E(X(\mathbf{s}_i))=\mu_i$, and $Var(X(\mathbf{s}_i))=\sigma^2_i$, is said to be regular if (Smith, 1980):

i) $\sigma_i^2 \geq \delta$ (non-degenerate uniform condition)

and (2.104)

ii) $E(|X(\mathbf{s}_i)-\mu_i|^{2+\delta}) \leq \beta$ (Lyapunov limitation condition)

Definition 40. A random field $\{X(\mathbf{s}), \mathbf{s}\in S\}$ with $E(X(\mathbf{s}_i))=\mu_i$, and $Var(X(\mathbf{s}_i)) = \sigma^2_i < \infty$, is said to be locally covariant if:

$$E(X(\mathbf{s}_i)\, X(\mathbf{s}_j))=\begin{cases} \neq 0 & \text{if } d_{ij} < d^* \\ = 0 & \text{otherwise} \end{cases} \tag{2.105}$$

with d_{ij} any measure of distance between sites $\mathbf{s}_i$ and $\mathbf{s}_j$ (see Section 2.2.1).

The concept of a locally covariant random field (due to Smith, 1980, and studied in detail by Arbia, 1989) is based on a form of restriction on the process dependence analogous to the asymptotic independence of temporary random processes.

In developing these concepts Smith (1980) demonstrated the following result.

Central Limit Theorem for Regular and Locally Covariant Fields. Let $\{X_n, n\geq 1\}$ be a regular and locally covariant random field such that $E(X_i)=\mu_i$, and $Var(X_i) = \sigma_i^2 < +\infty \quad \forall i=1,2,\ldots,n$. Moreover, consider the random variable $Y_n = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$ with $S_n = \sum_{i=1}^{n} X_i$, characterised by the cumulative distribution function $F_n(Y)$. Then:

$$Y_n \xrightarrow{D} Z \sim N(0,1) \tag{2.106}$$

Serfling (1980) further extended CLT to processes presenting more complex forms of dependence than those considered in the present context (see also Wooldridge and White, 1988).

3 Likelihood Function for Spatial Samples

3.1 Introduction

The most common estimation and hypothesis testing procedures in econometrics are those based on the notion of likelihood introduced last century by Sir Ronald Fisher in the 1920s and 1930s and subsequently extended by the work of several statisticians. Before introducing the specificity of spatial likelihood, let us first introduce the basic notation and concepts.

Definition 39. Let Φ be a probability model $\Phi = \left\{ f_{X_{s_1}, X_{s_2}, ..., X_{s_n}}(x_{s_1}, x_{s_2}, ..., x_{s_n}; \boldsymbol{\theta}), \right.$ $\left. s_i \in S, \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}$ and let $\mathbf{X}$ be a random vector, such that $\mathbf{X} \equiv (\mathbf{X}(s_1), \mathbf{X}(s_2), \ldots, \mathbf{X}(s_n))$, relating to an observed sample $\mathbf{x} \equiv (x_{s_1}, x_{s_2}, ..., x_{s_n})$, drawn according to a certain sampling model from the population. The joint probability density function of the sample is defined as $f_{X_{s_1}, X_{s_2}, ..., X_{s_n}}(x_{s_1}, x_{s_2}, ..., x_{s_n}; \boldsymbol{\theta})$, with $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the set of parameters to be estimated. This function represents the probability that, prior to the experiment, we had to draw the sample that we actually observed. The likelihood function is defined as:

$$\mathrm{L} = \mathrm{L}(\boldsymbol{\theta}) = \mathrm{L}(\boldsymbol{\theta}, \mathbf{x}) = c(\mathbf{x}) f_{X_{s_1}, X_{s_2}, ..., X_{s_n}}(x_{s_1}, x_{s_2}, ..., x_{s_n}; \theta) \tag{3.1}$$

with $c(\mathbf{x}) = c(x_{s_1}, x_{s_2}, ..., x_{s_n})$ a function of just the observed data.

Due to the exponential nature of many density functions it is often useful to operate with the logarithmic transformation of the likelihood (or log-likelihood) defined as $\ln[\mathrm{L}(\boldsymbol{\theta}, \mathbf{x})] = l(\boldsymbol{\theta}, \mathbf{x}) = l(\boldsymbol{\theta})$. Associated with the log-likelihood there is a set of so-called *likelihood quantities,* and a set of functions of their moments, that are of interest in statistical inference. In particular, the following definitions are essential in developing estimation and hypothesis testing procedures.

Definition 40. The first derivative of the log-likelihood function with respect to the unknown parameters is defined as the *score* function and is expressed as:

$$\mathbf{q}(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta}, X)}{\partial (\boldsymbol{\theta})} \tag{3.2}$$

Definition 41. The matrix of the second derivative of the log-likelihood function with respect to pairs of parameters is called the Fisher's observed information matrix and is expressed as:

$$\mathbf{j}(\boldsymbol{\theta}) = -\frac{\partial^2 l(\boldsymbol{\theta}, X)}{\partial \theta_l \partial \theta_k} \qquad \theta_l, \theta_k \in \boldsymbol{\theta} \tag{3.3}$$

Definition 42. The expected value of the matrix of the second derivative of the log-likelihood function with respect to pairs of parameters is called the Fisher's expected information matrix and is expressed as:

$$i(\boldsymbol{\theta}) = E\left[j(\boldsymbol{\theta})\right] = Var\left[q(\boldsymbol{\theta})\right] \tag{3.4}$$

Given the previous definitions, it is natural to introduce the following definition.

Definition 43. Given a likelihood function L(θ, x) the maximum likelihood estimator of the parameters θ is the Borel function $\hat{\boldsymbol{\theta}} : \mathbf{x} \to \boldsymbol{\Theta}$, such that:

$$L(\hat{\boldsymbol{\theta}}; \mathbf{x}) = \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}; \mathbf{x}) \tag{3.5}$$

In the case of absolutely continuous and differentiable density functions, the maximum likelihood estimator can be obtained as the solution to the equations:

$$\begin{cases} \dfrac{\partial L(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}} = q(\boldsymbol{\theta}) = \mathbf{0} \\ \dfrac{\partial^2 L(\boldsymbol{\theta}; \mathbf{x})}{\partial \boldsymbol{\theta}^2} < \mathbf{0} \end{cases} \tag{3.6}$$

The maximum likelihood estimators thus obtained satisfy a set of optimal properties. In particular, if independence between observations can be assumed in the sampling model, it has been proved that they are fully-efficient and (under certain regularity conditions on the probability model) consistent, asymptotically unbiased and asymptotically normally distributed (see Azzalini, 1996).

As we have already remarked in Section 2.1, if the observations represent a cross-section of individuals, it is often legitimate to assume the independence of the random variables so that the joint probability density function of the sample may be written as:

$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{X_i}(x_i; \boldsymbol{\theta}) \tag{3.7}$$

On this basis the log-likelihood of the sample may be written as:

$$l(\boldsymbol{\theta}) = \ln\left[c(\mathbf{x}) f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n; \boldsymbol{\theta})\right] = c(\mathbf{x}) + \ln\left[\prod_{i=1}^{n} f_{X_i}(x_i; \boldsymbol{\theta})\right] =$$

$$= c(\mathbf{x}) + \sum_{i=1}^{n} \ln\left[f_{X_i}(x_i; \boldsymbol{\theta})\right] \tag{3.8}$$

where the term $\ln\left[f_{X_i}(x_i; \boldsymbol{\theta})\right]$ represents the log-likelihood for a single observation x_i.

For observations drawn from a time series it is no longer legitimate to assume the independence of the component random variables. Nevertheless, it is possible to break down the joint probability density function of the sample (see again Section 2.1) as:

$$f_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}; \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{X_{t_i} | X_{t_1}, \ldots, X_{t_i - 1}}(x_{t_i} | x_{t_1}, x_{t_2}, \ldots, x_{t_{i-1}}; \boldsymbol{\theta}) \tag{3.9}$$

The log-likelihood can, therefore, be written as:

$$l(\boldsymbol{\theta}) = \ln\left[c(\mathbf{x}) f_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_{t_1}, x_{t_2}, \ldots, x_{t_n}; \boldsymbol{\theta}\right] =$$

$$= c(\mathbf{x}) + \ln\left[\prod_{i=1}^{n} f_{X_{t_i} | X_{t_1}, \ldots, X_{t_i - 1}}(x_{t_i} | x_{t_1}, x_{t_2}, \ldots, x_{t_{i-1}}; \boldsymbol{\theta})\right] =$$

$$= c(\mathbf{x}) + \sum_{i=1}^{n} \ln f_{X_{t_i} | X_{t_1}, \ldots, X_{t_{i-1}}}(x_{t_i} | x_{t_1}, x_{t_2}, \ldots, x_{t_{i-1}}; \boldsymbol{\theta}) \tag{3.10}$$

where $\log f_{X_{t_i} | X_{t_1}, \ldots, X_{t_{i-1}}}(x_{t_i} | x_{t_1}, x_{t_2}, \ldots, x_{t_{i-1}}; \boldsymbol{\theta})$ now represents the log-likelihood for a single observation x_{t_i} conditional on its past.

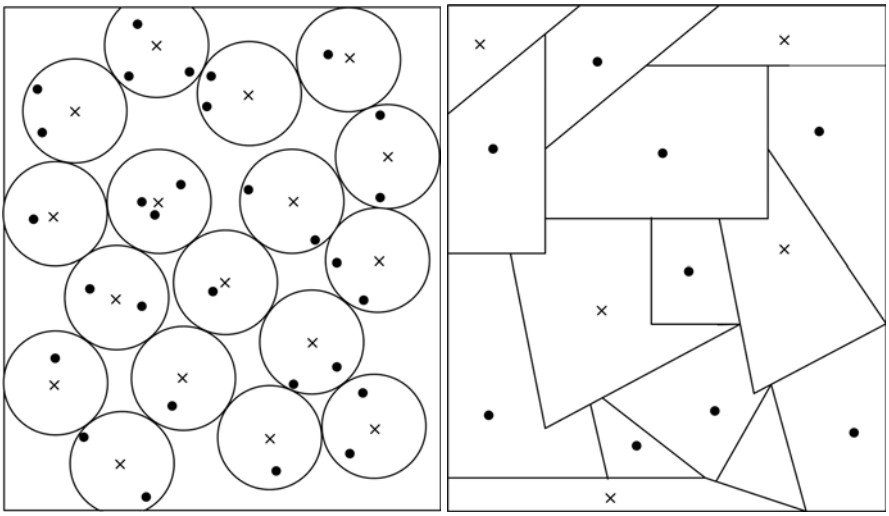
As regards observations drawn from a spatial random field, however, it is not possible to break down the joint probability density function of the sample in a manner similar to that seen for cross-sections or time series, given the greater complexity of dependence in space. Two alternative appoaches can be followed in this respect. The first consists of obtaining some approximate solutions in order to derive the likelihood from the marginal or the conditional densities of the random field. This approach will be followed in this chapter. The second approach consists of deriving the full likelihood on the basis of one of the models for random fields presented in the previous Section 2.4. This second approach will be presented in Chapter 4 when discussing the various estimation and hypothesis testing procedures within the context of regression modelling.

3.2 Some Approximations for the Likelihood of Random Fields

3.2.1 The Coding Technique

A first way of proceeding for deriving the likelihood function based on spatial samples is the so-called *coding technique* suggested by Besag (1972). The technique is based on the consideration that if a random field is Markovian (see Section 2.4.2) then pairs of random variables associated with two non-neighbouring locations (either points or regions) are mutually independent conditionally upon the remaining random variables. Consequently, by selecting *ad hoc* a subset of locations which are not neighbours, the likelihood can be derived in terms of the product of the conditional probability density functions.

To better illustrate the coding technique, let us look at Figure 3.1.



(a) (b)

Fig. 3.1. Coding model for a continuous-parameter field (a) and for a discrete-parameter field (b).

In Figure 3.1 each location is labelled either with a cross or with a circle in such a way that the locations labelled with a cross do not neighbour one another. Let us call Q the set of location labelled with a cross and let us assume that its cardinality is $\#[Q] = q$. In this way, if the observed field is Markovian, the random variables associated with the location cross-coded are all mutually

conditionally independent and, as a consequence, the joint probability density function of the sample can be expressed as:

$$f_{X_{s_1},X_{s_2},...,X_{s_n}}(x_{s_1},x_{s_2},...,x_{s_n};\boldsymbol{\theta}) = \prod_{\substack{i=1\\ i\in Q}}^{q} f_{X_{s_i}|X_{s_j}}(x_{s_i}\,|\,x_{s_j}, s_j \in N(i);\boldsymbol{\theta}) \tag{3.11}$$

In this way, we can obtain the conditional log-likelihood of the sample as:

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log\left[c(\mathbf{x}) f_{X_{s1},X_{s2},.....X_{sn}}(x_{s_1},x_{s_2},.....x_{s_n};\boldsymbol{\theta})\right] = \\ &\ln\left[\prod_{\substack{i=1\\ i\in Q}}^{q} c(\mathbf{x}) f_{X_{s_i}|X_{s_j}}(x_{s_i}\,|\,x_{s_j}, s_j \in N(i);\boldsymbol{\theta})\right] \\ &= c(\mathbf{x}) + \sum_{\substack{i=1\\ i\in Q}}^{q} \ln f_{X_{s_i}|X_{s_j}}(x_{s_i}\,|\,x_{s_j}, s_j \in N(i);\boldsymbol{\theta}) \end{aligned} \tag{3.12}$$

where $\ln f_{X_{s_i}|X_{s_j}}(x_{s_i}\,|\,x_{s_j}, s_j \in N(i);\boldsymbol{\theta})$ represents the log-likelihood of the single observation x_{s_i} (i=1, 2, ..., q) conditional on its neighbours.

Obviously, in any specific case, there are many possible alternatives for coding the locations and, as a result, all the inferential conclusions will be dependent on the choice made. It is, therefore, necessary to test how the results are robust to the choice made when applying this methodology to empirical cases.

The basic philosophy behind the technique described is that it is preferable to discard information (in our case, the locations marked as *circles*) rather than lose the optimal properties of the estimates and tests based on likelihood. In this sense, the technique is applied in cases where a large set of data is available (so that the loss of information does not compromise the number of degrees of freedom) from complete surveys (e. g. surveys related to individual economic agents, or complete coverage of regional data). The technique is also recommended in cases where the data come from sampling surveys and it is possible to control the sampling design by introducing non-neighbourhood constraints between the locations (see, e. g., the sampling plans suggested by Arbia and Switzer, 1994; Arbia, 1995a; and Arbia and Lafratta, 1997, 2002).

3.2.2 The Unilateral Approximation

An alternative way of deriving the likelihood function consists of constructing (starting from the observed field) a random field which has approximately the same probabilistic structure, but which is simpler to treat. In order to achieve this aim, let us first consider the following definitions.

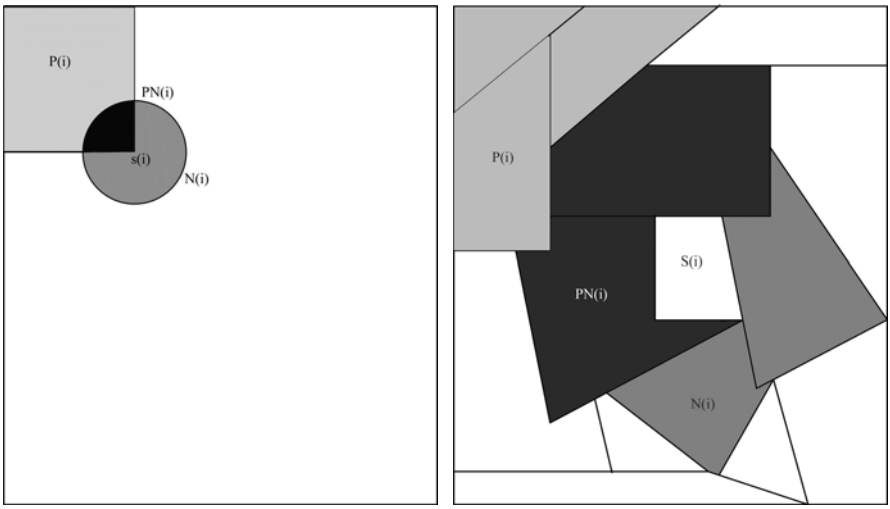
Definition 44. Given a set of locations ($\mathbf{s}_1$, $\mathbf{s}_2$,…, $\mathbf{s}_n$), $\mathbf{s}_i = (s_{1i}, s_{2i})$, then, for each location $\mathbf{s}_i$ let us define the set of *predecessors* of $\mathbf{s}_i$, say *P(i)*, as the set of locations $\mathbf{s}_j$ such that:

$$P(i) = \left\{ \mathbf{s}_j : (s_{1j} \le s_{1i}) \cap (s_{2j} < s_{2i}) \right\} \tag{3.13}$$

Definition 45. Given a set of locations ($\mathbf{s}_1$, $\mathbf{s}_2$, ..., $\mathbf{s}_n$), $\mathbf{s}_i = (s_{1i}, s_{2i})$, given the neighbouring structure $N(i)$ and the definition of predecessors $P(i)$, provided by Definition 44, for each location $\mathbf{s}_i$ let us define the set of *predecessors-neighbours* of $\mathbf{s}_i$, say $PN(i)$, as the set of locations $\mathbf{s}_j$ such that:

$$PN(i) = \left\{ \mathbf{s}_j : P(i) \cap N(i) \right\} \tag{3.14}$$

See Figure 3.2 for an illustration.



(a) (b)

Fig. 3.2. Predecessors of location $\mathbf{s}_i$ ($P(i)$), neighbours of $\mathbf{s}_i$ ($N(i)$) and predecessor-neighbours ($PN(i)$) of $\mathbf{s}_i$ in the case of continuous-parameter (a) and discrete-parameter (b) fields.

If we suppose that the random field under study is isotropic (see Section 2.3.1), then its dependence structure will be identical in all four directions originated in the four corners of the area (NW, NE, SE, SW). Thus it is not limiting to study the random field based on the random variables which fall (e. g.) into the upper left quadrant of every $\mathbf{s}_i$ (see Figure 3.2).

In this way, we can generate a unilateral random field $\{X(\mathbf{s}),\ \mathbf{s}\in\delta\}$ by specifying each component random variable conditionally upon the values of its *predecessors*. Moreover, if the field is of a Markov type, the distribution of each $X(\mathbf{s}_i)$ can be specified in terms of just those predecessors $P(i)$ which are also the neighbours of $\mathbf{s}_i$. This type of field represents the natural extension of the classic time series autoregressive process as long as we are able to specify a reasonable start-up as the initial value. (For these fields see Bartlett and Besag, 1969; Bartlett, 1971; Besag, 1972; 1974).

The great advantage of using a unilateral approximation lies in the fact that the likelihood function can now be written using a breakdown similar to that considered for time series, i. e.:

$$
\begin{aligned}
&l(\theta) = \ln\left[c(\mathbf{x}) f_{X_{s_1},X_{s_2},\ldots,X_{s_n}}(x_{s_1},x_{s_2},\ldots,x_{s_n};\boldsymbol{\theta})\right] = \\
&\ln\left[\prod_{i=1}^{n} c(\mathbf{x}) f_{X_{s_i}\left|X_{s_j}\right.}(x_{s_i}\left|x_{s_j};\mathbf{s}_j \in PN(i);\boldsymbol{\theta}\right)\right] = \\
&c(\mathbf{x}) + \sum_{i=1}^{n} \ln f_{X_{s_i}\left|X_{s_j}\right.}(x_{s_i}\left|x_{s_j};\mathbf{s}_j \in PN(i);\boldsymbol{\theta}\right)
\end{aligned}
\tag{3.15}
$$

where $f_{X_{s_i}\left|X_{s_j}\right.}(x_{s_i}\left|x_{s_j};\mathbf{s}_j \in PN(\mathbf{s}_i);\boldsymbol{\theta}\right)$ now assumes the meaning of the log-likelihood for a single observation x_{s_i} conditional on $\mathbf{s}_j$ which are *predecessors-neighbours* of $\mathbf{s}_i$.

While this technique provides an excellent solution to the problem of computing a likelihood in the case of continuous-parameter random fields, it does not seem advisable in the case of discrete-parameter random fields, given the necessary arbitrariness in the definition of predecessors in cases where the partition is particularly complex and the polygons are of irregular shape and size.

3.2.3 The Pseudo-Likelihood *à la Besag*

The previously discussed approximations are subject to some limitations. For data deriving from complete surveys or surveys based on rigorous statistical design and drawn from a Markov field, the coding technique seems to be the most appropriate. Conversely, in the case of Markov isotropic fields specified with a continuous-parameter, a unilateral approximation can be used.

If none of these conditions occurs, the only solution appears to be based on the so-called *Besag's pseudo-likelihood* (PL). As is well known, a pseudo-likelihood is a function that, even if not a properly defined likelihood, satisfies all or some of the properties of a likelihood function (Pace and Salvan, 1997). Besag (1974)

suggested the use in a spatial context of a pseudo-likelihood function simply defined as the product of the various conditional densities or:

$$Pl(\boldsymbol{\theta}) = \ln\left[c(\mathbf{x}) f_{X_{s_1}, X_{s_2}, \ldots, X_{s_n}}(x_{s_1}, x_{s_2}, \ldots, x_{s_n}; \boldsymbol{\theta})\right] =$$
$$\ln\left[\prod_{i=1}^{n} c(\mathbf{x}) f_{X_{s_i}|X_{s_j}}(x_{s_i} | x_{s_j}; \mathbf{s}_j \neq \mathbf{s}_i; \boldsymbol{\theta})\right] =$$

$$= c(\mathbf{x}) + \sum_{i=1}^{n} \ln f_{X_{s_i}|X_{s_j}}(x_{s_i} | x_{s_j}; \mathbf{s}_j \neq \mathbf{s}_i; \boldsymbol{\theta}) \tag{3.16}$$

Geman and Graffigne (1987) proved that the technique leads to consistent estimators of the unknown parameters. The element that appears to be most in favour of Besag's pseudo-likelihood, however, is the very practical circumstance that it provides good results in real cases (Ripley, 1990).

3.2.4 Computational Aspects

When the above methods were proposed during the 1970s, there were still huge computational limits when using the maximum likelihood method for parameters estimation and building hypothesis testing procedures. These problems now seem to be obsolete, since it is possible to solve the problem fairly quickly using numeric approximations (see Ripley, 1990).

In the case where the number of locations constituting the field is very high, however, it is still necessary to consider potential computational problems when deriving the likelihood function. By way of example, let us consider the case of an auto-normal (CAR) field (see Definition 28), that is a MVN($\boldsymbol{\mu}$, $\mathbf{V}$) field with mean vector $\boldsymbol{\mu} \equiv (\mu_1, \mu_2, \ldots, \mu_n)$ and variance-covariance matrix $\mathbf{V}$, such that $\mathbf{V} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Sigma}$. Up to a few years ago, calculation of the determinant $|\mathbf{I} - \mathbf{B}|$ was considered prohibitive even with a small number of locations. In fact, in many practical applicatons, the matrix $\mathbf{B}$ is scattered and quasi-singular (see e. g. Arbia, 1986). If we define λ_i as the eigenvalues of the matrix $\mathbf{W}$, Ord (1975) suggested using the breakdown:

$$|\mathbf{I} - \mathbf{B}| = |\mathbf{I} - \beta\mathbf{W}| = \prod_{i=1}^{n} (1 - \beta\lambda_i) \tag{3.17}$$

For other approximations see Mead (1967) and Whittle (1954) and, more recently, Smirnov and Anselin (2001), Griffith (2000, 2004), Pace and LeSage (2004), Pace and Zou (2000), Pace (1997), Pace and Barry (1997b, 1997c).

3.3 Maximum Likelihood Estimation Properties in Spatial Samples

As is known, the maximum likelihood estimators satisfy a set of optimal properties under the hypotheses of the probability model's regularity and the sampling model's independence (Azzalini, 1996). In the case of dependent observations, Bates and White (1985) and Heijmans and Magnus (1986) demonstrated that the maximum likelihood estimates maintain their properties of consistency, full efficiency, asymptotic unbiasedness and normality, even if they are based on observations drawn from random processes with non-independent components if, in addition, the following conditions hold:

(i) $L(\theta)$ exists

(ii) the following derivatives exist $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$

$$L'(\theta) = \frac{\partial L(\theta, X)}{\partial \theta} \; ; \; L''(\theta) = \frac{\partial^2 L(\theta, X)}{\partial^2 \theta} \text{ and } L'''(\theta) = \frac{\partial^3 L(\theta, X)}{\partial^3 \theta}$$

(iii) $L'(\theta) = \frac{\partial L(\theta, X)}{\partial \theta} < \infty$

(iv) $\mathbf{V}$ exists and is non-singular; with $\mathbf{V} = \{\gamma(\mathbf{s}_i, \mathbf{s}_j)\}$ the variance-covariance matrix of the random field generating the data.

From the Maximum Likelihood estimators' asymptotic efficiency it follows that their asymptotic variance achieves the Cramer and Rao lower bound:

$$\text{Var}(\boldsymbol{\theta}) = \mathbf{i}_{\infty}(\boldsymbol{\theta})^{-1} = [\lim_{n \to \infty}(\frac{1}{n} \mathbf{i}_n(\boldsymbol{\theta}))]^{-1} \tag{3.18}$$

with $\mathbf{i}_n(\boldsymbol{\theta})$ being Fisher's expected information matrix based on n observations. However, as stated by Anselin (1988), this expression can only be obtained in an explicit analytic form in very particular spatial fields.

3.4 Tests Based on Likelihood

The most commonly used statistical testing procedures in econometrics derive from the notion of likelihood and are incorporated in three fundamental criteria leading to three asymptotically equivalent uniformly most powerful tests. These are the likelihood ratio, the Lagrange multiplier (or score test) and Wald's criterion (Greene, 2003). Let us now summarize the main definitions related to these tests.

Definition 46. Let the likelihood function be $L(\boldsymbol{\theta}; X)$, θ_0 the vector of the parameters under the null hypothesis H_0 and $\hat{\boldsymbol{\theta}}$ the maximum likelihood estimator of $\boldsymbol{\theta}$. The ratio:

$$\lambda(\mathbf{X}) = \frac{L(\boldsymbol{\theta}_0; X)}{L(\hat{\boldsymbol{\theta}}; X)} \tag{3.19}$$

is termed the *likelihood ratio*. Since a monotonic transformation does not alter the conclusions of the test, it is equivalent to consider the transformation:

$$LRT = -2 \log \lambda(\mathbf{X}) \tag{3.20}$$

which represents the *likelihood ratio test.* It can be proved that, under the hypotheses of regularity required for the optimality of the maximum likelihood estimators, the likelihood ratio test is distributed asymptotically under H_0 as a $\chi^2(m)$ with m degrees of freedom, where m is the number of constraints under the null hypothesis.

Definition 47. Under the usual assumptions of regularity, by expanding the likelihood ratio test according to Taylor's series around $\hat{\boldsymbol{\theta}}$ and stopping at the second term, we obtain:

$$LRT = n\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^2 i(\boldsymbol{\theta}_0) + o_p(1) \tag{3.21}$$

with $\mathbf{i}(\boldsymbol{\theta})$ being Fisher's expected information and $o_p(1)$ a term asymptotically of order 1. The term

$$WT = n\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^2 \mathbf{i}(\boldsymbol{\theta}_0) \tag{3.22}$$

is an alternative test called the *Wald test*. It can be seen from Equation (3.21) that WT is asymptotically equivalent to the LRT and, as a consequence, it is also distributed asymptotically as a $\chi^2(m)$.

Definition 48. A third alternative testing procedure can be obtained by expanding the score function according to Taylor's series. This leads to:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \frac{\mathbf{q}(\boldsymbol{\theta}_0)}{\sqrt{n}\mathbf{i}(\boldsymbol{\theta}_0)} + o_p(1) \tag{3.23}$$

which, substituted into Equation (3.21) gives:

$$WT = \frac{\mathbf{q}(\boldsymbol{\theta}_0)^2}{n\mathbf{i}(\boldsymbol{\theta}_0)} + o_p(1) \tag{3.24}$$

The term

$$LMT = \frac{\mathbf{q}(\boldsymbol{\theta}_0)^2}{\mathbf{n}i(\boldsymbol{\theta}_0)} \tag{3.25}$$

is asymptotically equivalent to the LRT and the WT. It was termed the "*score test*" by Rao (1948) and the "*Lagrange multiplier test*" by Aitchinson and Silvey (1958, 1960) and by Silvey (1959). Due to the asymptotical equivalence to the other two tests, the LMT, too, has an asymptotic distribution of $\chi^2(m)$.

There are many reasons for considering all three procedures even though they lead to asymptotically equivalent inferential results. First of all, it is certainly true that in large samples if we apply any of them to the same system of hypotheses will yield exactly the same results in terms of the statistical decision. It should not be forgotten, however, that the three tests can lead to quite different results when dealing with small samples. Secondly, none of the three tests dominates in absolute the others in terms of their power when dealing with small samples. Hence, different situations may require different choices. Thirdly, there are some relative advantages/disadvantages in choosing one test over the other in particular circumstances. For instance, the Wald test is not invariant to transformations and this may be seen as a drawback in some inferential procedures. Similarly, the score test has the advantage of not requiring the evaluation of the point estimators prior to its computation. This advantage is only an apparent one because in most inferential processes the point estimation phase often preceeds that of hypothesis testing.

The asymptotic distribution of the three tests presented above is based on the optimal properties of the maximum likelihood estimator and, in particular, on those properties deriving from the estimators' asymptotic normality.

Therefore, if the conditions for the optimality of the maximum likelihood estimators are valid and, furthermore, the conditions for applying the limit theorems to the case of random fields are verified (see Section 2.5), the tests also maintain their validity in the case of spatial observations.

The three criteria introduced here will be used in the next chapter to test hypotheses on the linear model with spatial data. Most of the procedures employed in random fields hypothesis testing are based on the likelihood ratio and Wald's criteria. Tests based on the Lagrange multiplier have been proposed at a theoretical level (Anselin, 1988), but, to date, have never been used in practical cases.

3.5 Tests Based on Residual Sums of Squares

A very popular way of deriving test statistics in regression analysis is based on the comparison between the residual sums of squares obtained under two constrasting situations; i. e. the case in which the null hypothesis holds true and that in which a properly specified alternative hypothesis is valid.

If $\hat{\boldsymbol{\beta}}_0$ denotes the set of estimates under the null hypothesis and $\hat{\boldsymbol{\beta}}_1$ the set of estimates under the alternative hypothesis, then an F statistic for testing H_1 against H_0 can be computed by

$$\frac{n-r}{r}\frac{\left[RSS(\hat{\boldsymbol{\beta}}_1)-RSS(\hat{\boldsymbol{\beta}}_0)\right]}{RSS(\hat{\boldsymbol{\beta}}_0)} \tag{3.26}$$

with n the number of observations, r the number of constraints under the null hypothesis, $RSS(\hat{\boldsymbol{\beta}}_1)$ the residuals' sum of squares under the alternative hypothesis and $RSS(\hat{\boldsymbol{\beta}}_0)$ the residuals' sums of squares under the null.

In cases when the random variables involved in the regression are jointly normal, the quantity expressed in Equation (3.26) is distributed as a Fisher's F with r and n-r degrees of freedom.

4 The Linear Regression Model with Spatial Data

4.1 Introduction

The linear regression model represents the benchmark for other, more sophisticated statistical models of econometric relationships. In this chapter we will begin by presenting the basic concepts of linear regression and then introduce the specificity arising when estimating it on the basis of spatially distributed data. We shall be discussing into details only the standard linear regression model and its more common violations in the spatial context. We are nevertheless confident that the framework presented here is sufficiently general for it to be easily extended to more complex models simply by adapting the notions introduced in Chapter 2 when discussing models of random fields.

Let $Y(\mathbf{s}_i)$ be the dependent variable of the model at location $\mathbf{s}_i$, $\mathbf{X}(\mathbf{s}_i)$ a vector of esplicative variables of dimension k (including the constant term) and $\mathbf{Z}(\mathbf{s}_i) = [Y(\mathbf{s}_i), \mathbf{X}(\mathbf{s}_i)^T]^T$ a collection of random variables belonging to the vector random field $\{\mathbf{Z}(\mathbf{s}) \ \mathbf{s} \in \delta\}$ defined on the probability space $(\Omega, B, P(.))$ which generates a set of data observed in n locations of coordinates $(\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n)$ on a continuous or a discrete space. Let us assume that we want to build up a model which explains the behaviour of the economic variable $Y(\mathbf{s}_i)$ in location $\mathbf{s}_i$ in terms of the behaviour of the other random variables $\mathbf{X}(\mathbf{s}_i)$ constituting the random field. Having clarified the meaning of different spatial indices $\mathbf{s}_i$ in the previous chapters, henceforth we shall simplify the notation by indicating the random field $\mathbf{Z}(\mathbf{s}_i) = [Y(\mathbf{s}_i), \mathbf{X}(\mathbf{s}_i)^T]^T$ as $\mathbf{Z}_i = (Y_i, \mathbf{X}_i^T)^T$ and the sample observations $\mathbf{z}(\mathbf{s}_i) = [y(\mathbf{s}_i), \mathbf{x}(\mathbf{s}_i)^T]^T$ as $\mathbf{z}_i = [y_i, \mathbf{x}_i^T]^T$.

4.2 Specification of a Linear Regression Model

In this section we will summarize the basic assumptions underlying a linear regression model. Two alternative specifications of the model will be presented. In Section 4.2.1 we will present a model based on the full specification of the joint behaviour of the variables $\mathbf{Z}_i = (Y_i, \mathbf{X}_i^T)^T$ in which we model the systematic part through conditional expectations. We will refer to this as the conditional specification. In Section 4.2.2 we will consider the more standard textbook specification based on the modelling of the stochastic error component. Both specifications will be then used when discussing the violations of the hypotheses when using spatial samples.

4.2.1 The *Conditional* Specification

We will now summarize the basic assumptions on which a linear regression model is based. We will group the hypotheses into three blocks: (i) hypotheses concerning the probability model (PM), (ii) hypotheses concerning the statistical generating mechanism (GM), and (iii) hypotheses on the data sampling model (SM).

4.2.1.1 Hypotheses on the Probability Model (PM)

The fundamental assumption on which the standard linear regression model is based is that the joint distribution of the random variables involved (both Y_i and the explicative variables $\mathbf{X}_i$) is multivariate Gaussian, i. e.:

$$\text{PM } \Phi = \left\{ f_{\mathbf{Z}_i}(\mathbf{Z}_i = \mathbf{z}_i, \boldsymbol{\theta}_i); \boldsymbol{\theta} \in \Theta; \Theta \subset \Re^{k+1} \right\} \qquad \text{and} \qquad Z_i \sim \text{MVN}$$

where Φ represents a parametric family of density functions, θ the associated parametrization and Θ the parametric space.

All other hypotheses concerning the probability model are consequences of this basic assumption. In fact the normality of the conditional distributions follows directly from the joint normality. Hence

$$\text{PM1} \quad f_{Y_i|\mathbf{X}_i}(Y_i = y_i | \mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_i) \sim \text{N}$$

And, from PM1, the linearity of the expected value (regression function),

$$\text{PM2} \quad E(Y_i = y_i | \mathbf{X}_i = \mathbf{x}_i) = \boldsymbol{\beta}^T \mathbf{x}_i$$

$\boldsymbol{\beta}$ being a *k*-by-1 column vector of parameters and $\mathbf{x}_i$ a *k*-by-1 column vector of observations of the random variables $\mathbf{X}_i$. From PM1 it also follows the constancy of the conditional variance (also called the skedasticity function) independently of the value of $\mathbf{x}_i$ (homoskedasticity):

$$\text{PM3} \quad \text{Var}(Y_i | \mathbf{X}_i = \mathbf{x}_i) = \sigma^2 \qquad \forall \mathbf{x}_i$$

and, finally, the hypothesis of the spatial invariance of the parameters

$$\text{PM4} \quad \boldsymbol{\theta}_i \equiv (\boldsymbol{\beta}, \sigma^2) = \boldsymbol{\theta} \qquad \forall i$$

PM1 to PM4 summarize all the aspects of the Probability Model that are necessary to build the statistical inferential procedures.

4.2.1.2 Hypotheses on the Statistical Generating Mechanism (GM)

The basic assumption behind the statistical generating mechanism is that it is constituted by a systematic (forecastable) component and a non-systematic (non forecastable) component. In the basic linear regression model, the two components are combined linearly. As it is obvious, this linearity has nothing to do with the linearity of the conditional mean assumed under PM2 as a consequence of the postulated multinormality. In particular, the systematic component (say μ_i) is represented by the conditional expectation of y_i given $\mathbf{x}_i$ whilst the non-systematic component is simply the unexplained part of the model, measured by the difference between the observed value and the systematic component. We have therefore:

GM1 $\mu_i = E(Y_i | \mathbf{X}_i = \mathbf{x}_i)$
$u_i = Y_i - \mu_i$
$Y_i = \mu_i + u_i = E(Y_i | X_i = x_i) + u_i$

From the linearity of the mean, assumed in PM2, we then have :

$Y_i = \boldsymbol{\beta}^T \mathbf{x}_i + u_i$

From the normality postulated in the PM, we also have that the model's only parameters of interest are represented by the vector $\boldsymbol{\beta}$ and the conditional variance of y given $\mathbf{X}$ (say σ^2).

GM2 $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}; \sigma^2)$; $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \Re^{k+1}$

Of course the parametrization depends on the specific hypotheses made for the PM. Changing the PM (as we will do when introducing some of the procedures to treat spatial data) will, in general, result in a different parametrization.

The fact of concentrating on modelling the conditional distribution of Y_i given $\mathbf{X}_i = \mathbf{x}_i$ (and, specifically, the expected value of this distribution) rather than modelling their joint distribution directly, implies that in the identity:

$$f_{\mathbf{Z}_i}(\mathbf{z}_i, \boldsymbol{\theta}_i) = f_{Y_i|\mathbf{X}_i}(y_i | \mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_i)\ f_{X_i}(\mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_i)$$

we deliberately decide to neglect all relevant information contained in the marginal density $f_{\mathbf{x}_i}(\mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_i)$. Operationally, this means that all information contained in $f_{\mathbf{x}_i}(\mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_i)$ is inferentially irrelevant with respect to the problem of estimating the vector of parameters of interest $\boldsymbol{\theta}$. This simplifying hypothesis leads to the following assumption:

GM3 $\mathbf{X}_i$ is weakly exogenous with respect to $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}; \sigma^2)$

See Engle et al. (1983) for details on exogeneity.

Furthermore, no restrictions are imposed a priori on the range (if deterministic) or on the distribution (if stochastic) of the parameters $\boldsymbol{\theta}$.

GM4 We do not have a priori information on $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}; \sigma^2)$.

Finally, for reasons related to the estimation procedures and, in particular, to avoid the singularity of the variance-covariance matrix, we assume that the observed data matrix is of full rank, whatever the observed sample, i. e.:

GM5 Rank $(\mathbf{X}) = k$

with $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2,...,\mathbf{x}_n)^T$, $\mathbf{x}_i$ a k-by-1 vector of observations and $\mathbf{X}$ an n-by-k data matrix such that $n > k$.

Hypotheses GM1 to GM5 represent what we call the Statistical Generating Mechanism.

4.2.1.3 Hypotheses on the Sampling Model (SM)

The basic assumption behind the sampling model is that the data are drawn with a simple random criterion from the conditional distribution of Y given $\mathbf{X}$. In other words:

SM1 $\mathbf{y} \equiv (y_1, y_2,...,y_n)^T$ is an independent sample drawn from the density $f_{Y_i|\mathbf{X}_i}(y_i|\mathbf{X}_i = \mathbf{x}_i, \boldsymbol{\theta}_i)$, i = 1,.....,n

The set of hypotheses presented above defines all the elements required to implement statistical inferential procedures. The distinction between assumptions regarding the probability sphere, the data generation process and the sampling scheme is purely for clarity of presentation and to facilitate the discussion of the violations of the basic assumptions. However, the reader may be more familiar with an alternative and equivalent specification of the linear regression more commonly found in standard econometric textbooks, in which the three aspects are treated together. Before discussing the basic hypotheses in the case of samples drawn from a spatial field, we will therefore review this second specification and compare it with the one presented above.

4.2.2 Standard Textbook Specification

An equivalent formulation of the linear regression model more commonly used in econometrics manuals (and which we will use in some circumstances in this book) concentrates on the modelling of a stochastic error component and derives from it all the hypotheses on the PM and on the SM.

In this second formulation, a linear relationship between y_i and $\mathbf{x}_i$ is assumed *a priori* (and not derived from the Probability Model as before) and all the probabilistic assumptions are concentrated on a stochastic error term rather than directly on the stochastic relationship between y_i and the $\mathbf{x}_i^T$.

In particular, the linearity assumption first of all implies

A1 $y_i = \boldsymbol{\beta}^T \mathbf{x}_i + u_i$

or, more compactly:

A1 $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$

with $\mathbf{u}$ a n-by-1 column vector of the error terms and $\mathbf{X}$ an n-by-k data matrix ($n > k$). Assumption A1 parallels assumption GM1 in the previous specification.

Furthermore, we assume that the error component, conditionally upon the observed values of $\mathbf{X}$, is distributed in a Gaussian form with zero expected value and constant variance

A2 $f_{\mathbf{u}}(\mathbf{u}|\mathbf{X}) \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$

$\mathbf{I}_n$ being an n-by-n identity matrix.

Note that A2 implies assumptions PM1 to PM4 and SM1 of the previous alternative formulation.

Finally, parallel to assumption GM4 of the previous formulation, we assume that no restrictions are imposed *a priori* on the parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \Re^{k+1}$

A3 we do not have any prior information on $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \sigma^2)$

And, parallel to GM5, we postulate that the observed data matrix is of full rank whatever is the observed sample,

A4 Rank $(\mathbf{X}) = k$

with $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n)^T$ an n-by-k matrix containing the sample information ($n > k$).

In line with our proposed aim of describing the basic methodologies for testing spatial relationships, we are not going to discuss here the whole set of assumptions on which the linear regression model is based in the two alternative formulations. On the contrary, we are going to limit ourselves to considering only those violations which assume a precise meaning and have an easy interpretation in cases where the empirical data come from spatial observations.

In particular, we are not going to discuss the assumptions relating to the data generation statistical model (GM). The next sections of this Chapter will concentrate on the assumptions relating just to the Probability Model (PM) and the Sampling Model (SM). More specifically, Section 4.3 will discuss the most crucial hypothesis (systematically violated with spatial data) of independence of the spatial observations. Section 4.4 will discuss the violations of the hypotheses behind the probability model and, in particular, those connected with spatial homoskedasticity and the spatial invariance of parameters that are more frequently encountered in empirical instances in economic analysis.

4.3 Violation of the Hypotheses on the Sampling Model

4.3.1 Introduction

The SM1 assumption of simple random sampling is certainly the most important among those the linear regression model is based on. In practice none of the results relating to the estimation and hypothesis testing remain valid if they are rejected on the basis of empirical data. More precisely, the implications of the violation of the simple random sampling hypothesis on SM are that the OLS estimates of β and σ^2 are inefficient and inconsistent even if still unbiased. Moreover, the sampling variances are biased and in most cases significantly underestimated. As a consequence, the coefficient of determination (R^2) as well as the test statistics t and F tend to be inflated, leading to accept the model more frequently than it should (Maddala, 2001). As stated in Section 1.1, many econometric textbooks are aware of the problem of non independence arising when dealing with spatially collected data. Kmenta (1997) warns that: "In many circumstances the most questionable assumption […] is that the cross-sectional units are mutually independent […] when the cross-sectional units are geographical regions […] we would not expect this assumption to be well satisfied". Similarly, Maddala (2001) points out that: "In cross-section data [correlation] can arise among contiguous units […] Similarly, if our data are on states, the error terms for contiguous states tend to be correlated". Woolridge (2002a), too, is aware of the problem when he says that "… cross section data that are not the result of independent sampling can be difficult to handle. *Spatial correlation*, or, more generally, *spatial dependence*, typically occurs when […] data are collected at the county, state, province, or country level. Outcomes from adjacent units are likely to be correlated". In this respect, the quotes from Baltagi (2001), Gujarati (2003), Johnston (1991) and Kennedy (2003) cited in Chapter 1.1 should also be recalled.

At the level to which this book has been kept we will limit ourselves to the Gaussian paradigm. Thus the study of how the ideal situation of independence in space is violated reduces to the simpler study of the correlation between spatial units or *spatial correlation*. Historically, the study of spatial correlation (or spatial autocorrelation, as we called it in Chapter 2) was the first topic addressed by spatial statistics (see Whittle, 1954; Besag; 1974; Cliff and Ord, 1981). It led to the specification of a first spatial independence test that is still very popular in the spatial econometric literature. This is the so-called Moran's I spatial autocorrelation test named after the statistician who introduced it in the 1950s (Moran, 1950). Moran's I has several disadvantages. First of all, it is not a correlation coefficient in that it does not range between –1 and 1. Secondly, it is not a proper statistical test because, although based on the implicit null hypothesis of spatial independence, it does not consider an explicit alternative hypothesis. Other critiques to the Moran's I may be found in Arbia (1989). On the other hand Moran's I does represent the simplest and most commonly used test statistics in the spatial econometric literature. Furthermore, the fact it does not require the specification of an alterna-

tive hypothesis may turn out to be an advantage when initially approaching problems in an exploratory way, before specifying a definite random field model as an alternative to the basic assumptions. Finally, Burridge (1980) proved that Moran's *I* test is asymptotically equivalent to a Lagrange multiplier test if the alternative to the null hypothesis of spatial independence is expressed by a SAR or a SMA random field model.

For these reasons, we will devote the next section to a discussion of this test.

4.3.2 A *General-Purpose* Testing Procedure for Spatial Independence

While studying spatial autocorrelation Moran (1950) and Cliff and Ord (1972) proposed a test that can be applied to the study of the dependence pattern of a linear regression model's residuals at least in an exploratory analysis. The test assumes the form:

$$I = h\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\hat{e}_i\hat{e}_j}{\sum_{i=1}^{n}\hat{e}_i^2} \tag{4.1}$$

with n the number of spatial observations, $\hat{e}_i = y_i - \hat{\boldsymbol{\beta}}^T\mathbf{x}_i$ the OLS regression residuals, $w_{ij} \in \mathbf{W}$ and $\mathbf{W}$ a properly defined n-by-n contiguity matrix, $\mathbf{x}_i$ a k-by-1 data vector, $\hat{\boldsymbol{\beta}}$ the OLS estimates of the (k-by-1) parameters' vector $\boldsymbol{\beta}$, and h a normalizing factor such that $h = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$.

In matrix notation, Equation (4.1) becomes

$$I = h\left(\hat{\mathbf{e}}^T\hat{\mathbf{e}}\right)^{-1}(\hat{\mathbf{e}}^T\mathbf{W}\hat{\mathbf{e}}) \tag{4.2}$$

with $\hat{\mathbf{e}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\mathbf{X}$ an n-by-k data matrix.

Equivalently, by referring to the standardized connectivity matrix $\mathbf{W}^*$ (see Section 2.2.1.2), the test statistic simplifies as:

$$I = \left(\hat{\mathbf{e}}^T\hat{\mathbf{e}}\right)^{-1}(\hat{\mathbf{e}}^T\mathbf{W}^*\hat{\mathbf{e}}) \tag{4.3}$$

Let us first of all note that the most celebrated analogous time-series test for temporal autocorrelation, the Durbin-Watson test, can be expressed in a like manner. It is sufficient to show this point to define an appropriate connectivity matrix which accounts for the temporal dependence pattern.

In particular, if we substitute the matrix $\mathbf{W}^* = \mathbf{L}\mathbf{L}^T$ in Equation (4.3) with $\mathbf{L}$ such that:

$$\mathbf{L} = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & . & . & 0 \\ 0 & 1 & -1 & 0 & 0 & . & . & 0 \\ 0 & 0 & 1 & -1 & 0 & . & . & 0 \\ . & . & & . & & & & \\ & & & & & & & . \\ & & & & & & . & \\ 0 & 0 & 0 & 0 & 0 & . & 1 & -1 \end{pmatrix} \tag{4.4}$$

and if we, further, replace the spatial index i (i = 1, 2, …….n) with a time index t (t = 1, 2,…, T), and, finally, set h = 1, Equation (4.2) becomes:

$$I = \frac{\sum_{t=2}^{T} (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=2}^{T} \hat{e}_t^2} \tag{4.5}$$

which is the familiar expression of the Durbin-Watson test (Maddala, 2001).

Cliff and Ord (1972, 1981) studied the asymptotic distribution of the statistic I computed on OLS regression residuals. If the residuals are normally distributed (as postulated explicitly in Section 4.2.2) the test statistic is also asymptotically normal with expected value

$$E(I) = \frac{k-1}{n-k-1} tr(\mathbf{M}\mathbf{W}^*) \tag{4.6}$$

and variance

$$Var(I) = \frac{tr(\mathbf{M}\mathbf{W}^*\mathbf{M}\mathbf{W}^{*T}) + tr(\mathbf{M}\mathbf{W}^*)^2 + \left[tr(\mathbf{M}\mathbf{W}^*)\right]^2}{(n-k-1)(n-k+1)} - E(I)^2 \tag{4.7}$$

with the symbol $tr(.)$ denoting the trace of a matrix, $\mathbf{M}$ the idempotent projection matrix derived from the data matrix $\mathbf{X}$ in such a way that $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, and k the number of independent variables of the model including the constant term.

Where normality cannot be assumed for the residuals, Cliff and Ord (1972) have proved the normality of the test using a non-parametrical permutational argument.

4.3.3 The Respecification of the Linear Regression as a Multivariate CAR Field

4.3.3.1 Introduction

In this section we will refer to a linear regression model specified in a conditional form as described in Section 4.2.1. If we maintain the basic assumptions about the probability model (i. e. that the vector of random variables **Z** is multivariate Gaussian), we can consider the particular instance where the assumption SM1 of independence of the sample observations is violated.

In these circumstances, we need to re-specify our model as a vector Gaussian random field (see Section 2.2.2) for which we have:

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \\ \\ \mathbf{X}_n \end{pmatrix} \sim MVN \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \\ \\ \boldsymbol{\mu}_n \end{pmatrix} \begin{pmatrix} \mathbf{V}_1 & \mathbf{C}_{12} & . & . & \mathbf{C}_{1n} \\ . & \mathbf{V}_2 & & & . \\ . & & . & & . \\ . & & & . & . \\ \mathbf{C}_{n1} & . & . & . & \mathbf{V}_n \end{pmatrix} \tag{4.8}$$

where each $\boldsymbol{\mu}_i$ represents a *k*-by-1 vector of expected values at site *i*, and $\mathbf{V}_i$ and $\mathbf{C}_{ij}$ the matrices of cross-covariance and, respectively, of spatial auto-covariance and spatial cross-covariance between pairs of sites defined by:

$$\mathbf{V}_i = \begin{pmatrix} \gamma_{11}(ii) & \gamma_{12}(ii) & . & . & \gamma_{1k}(ii) \\ . & . & & & . \\ . & & . & & . \\ . & & & & . \\ \gamma_{k1}(ii) & . & . & . & \gamma_{kk}(ii) \end{pmatrix} \tag{4.9}$$

and

$$\mathbf{C}_{ij} = \begin{pmatrix} \gamma_{11}(ij) & \gamma_{12}(ij) & & \gamma_{1k}(ij) \\ & & & \\ & & & \\ \gamma_{k1}(ij) & & & \gamma_{kk}(ij) \end{pmatrix} \tag{4.10}$$

$\gamma_{kl}(ii)$ being the covariance between the random variables X_k and X_l at location *i* and $\gamma_{kl}(ij)$ the cross- covariance between X_k and X_l at locations *i* and *j*. (see Section 2.2.2).

However, as already noted, such a definition of a random field is too general and requires a very high number of parameters (all the elements in the matrices). The latter cannot be estimated on the basis of a single realization unless we introduce some restrictions on both the heterogeneity and the structure of dependence of the field.

In practice, in order to obtain an operational sampling model, it is useful to limit ourselves to one of the random fields introduced in Section 2.4 the properties of which are known. Given the continuous nature of many economic variables, and due to their simplicity, the spatial econometric literature has concentrated almost exclusively on random fields that obey the simultaneous autoregressive field. The framework presented here, however, is general enough to allow application to any other random fields in those cases when the phenomenon under study requires a different specification. In the present section, we shall redefine the basic hypothesis of the linear regression model in the case of a non-independent sampling model by making explicit reference to the auto-normal random field. In the following section we shall exploit a framework based on the simultaneous autoregressive field.

4.3.3.2 Respecification of the PM, GM and SM Hypotheses

Probability Model

A first way of redefining the linear regression model to take account of the spatial nature of data is by redefining the probability model in such a way that the vector of random variables involved is assumed to obey an autonormal field, that is:

$$\text{PM} \qquad \Phi = \left\{ f_{\mathbf{Z}_i}(\mathbf{Z}_i = \mathbf{z}_i, \boldsymbol{\theta}_i); \boldsymbol{\theta} \in \Theta; \Theta \subset \Re^{2(k+1)} \right\} \qquad \text{and} \qquad \mathrm{Z_i} \sim \mathrm{MVN}$$

From this, fundamental, assumption we can derive a series of consequences in terms of the probability model. First of all, we have the normality of the conditional distributions:

$$\text{PM1} \qquad f_{Y_i|\mathbf{X}_i, Y_j}(y_i | \mathbf{X}_i = \mathbf{x}_i, Y_j = y_j; j \in N(i); \boldsymbol{\theta}_i) \sim \mathrm{N}$$

As a consequence of PM1, we have the linearity of the expected values that (recalling the definition of an autonormal random field set out in Section 2.4.2.7) can now be expressed as:

$$\text{PM2} \qquad E(Y_i | \mathbf{X_i} = \mathbf{x_i}; Y_j = y_j; j \in N(i); \boldsymbol{\theta}) = \boldsymbol{\alpha}^{\mathbf{T}} \sum_{j=1}^{n} w_{ij} \mathbf{x}_j + \boldsymbol{\beta}^T \mathbf{x}_i + \rho \sum_{j=1}^{n} w_{ij} y_j$$

where $\boldsymbol{\alpha} \equiv (\alpha_1, \alpha_2, \ldots, \alpha_k)^T$, $\sum_{j=1}^{n} w_{ij} \mathbf{x}_j = \left(\sum_{j=1}^{n} w_{ij} x_{1j}, \ldots, \sum_{j=1}^{n} w_{ij} x_{kj} \right)^T$ and $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and ρ are the parameters to be estimated. In particular, $\boldsymbol{\alpha}$ and ρ are the parameters that

regulate the amount of spatial dependence in the independent variables and in the lagged dependent variable respectively. A second consequence of PM1 is the constancy of the conditional variance (homoskedasticity) i. e.:

PM3 $\quad \mathrm{Var}(Y_i | \mathbf{X}_i = \mathbf{x}_i; Y_j = y_j; j \in N(i), \boldsymbol{\theta}) = \sigma^2 \qquad \forall \mathbf{x}_i\, y_i$

We also derive the constancy of all parameters with respect to space, i. e.:

PM4 $\quad \boldsymbol{\theta}_i (\boldsymbol{\alpha}^{\mathbf{T}}, \boldsymbol{\beta}^{\mathbf{T}}, \rho, \sigma^2) = \boldsymbol{\theta} \qquad \forall i$

Statistical Model

Here we can maintain the main hypothesis that the observations of the random variable Y_i are generated by a linear combination of a systematic component and a non-systematic component

$$Y_i = \mu_i + u_i \tag{4.11}$$

The systematic component μ_i is constituted by the conditional expectation of the variable Y in location *i*, conditional upon the variable Y in the surrounding locations and on the variables **X** recorded at the same location *i* and in the surrounding locations (see Section 2.4.2.8):

$$\mu_i = E(Y_i \left| \mathbf{X_i} = \mathbf{x_i}; Y_j = y_j; j \in N(i); \boldsymbol{\theta} \right.)$$

From the linearity of the conditional expectation derived from PM2 and substituting into Equation (4.11) we obtain

GM1 $\quad Y_i = \boldsymbol{\alpha}^{\mathbf{T}} \sum_{j=1}^{n} w_{ij} \mathbf{x}_j + \boldsymbol{\beta}^T \mathbf{x}_i + \rho \sum_{j=1}^{n} w_{ij} y_j + u_i$

From the normality postulated in the PM we also have that the parameters of interest are

GM2 $\quad \boldsymbol{\theta} \equiv (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T, \rho, \sigma^2)$

We also keep the hypothesis of weak exogeneity of $\mathbf{X}_i$ with respect to $\boldsymbol{\theta}$**:**

GM3 $\quad \mathbf{X}_i$ is weakly exogenous with respect to $\boldsymbol{\theta}$

The model considered here imposes some restrictions on the parameters connected with the symmetry of the variance-covariance matrix. In particular, we have:

GM4 The parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\alpha}^{\mathbf{T}}, \boldsymbol{\beta}^{\mathbf{T}}, \rho, \sigma^2)$ are subject to the restrictions:
$\alpha_l \sigma_i^2 w_{ij} = \alpha_l \sigma_j^2 w_{ji}$; $\beta_l \sigma_i^2 w_{ij} = \beta_l \sigma_j^2 w_{ji}$; $\rho \sigma_i^2 w_{ij} = \rho \sigma_j^2 w_{ji}$; $\forall i, j, l$

ensuring that the variance-covariance matrix of the field is symmetrical. In practical terms, due to the PM3 assumption of constant variance, this restriction only requires the choice of a symmetrical **W** matrix, a condition that is almost invariably respected in the generality of geographical applications.

Finally, in order to avoid singularity of the variance-covariance matrix, we need also to assume that $\mathbf{X} \equiv (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)^T$ is a full rank (n-by-k) matrix for all the observed values of the random variables **X**:

GM5 Rank (**X**) = k

Sampling Model

Finally, in the sampling model, we shall assume that $\mathbf{y} \equiv (y_1, y_2, ..., y_n)^T$ is a sample drawn from a stationary random field characterised by the conditional distribution

$$f_{Y_i|\mathbf{X}_i, y_j}(y_i|\mathbf{X}_i = \mathbf{x}_i, Y_j = y_j; j \in N(i); \boldsymbol{\theta}) \text{ that is:}$$

SM1 $y \equiv (y_1, y_2, \ldots, y_n)^T$ is independently drawn from

$$f_{Y_i|\mathbf{X}_i, y_j}(y_i|\mathbf{X}_i = \mathbf{x}_i, Y_j = y_j; j \in N(i); \boldsymbol{\theta})$$

The model thus specified will be referred to as the *multivariate CAR spatial linear regression model*. Notwithstanding its sound probabilistic foundations, it is not encountered in the applied spatial econometric literature. It should also be remarked that part of the problem may reside in the fact that there is no routine currently available for estimation and hypothesis testing based on this modelling framework (see Chapter 6 below).

4.3.3.3 Likelihood of a Bivariate CAR Spatial Linear Regression Model

In this section we shall limit ourselves simply to bivariate auto-normal fields $\mathbf{Z}_i = (Y_i, X_i)^T$. The extension to higher dimensional fields is more complicated in its formalism and is outlined in the next section.

If Y and X are jointly distributed as a bivariate random field we know (from Section 2.4.2.8) that we can express the conditional expected value of Y_i as:

$$E\left[Y_i \middle| X_j; Y_j; X_i\right] = {}_Y\mu_i + \sum_{i \neq j} \alpha_{ij}(X_j - {}_X\mu_j) + \beta(X_i - {}_X\mu_i) + \sum_{i \neq j} \rho_{ij}(Y_j - {}_Y\mu_j) \tag{4.12}$$

To simplify a notation that runs the risk of becoming too cumbersome, we shall express (without loosing in generality) each random variable as a deviation from its respective expected value. Equation (4.12) now becomes

$$E\left[Y_i \middle| X_j; Y_j; X_i\right] = \sum_{i \neq j} \alpha_{ij} X_j + \beta X_i + \sum_{i \neq j} \rho_{ij} Y_j \tag{4.13}$$

with $\alpha_{ij} = \alpha w_{ij}$, $\rho_{ij} = \rho w_{ij}$, α, β and ρ parameters and $w_{ij} \in \mathbf{W}$. the elements of a contiguity matrix properly defined.

This expression of the expected conditional value provides an operational form for the model's systematic component. Thus if we define the non-systematic component as:

$$u_i = Y_i - E\left[Y_i \middle| X_j; Y_j; X_i\right] = Y_i - \sum_{i \neq j} \alpha_{ij} X_j - \beta X_i - \sum_{i \neq j} \rho_{ij} Y_j \tag{4.14}$$

and we redefine the data generation statistical model as the sum of the systematic and the non-systematic components:

$$Y_i = E\left[Y_i \middle| X_j; Y_j; X_i\right] + u_i \tag{4.15}$$

we have:

$$Y_i = \alpha \sum_{i \neq j} w_{ij} X_j + \beta X_i + \rho \sum_{i \neq j} w_{ij} Y_j + u_i \tag{4.16}$$

In this way each observation of the random variable Y at location *i* is expressed as a function of the observation of the variable X in the same location (as in a standard linear regression model), but also of the spatially lagged values of the variable X and of the variable Y. In other words it is expressed as a function of the mean of the neighbouring values for both variables.

We know that a bivariate CAR field has a variance-covariance matrix given by Equation (2.66) (reported here again for convenience):

$$\mathbf{Q} = \mathbf{Q}(\alpha, \beta, \rho, \sigma_x^2, \sigma_y^2) = \left\{ \mathbf{I} * - \begin{bmatrix} \rho\mathbf{W} & \alpha\mathbf{W} + (\beta - \alpha)\mathbf{I} \\ \alpha\mathbf{W} + (\beta - \alpha)\mathbf{I} & \rho\mathbf{W} \end{bmatrix} \right\}^{-1} \mathbf{\Sigma} \tag{4.17}$$

with $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{I}_n \sigma_x^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n \sigma_y^2 \end{pmatrix}$, $\mathbf{I}_l$ are *l*-by-*l* identity matrices and σ_x^2 and σ_y^2 are the unconditional variances of X and Y respectively.

Therefore it is immediate to redefine the likelihood function of the sample as a bivariate Gaussian density function having $\mathbf{Q} = \mathbf{Q}(\alpha, \beta, \rho, \sigma_x^2, \sigma_y^2)$ as a variance-covariance matrix, i. e.:

$$L(\alpha, \beta, \rho, \sigma_x^2, \sigma_y^2; \mathbf{z}) = c(\mathbf{z}) \left|\mathbf{Q}\right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \mathbf{z}^T \mathbf{Q}^{-1} \mathbf{z} \right\} \tag{4.18}$$

and, consequently, the log-likelihood is defined as:

$$l(\alpha,\beta,\rho,\sigma_x^2,\sigma_y^2;\mathbf{z}) = c(\mathbf{z}) - \frac{1}{2}\ln|\mathbf{Q}| - \frac{1}{2}\mathbf{z}^T\mathbf{Q}^{-1}\mathbf{z} \tag{4.19}$$

This expression is highly non-linear in the parameters and can therefore only be maximized by using numerical algorithms (see e. g. LeSage, 1998) in order to obtain maximum likelihood estimators. The likelihood thus derived also constitutes the basis for building various hypothesis testing procedures, as we will show in the next section.

4.3.3.4 Hypothesis Testing in the Bivariate CAR Spatial Linear Regression Model

As we said in Section 3.4, the most popular general tests used in econometrics are based on the three asymptotically equivalent statistics deriving from the concept of likelihood, namely, the Likelihood ratio (LRT), the Wald test (WT) and the Lagrange multiplier test (LMT). Now that we have fully specified the alternative hypothesis to the null hypothesis of spatial independence in the regression model, we are in a position to apply these general procedures to this particular instance. This is the aim of the present section.

Indeed, once the model is respecified (as indicated in Section 4.3.3.1 above), the system of hypotheses can be explicitly obtained by contrasting the null hypothesis $H_0 : \alpha_0 = \rho_0 = 0$, with the alternative hypothesis $H_1 : \alpha \neq 0; \rho \neq 0$

In terms of the likelihood we have, under the null:

$$L(\beta_0,\sigma_{x,0}^2,\sigma_{y,0}^2;\mathbf{z}) = c(\mathbf{z})|\mathbf{Q}_0|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{z}^T\mathbf{Q}_0^{-1}\mathbf{z}\right\} \tag{4.20}$$

with $\mathbf{Q}_0$ the variance-covariance matrix that, in this case, is $\mathbf{Q}_0 = \mathbf{Q}(\beta_0,\sigma_{x,0}^2,\sigma_{y,0}^2)$

$$= \left\{\mathbf{I}^* - \begin{bmatrix} 0 & \boldsymbol{\beta}\mathbf{I} \\ \boldsymbol{\beta}\mathbf{I} & 0 \end{bmatrix}\right\}^{-1}\boldsymbol{\Sigma}$$, implying independence between observations.

As a consequence, the log-likelihood can be expressed as:

$$l(\beta_0,\sigma_{x,0}^2,\sigma_{y,0}^2;\mathbf{z}) = c(\mathbf{z}) - \frac{1}{2}\ln|\mathbf{Q}_0| - \frac{1}{2}\mathbf{z}^T\mathbf{Q}_0^{-1}\mathbf{z} \tag{4.21}$$

In contrast, under the alternative hypothesis of a bivariate CAR random field, the likelihood assumes the expression

$$L(\alpha,\beta,\rho,\sigma_x^2,\sigma_y^2;\mathbf{z}) = c(\mathbf{z})|\mathbf{Q}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\mathbf{z}^T\mathbf{Q}^{-1}\mathbf{z}\right\} \tag{4.22}$$

with $\mathbf{Q} = \mathbf{Q}(\alpha,\beta,\rho,\sigma_x^2,\sigma_y^2)$ provided by Equation (2.62). Consequently, under the alternative hypothesis, the log-likelihood becomes:

$$l(\alpha,\beta,\rho,\sigma_x^2,\sigma_y^2;z) = c(\mathbf{z}) - \frac{1}{2}\ln|\mathbf{Q}| - \frac{1}{2}\mathbf{z}^T\mathbf{Q}^{-1}\mathbf{z} \tag{4.23}$$

A test of spatial independence can therefore easily be derived from the likelihood ratio test criterion, by simply taking minus the log-difference between (4.20) and (4.22). This leads to:

$$LRT = -2\ln\frac{L(\boldsymbol{\theta}_0;\mathbf{z})}{L(\boldsymbol{\theta}_1;\mathbf{z})} \tag{4.24}$$

with $\boldsymbol{\theta}_0 \equiv \left(\beta_0, \sigma_{x,0}^2, \sigma_{y,0}^2\right)$, and $\boldsymbol{\theta}_1 \equiv (\alpha,\beta,\rho,\sigma_x^2,\sigma_y^2;\mathbf{z})$. Substituting (4.21) and (4.23) into (4.24) we obtain:

$$\begin{aligned} LRT &= -2\left[\ln L(\boldsymbol{\theta}_0; X) - \ln L(\boldsymbol{\theta}_1; X)\right] = \\ &2\left[l(\beta_0,\sigma_{x,0}^2,\sigma_{y,0}^2;\mathbf{z}) - l(\alpha,\beta,\rho,\sigma_x^2,\sigma_y^2;\mathbf{z})\right] = \\ &= \left[\ln|\mathbf{Q}_0| + \mathbf{z}^T\mathbf{Q}_0^{-1}\mathbf{z} - \ln|\mathbf{Q}| - \mathbf{z}^T\mathbf{Q}^{-1}\mathbf{z}\right] \end{aligned} \tag{4.25}$$

that represents the formal expression of a likelihood ratio test in the case of a bivariate Gaussian field.

An alternative way of obtaining a test statistics for the hypothesis of independence in a bivariate CAR linear regression model lies in considering the different formulations assumed by the data generation mechanism according to whether or not the random sampling hypothesis is verified. Indeed, when such a condition is respected, the data generation model can be expressed as:

$$Y_i = \beta x_i + u_i \tag{4.26}$$

Whereas, in the second case, if we postulate a bivariate auto-normal random field as an alternative, it is conversely given by:

$$Y_i = \alpha\sum_{i\neq j} w_{ij}x_j + \beta x_i + \rho\sum_{i\neq j} w_{ij}y_j + u_i \tag{4.27}$$

Consequently, a simple test for the independence hypothesis can be constructed using the general test statistic introduced in Section 3.5:

$$F = \frac{n-r}{r}\frac{\left[RSS_1 - RSS_0\right]}{RSS_0} \tag{4.28}$$

with n the number of observations, r the number of constraints under the null hypothesis, RSS_1 the residuals' sum of squares of model (4.27) and RSS_0 the residuals' sums of squares of model (4.26). Expression (4.28) has a Fisher central F-distribution with r and n-r degrees of freedom.

4.3.3.5 *Likelihood of a Multivariate CAR Spatial Linear Regression Model*

Let us now consider the more general case of a multivariate linear regression model and the extension of the derivation of the likelihood considered in Equation (4.19).

From Equation (2.77), the likelihood of the multivariate CAR model is given by:

$$L(\boldsymbol{\mu},\boldsymbol{\Omega};\mathbf{z}) = c(\mathbf{z})|\boldsymbol{\Omega}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\} \tag{4.29}$$

and, hence, the log-likelihood by

$$l(\boldsymbol{\mu},\boldsymbol{\Omega};\mathbf{z}) = c(\mathbf{z}) - \frac{1}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1}(\mathbf{z}-\boldsymbol{\mu}) \tag{4.30}$$

Equation (4.30) can be expressed differently way by assuming the following reparametrizations for $\boldsymbol{\mu}_i$, $\mathbf{C}_{ij}$ and $\mathbf{V}_i : \forall i \; : \; \boldsymbol{\mu}_i = \boldsymbol{\lambda}$ and $\mathbf{V}_i = \mathbf{V} = \text{diag}\left(\upsilon_1^2 \quad \cdots \quad \upsilon_p^2\right)$, $\forall i, j \; : \; j \neq i \Rightarrow \mathbf{C}_{ij} = \boldsymbol{\Phi} w_{ij}$, where $\boldsymbol{\Phi}$ is a *k*-by-*k* symmetric matrix and $w_{ij} \in \mathbf{W}$ is the generic element of a weights matrix.

Under these assumptions, we can write:

$$\boldsymbol{\Omega}^{-1} = \mathbf{I}_n \otimes \mathbf{V}^{-1} - \mathbf{W} \otimes \mathbf{V}^{-1}\boldsymbol{\Phi}, \tag{4.31}$$

so that the log-likelihood becomes:

$$\begin{aligned} l(\boldsymbol{\lambda},\boldsymbol{\Phi},\mathbf{V};\mathbf{z}) &= c(\mathbf{z}) + \frac{1}{2}\ln\left|\mathbf{I}_n \otimes \mathbf{V}^{-1} - \mathbf{W} \otimes \mathbf{V}^{-1}\boldsymbol{\Phi}\right| + \\ &- \frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\lambda})^T \mathbf{V}^{-1}(\mathbf{x}_i - \boldsymbol{\lambda}) + \\ &+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j \; : \; j \neq i}(\mathbf{x}_i - \boldsymbol{\lambda})^T \mathbf{V}^{-1}\boldsymbol{\Phi} w_{ij}(\mathbf{x}_j - \boldsymbol{\lambda}). \end{aligned} \tag{4.32}$$

and can be used in the estimation and hypothesis testing procedures.

4.3.4 The Respecification of the Linear Regression with SAR Residuals (the *Spatial Error Model*)

4.3.4.1 *Introduction*

In the previous Section 4.3.2 we attacked the problem of non-independence in the sampling model by considering the spatial observations as drawn from a multivariate CAR field. Let us now approach the problem following a different strategy.

In this section, we will start from the alternative formulation of the linear regression model introduced in Section 4.2.2. Following this formulation, the model can be written as:

$$\text{A1} \qquad y_i = \boldsymbol{\beta}^{\mathbf{T}} \mathbf{x}_i + u_i$$

Or, more compactly, as

$$\text{A1} \qquad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

If we use such a specification, we know (from Section 4.2.2) that the hypothesis of simple random sampling is incorporated in the error component by assuming it to be conditionally distributed as a Gaussian spatial white-noise:

$$\text{A2} \qquad f_{\mathbf{u}}(\mathbf{u}|\mathbf{X}) \sim \mathrm{N}(0, \sigma 2\mathrm{In})$$

with $\mathbf{I}_n$ an *n*-by-*n* identity matrix.

A way to express formally the violation of this unrealistic condition is to considering, as an alternative, a random field $\{\mathbf{u}\}$ characterised by a particular known distribution. In this way, the problem shifts from the direct modelling of the random field $\mathbf{Z}_i = (Y_i, \mathbf{X}_i{}^{T})^T$ to the simpler problem of postulating a plausible form of dependence for the random field $\{\mathbf{u}_i\}$. Of course, at least in principle, any of the random field models introduced in Section 2.4 could be used to achieve this aim. However, one of the most popular alternatives to the spatial white noise hypothesis adopted in the econometeric literature is that of postulating a SAR model for the non-systematic component. This formulation is referred to in the econometric literature as the "Spatial Error Model" or SEM (see Anselin and Bera, 1998; Anselin et al., 2004). It is without doubt the model most commonly used in applied spatial econometrics partly because specific software and routines are readly available for estimation purposes (see Chapter 6 below). The terminology used in most of the spatial econometric literature, however, runs the risk of creating confusion in those who are more familiar with the spatial statistical literature. In fact, the acronym SAR adopted here, is used (in accordance with the spatial statistics literature; see e. g. Whittle, 1954; Cressie, 1991) to indicate the Simultaneous AutoRegressive field model discussed in Section 2.4.3.1. In the spatial econometric literature, on the other hand, the term usually indicates the so-called Spatial AutoRegressive model (see Anselin and Bera, 1998; Anselin *et al.*, 2004; LeSage, 1999). The latter has a completely different formulation (and a less justified one, under the probabilistic point of view) with respect to the one presented here and that will be examined in Section 4.3.6 below.

4.3.4.2 Derivation of the Likelihood

If we decide to model the non-systematic component as a SAR random field, we need to redefine the linear regression model by supplementing the fundamental equation:

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + e_i \tag{4.29}$$

with the simultaneous autoregressive expression (see Section 2.4.3.1) for the error term:

$$e_i = \rho \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij} e_j + u_i \tag{4.30}$$

where u_i is a Gaussian spatial white noise, $w_{ij} \in \mathbf{W}$ and $\mathbf{W}$ a properly defined weights' matrix.

In compact matrix notation, Equations (4.29) and (4.30), become respectively

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{4.31}$$

with $\mathbf{X}$ an n-by-k matrix of observations and

$$\mathbf{e} = \rho \mathbf{W}\mathbf{e} + \mathbf{u} \tag{4.32}$$

Since $\mathbf{u}$ is a Gaussian white noise field, then $\mathbf{e}$ is also Gaussian. On these assumptions, the problem is transformed into a situation where it is necessary to make inference on the vector of unknown parameters $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}, \sigma^2, \rho)$ and, with the additional feature of normality introduced through the hypothesis on the random field $\mathbf{u}$, the violation of the random sampling hypothesis is reduced to the study of the spatial autocorrelation which is present in the non-systematic component.

We know (from Section 2.4.3.1) that a SAR process is characterised by a variance-covariance matrix:

$$\mathbf{V} = (\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Sigma} (\mathbf{I} - \mathbf{B})^{-\mathbf{T}} \tag{4.33}$$

with $\mathbf{B} \equiv \{ \rho_{ij} \}$, $\rho_{ij} = \rho\, w_{ij}$ and $\boldsymbol{\Sigma}$ a diagonal matrix of generic element $\sigma_i^2 = \mathrm{Var}(u_i)$.

We will refer to this matrix as $\mathbf{V}$ (ρ, σ_i^2) in order to make explicit reference to the parameters of inferential interest. In the case of constant variances $\sigma_i^2 = \sigma^2$, matrix $\mathbf{V}$ in Equation (4.33) becomes

$$\mathbf{V} = \mathbf{V}(\rho, \sigma^2) = \sigma^2 (\mathbf{I} - \mathbf{B})^{-1} (\mathbf{I} - \mathbf{B})^{-T} \tag{4.34}$$

From Equation (4.31) we then derive

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \tag{4.35}$$

and, since $\mathbf{e}$ is assumed to be distributed as a Gaussian SAR random field, we easily obtain the likelihood function given by:

$$L(\rho, \sigma^2, \boldsymbol{\beta}; \mathbf{e}) = c(\mathbf{e})\left|\mathbf{V}(\rho, \sigma^2)\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\mathbf{e}^T\mathbf{V}(\rho, \sigma^2)^{-1}\mathbf{e}\right\} \tag{4.36}$$

Substituting the expression $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ in this last equation we obtain:

$$L(\rho, \sigma^2, \boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) = c(y, \mathbf{X})\left|\mathbf{V}(\rho, \sigma^2)\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\mathbf{V}(\rho, \sigma^2)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} \tag{4.37}$$

and by substituting the explicit expression for the matrix $\mathbf{V}$ reported in Equation (4.34) we can write:

$$\begin{aligned} &L(\rho, \sigma^2, \boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) = \\ &= c(\mathbf{y}, \mathbf{X})\left(\sigma^2\right)^{-\frac{n}{2}}\left|(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B})^{-T}\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\left[(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B})^{-T}\right]^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} \end{aligned} \tag{4.38}$$

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

$$\begin{aligned} &l(\rho, \sigma^2, \boldsymbol{\beta}; \mathbf{y}, \mathbf{X}) = \\ &= c(\mathbf{y}, \mathbf{X}) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\ln\left|(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B})^{-T}\right| - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T\left[(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B})^{-T}\right]^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned} \tag{4.39}$$

Equation (4.39) cannot be maximized analytically due to the high degree of non-linearity in the parameters. It can, however, be maximized numerically in order to produce estimates of the parameters $\boldsymbol{\beta}$, σ^2 and ρ. For the computational aspects, the reader is referred to LeSage (1999). We must, however, remark that the computational procedures employed in the available software are all approximated in that they are based on a pseudo-likelihood version of Equation (4.30) and, more specifically, on a partial likelihood function (Pace and Salvan, 1997). The log-likelihood derived in Equation (4.39) can also be used to construct hypothesis testing procedures.

4.3.4.3 Equivalence of the Statistical Model Implied by the Bivariate CAR and the SAR Residual

It is instructive to show the relationship existing between the SAR residual model and the previously introduced bivariate CAR modelling. In order to achieve this aim, let us consider the row-standardized connectivity matrix of generic element $w_{ij}^* \in \mathbf{W}^*$ introduced in Equation (2.4) and let us re-write the SAR residual model as:

$$y_i = \beta x_i + e_i \tag{4.40}$$

and

$$e_i = \rho \sum_{\substack{j=1 \\ i \neq j}}^{n} w^*_{ij} e_j + u_i = \rho L[e_i] + u_i \tag{4.41}$$

with $L[e_i]$ the spatially lagged value of e_i (See Definition 5). From Equation (4.41), rearranging the elements, we have:

$$e_i - \rho L[e_i] = u_i \tag{4.42}$$

and, finally:

$$u_i = (1 - \rho L) e_i \tag{4.43}$$

Now let us multiply both sides of Equation (4.40) by $(1 - \rho L)$. We obtain:

$$(1 - \rho L) y_i = (1 - \rho L) \beta x_i + (1 - \rho L) e_i =$$

$$y_i = \rho L y_i + \beta x_i - \rho \beta L x_i + (1 - \rho L) e_i = \tag{4.44}$$

$$y_i = \rho L y_i + \beta x_i - \rho \beta L x_i + u_i$$

and, by writing the explicit expression of a spatial lag, this becomes:

$$y_i = -(\rho\beta) \sum_{j=1}^{n} w^*_{ij} x_j + \beta x_i + \rho \sum_{j=1}^{n} w^*_{ij} y_j + u_i \tag{4.45}$$

This expression is formally equivalent to a bivariate CAR model with the reparametrization $\alpha = -\rho\beta$. However, even if the statistical model is equivalent in the two specifications, the implications in terms of statistical inference are quite different as is patent by comparing the two likelihood functions derived in Equations (4.19) and (4.39). In particular, from a comparison it emerges that the riparametrization $\alpha = \rho\beta$ imposes a restriction in Equation (4.39) that is not present in Equation (4.19). Furthermore, in the bivariate CAR, we have two parameters referring to the unconditional variances of the random fields X and Y (σ_X^2 and σ_y^2) instead of the single parameter σ^2 appearing in the residual SAR model representing the variance of the non-systematic component **u**. Thus the results obtained in the two modelling frameworks are different in terms of point estimation, the estimates' properties and hypothesis testing procedures.

4.3.4.4 Hypothesis Testing in the Spatial Error Model

If we adopt the approach based on the residual autocorrelation modelling, it is possible to construct independence tests starting from the likelihood function set out in Equation (4.30). Indeed, if we consider that the model's non-systematic component obeys a SAR random field as postulated in this section, the test can be built up by considering the null hypothesis $\rho=0$ against the alternative hypothesis $\rho\neq 0$.

Under the alternative hypothesis, the likelihood of the model was obtained in Equation (4.38) and is given by:

$$\begin{aligned} &L(\rho,\sigma^2,\boldsymbol{\beta};\mathbf{y},\mathbf{X}) = \\ &= c(\mathbf{y},\mathbf{X})\left(\sigma^2\right)^{-\frac{n}{2}}\left|(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B})^{-T}\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\left[(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B})^{-T}\right]^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\} \end{aligned} \tag{4.46}$$

Consequently, the log-likelihood is given by:

$$\begin{aligned} &l(\rho,\sigma^2,\boldsymbol{\beta};\mathbf{y},\mathbf{X}) = \\ &= c(\mathbf{y},\mathbf{X}) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\ln\left|(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B})^{-T}\right| - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\left[(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B})^{-T}\right]^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \end{aligned} \tag{4.47}$$

In contrast, under the null hypothesis of spatial independence, the likelihood is defined as:

$$L_0(\sigma^2,\boldsymbol{\beta};\mathbf{y},\mathbf{X}) = c(y,\mathbf{X})\left(\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\} \tag{4.48}$$

and the log-likelihood is thus defined by:

$$l_0(\sigma^2,\boldsymbol{\beta};\mathbf{y},\mathbf{X}) = c(y,\mathbf{X}) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) \tag{4.49}$$

From Equations (4.47) and (4.49), a likelihood ratio test for the hypothesis of spatial independence can thus be easily obtained as:

$$LRT = -2\left[l\left(\rho,\sigma^2,\boldsymbol{\beta};\mathbf{y},\mathbf{X}\right) - l_0\left(\rho,\sigma^2,\boldsymbol{\beta};\mathbf{y},\mathbf{X}\right)\right] \tag{4.50}$$

Substituting Equations (4.47) and (4.49) into Equation (4.50) we have:

$$\begin{aligned} &LRT = \\ &= -2\left\{-\frac{n}{2}\ln\sigma^2 - \frac{1}{2}\ln\left|(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B})^{-T}\right| - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\left[(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B})^{-T}\right]^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + \right. \\ &\left. + \frac{n}{2}\ln\sigma^2 + \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\} \end{aligned} \tag{4.51}$$

And, after some algebra, one eventually obtains:

$$LRT = -\ln\left|(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B})^{-T}\right| - \frac{1}{\sigma^2}(\mathbf{y}-\mathbf{X\beta})^T\left[(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B})^{-T}\right]^{-1}(\mathbf{y}-\mathbf{X\beta})$$
$$+\frac{1}{\sigma^2}(\mathbf{y}-\mathbf{X\beta})^T(\mathbf{y}-\mathbf{X\beta}) \tag{4.52}$$

The above test is distributed asymptotically as a χ^2 with one degree of freedom.

Based on the same hypotheses considered above, Anselin (1988) derived two further tests of spatial independence using the general expressions of the Wald test and the Lagrange multiplier Test introduced in Section 3.4. The Wald test assumes the following expression:

$$WT = \hat{\rho}^2\ [a + b - c^2/n] \tag{4.53}$$

with $\hat{\rho}$ the maximum likelihood estimator of the non-systematic component's auto-regressive parameter ρ of the and with a = tr{ $[\mathbf{W}(\mathbf{I}-\hat{\rho}\ \mathbf{W})^{-1}]^T\ [\mathbf{W}(\mathbf{I}-\hat{\rho}\ \mathbf{W})^{-1}]$ }, b = tr $[\mathbf{W}(\mathbf{I}-\hat{\rho}\ \mathbf{W})^{-1}]^2$, and c = tr $[\mathbf{W}(\mathbf{I}-\hat{\rho}\ \mathbf{W})^{-1}]$.

The Lagrange multiplier test is, instead, given by:

$$\text{LMT} = \frac{\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)^T\mathbf{W}\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)}{\sqrt{J\sigma^2}} \tag{4.54}$$

with $J = tr\left[\left(\mathbf{W}+\mathbf{W}^T\right)\mathbf{W}\right]$. Like the LRT, both the WT and the LMT are asymptotically distributed as a χ^2 with one degree of freedom.

4.3.4.5 Generalized Least Squares Estimators

We have seen in the previous pages that, by following the approach based on the SAR modelling of the non-systematic component, it is possible to obtain maximum likelihood estimators by using the likelihood function derived in Equation (4.36). In this case, however, it is also possible to use the estimators based on the Generalised Least Squares (GLS) as an alternative.

First, let us remember that, in this case, the model is specified as:

$$y_i = \boldsymbol{\beta}^T\mathbf{x}_i + e_i \tag{4.55}$$

and

$$e_i = \rho\ \Sigma i{\neq}j\ wij\ ej + u_i \tag{4.56}$$

or, equivalently, in more compact matrix notation, as:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{e} \tag{4.57}$$

and

$$\mathbf{e} = \mathbf{B}\,\mathbf{e} + \mathbf{u} \tag{4.58}$$

with **u** a white noise field, $\mathbf{B} = \rho\mathbf{W}$ and $\mathbf{W} \equiv \{w_{ij}\}$ a properly defined contiguity matrix.

Therefore, by minimising the GLS equation with respect to $\boldsymbol{\beta}$:

$$gls(\boldsymbol{\beta}) = (\mathbf{y} - \boldsymbol{\beta}\mathbf{X})^T \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\beta}\mathbf{X}) \tag{4.59}$$

and using the variance-covariance matrix **V** as specified in Equation (4.34) above, we obtain:

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{X}^T (\mathbf{I} - \mathbf{B})^T (\mathbf{I} - \mathbf{B})\mathbf{X}\right]^{-1} \mathbf{X}^T (\mathbf{I} - \mathbf{B})^T (\mathbf{I} - \mathbf{B})\mathbf{y} \tag{4.60}$$

It is easy to convince ourselves that, in numerical terms, this last expression is equivalent to the OLS estimator applied to a transformation of the variables X and Y. Let us define a new pair of variables $\mathbf{X}^*$ and $\mathbf{Y}^*$ such that

$$\mathbf{X}^* = (\mathbf{I} - \mathbf{B})\mathbf{X} \tag{4.61}$$

and

$$\mathbf{Y}^* = (\mathbf{I} - \mathbf{B})\mathbf{Y} \tag{4.62}$$

From Equations (4.61) and (4.62) we can interpret the transformed variables $\mathbf{X}^*$ and $\mathbf{Y}^*$ as the original variables with the spatial effects "filtered" away (on filtering spatially correlated variables see Griffith, 2003). By deriving the OLS estimators of the vector of parameters $\hat{\boldsymbol{\beta}}$ for the transformed variables, one obtains:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{*T}\mathbf{X}^*\right)^{-1} \mathbf{X}^{*T}\mathbf{Y}^* = (\mathbf{X}^{*\mathrm{T}}\,\mathbf{X}^*)^{-1}\,\mathbf{X}^{*\mathrm{T}}\,\mathbf{Y}^* \tag{4.63}$$

and, by substituting Equations (4.61) and (4.62) into Equation (4.63) we again obtain Equation (4.60).

The fundamental problem in using of Generalised Least Squares is that they assume a previous knowledge of the parameter ρ which represents the argument of the matrix **B**. In practice, however, this condition is hardly ever realized unless we avail a preliminary estimate based on previous surveys or a pre-sample. Some of the proposals for overcoming this limitation will be reviewed in the next section.

4.3.4.6 Approximate Estimation Techniques

In this section we will review a set of techniques designed to overcome the GLS's procedure above mentioned limitations. These are two iterative techniques proposed by Hordijk (1974), Bartels (1979) and Anselin (1980) by analogy with those proposed by Cochrane and Orcutt (1949) and by Durbin (1960) in the context of time series analysis.

A first technique derives from the analogous procedure introduced by Durbin (1960) for time series autoregressive processes. It suggests a way of deriving an estimate of the unknown parameter ρ necessary to activate the GLS procedure described in the previous section. In order to introduce Durbin's spatial technique, let us write once again set out the basic formulation of the model written in matrix notation:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{e} \tag{4.64}$$

and

$$\mathbf{e} = \mathbf{Be} + \mathbf{u} \tag{4.65}$$

By pre-multiplying both sides of the Equation (4.64) by the term (**I** – **B**) we obtain:

$$(\mathbf{I} - \mathbf{B})\mathbf{y} = (\mathbf{I} - \mathbf{B})\mathbf{X\beta} + (\mathbf{I} - \mathbf{B})\mathbf{e} \tag{4.66}$$

and, since $(\mathbf{I} - \mathbf{B})\mathbf{e} = \mathbf{u}$, this reduces to:

$$(\mathbf{I} - \mathbf{B})\mathbf{y} = (\mathbf{I} - \mathbf{B})\mathbf{X\beta} + \mathbf{u} \tag{4.67}$$

From Equation (4.67), by developing the matrix products, we finally obtain:

$$\mathbf{y} = \mathbf{By} + \mathbf{X\beta} + + \mathbf{BX\beta} + \mathbf{u} \tag{4.68}$$

Written as in Equation (4.68) the model is formally equivalent to a simple linear regression model characterized by a non-systematic component **u** that is a white noise. As a consequence the parameters can be estimated with the Ordinary Least Squares subject to the constraint $\mathbf{B} = \rho\mathbf{W}$. Thus we obtain an estimate of the parameter ρ, say $\hat{\rho}$.

In a second step, we can use this estimate to filter the variables X and y as indicated in Equations (4.61) and (4.62) and we can apply the OLS to the transformed data. In this way, we obtain the GLS estimates of the parameters' vector β.

The second approximate method (known as the *Cochrane – Orcutt spatial method*) is an iterative technique proposed by Hordijk (1974) for obtaining

estimates of the parameters of a spatial regression model. The procedure develops through a number of steps.

In the first step, an a-spatial regression model is assumed obeying to the equation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{4.69}$$

and the Ordinary Least Squares estimates are obtained for the vector of parameters $\boldsymbol{\beta}$ through the familiar OLS expression:

$$\hat{\boldsymbol{\beta}}_{(1)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{4.70}$$

Correspondingly, the residuals are obtained as:

$$\mathbf{e}_{(1)} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{(1)} \tag{4.71}$$

In a second step, the residuals thus obtained are used to derive an estimate of ρ (the parameter that defines the spatial autocorrelation of the non-systematic component in the supplementary equation), say $\hat{\rho}_{(1)}$, through the equation:

$$\hat{\rho}_{(1)} = \left(\mathbf{e}_{(1)}{}^T\mathbf{e}_{(1)}\right)^{-1}\left(\mathbf{e}_{(1)}{}^T\mathbf{W}\mathbf{e}_{(1)}\right) \tag{4.72}$$

with **W** the usual contiguity matrix.

Finally, in the third step the parameter estimate $\hat{\rho}_{(1)}$ is used to obtain a new estimate of the vector of parameters **β,** say $\hat{\boldsymbol{\beta}}_{(2)}$ by applying the Generalised Least Squares technique described in Section 4.3.3.4. In particular, by applying Equation (4.60), one obtains:

$$\hat{\boldsymbol{\beta}}_{(2)} = \left[\mathbf{X}^T(\mathbf{I}-\mathbf{B})^T(\mathbf{I}-\mathbf{B})\mathbf{X})\right]^{-1}\left[\mathbf{X}^T(\mathbf{I}-\mathbf{B})\mathbf{y}\right] \tag{4.73}$$

with $\mathbf{B} = \hat{\rho}_{(1)}\mathbf{W}$.

The three steps are then iterated so that, at the *k*-th step, we have:

$$\hat{\boldsymbol{\beta}}_{(k)} = f\left[\hat{\rho}_{(k-1)}\right] \tag{4.74}$$

and

$$\hat{\rho}_{(k+1)} = g\left[\hat{\boldsymbol{\beta}}_{(k)}\right] \tag{4.75}$$

The process is iterated until convergence is achieved and we obtain:

$$\hat{\boldsymbol{\beta}}_{(k)} \approx \hat{\boldsymbol{\beta}}_{(k-1)} \tag{4.76}$$

and

$$\hat{\rho}_{(k)} \approx \hat{\rho}_{(k-1)} \tag{4.77}$$

At each step therefore, the new values of $\hat{\rho}_{(k)}$ and $\hat{\boldsymbol{\beta}}_{(k)}$ are the ones that minimize the sums of squared residuals conditional upon the previous value. It can be proved that the technique eventually converges to a local minimum (for the proof in time series analysis see Sargan, 1964 and Oberhofer and Kmenta, 1971). It should be remarked, however, that there is no guarantee that this local minimum will also be the desired absolute minimum.

4.3.5 The Re-specification of the Linear Regression by Adding a Spatial Lag (the *Spatial Lag* Model)

4.3.5.1 Introduction

The two models presented in the previous sections (i. e. the bivariate CAR model and the residual SAR) represent a formal way of tackling the problem of spatial data dependence by replacing the classical paradigm of random sampling with two alternative random field models. In this sense, the two models provide a sound probabilistic background to a spatial linear regression model.

A further alternative, particularly popular in the spatial econometric literature, is not based on any specific random field model. It consists rather of a technical expedient that seeks to account for the spatial dependence between data by adding the spatially lagged value of *y* as an extra independent variable in a manner similar to the inclusion of a serially autoregressive term in a time series context. This model is often referred to as the *spatial lag* model (e. g. in Anselin and Bera, 1998), or as the *mixed regressive spatial autoregressive* model (Anselin, 1988) or, finally, as the *spatial autoregressive* or *SAR* model (LeSage, 1999). As already observed in Section 4.3.4.1, this last definition is particularly misleading because the acronym SAR is used in spatial statistics to indicate a specific random field: the Simultaneous AutoRegressive field (see Section 2.4.3.1). Due to the popularity of this model, this section we will focus on describing this alternative specification of a spatial regression model.

4.3.5.2 Derivation of the Likelihood

For the purpose of introducing of the spatial lag model, we will refer to the alternative formulation of a linear regression model introduced in Section 4.2.2. Following this formulation the model can be written as the set of the following hypotheses:

A1 $$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + \rho \sum_{j=1}^{n} w_{ij} y_j + u_i$$

where $\boldsymbol{\beta}$ is the usual *k*-by-1 vector of regressive parameters, $\mathbf{x}_i$ the *k*-by-1 vector of explicative variables at site *i*, ρ an autoregressive parameter, $w_{ij} \in \mathbf{W}$ the elements of a (possibly row-standardized) weight matrix and **u** a Gaussian spatial random field such that:

A2 $$f_{\mathbf{u}}(\mathbf{u}|\mathbf{X}) \sim \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

with $\mathbf{I}_n$ an n-by-n identity matrix.

Unlike the case of the time series analogous specification, the presence of the spatially lagged term amongst the explicative variables induces a correlation between the error term and the lagged variable itself (see Anselin and Bera, 1998). Thus Ordinary Least Squares do not provide consistent estimators in this specification. It is important to note that this specific result (proved by Anselin, 1988) does not depend on assumption A2 and holds irrespectively of the properties of the non-systematic component.

Let us write assumption A1 in a more compact matrix notation as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho \mathbf{W}\mathbf{y} + \mathbf{u} \tag{4.78}$$

with **X** now indicating a *n*-by-*k* matrix of observations.

We will now provide a probabilistic justification of Equation (4.78), using concepts of the random fields theory, before deriving the likelihood function for this alternative spatial regression model. Equation (4.78) can be interpreted as a non-stochastic linear regression where the matrix of observations **X** is assumed to be a fixed set of numbers. As a consequence of the lack of probabilistic assumptions on the dependent variables **X**, Equation (4.78) can be interpreted as a stochastic differential equation leading to a simultaneous Autoregressive (SAR) random field (see Section 2.4.3.1) in which the additional constant term $\mathbf{X}\boldsymbol{\beta}$ appears. The introduction of this term only affects the expected value of the random field: neither its variance nor its dependence structure are altered. We can therefore exploit the results relating to the SAR field (and, more specifically, use the variance-covariance matrix defined in Equation (2.69)) once we have introduced the necessary amendments.

Formally, let us re-write our model as:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon} \tag{4.79}$$

with a non-systematic component $\boldsymbol{\varepsilon}$ defined as $\boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ and with $\mathbf{u}$ a spatial white noise field such that $\mathbf{u} \approx N(\mathbf{0}; \sigma^2 \mathbf{I}_n)$. As a consequence of the Gaussian assumption on the white noise component, we have that $\boldsymbol{\varepsilon}$ is also Gaussian, but with non-zero expected values such that $\boldsymbol{\varepsilon} \approx N(\mathbf{X}\boldsymbol{\beta}; \sigma^2 \mathbf{I})$.

Let us now isolate the variable $\mathbf{y}$ in Equation (4.79) and reformulate the model as:

$$\mathbf{y} = \left(\mathbf{I} - \rho\mathbf{W}\right)^{-1} \boldsymbol{\varepsilon} \tag{4.80}$$

It is now possible to derive the properties of the random field $\mathbf{y}$ thus generated (and, consequently, the likelihood associated with a set of empirical observations) in the following way.

To start with, the expected value of the random field $\mathbf{y}$ is given by:

$$E(\mathbf{y}) = E\left[\left(\mathbf{I} - \rho\mathbf{W}\right)^{-1} \boldsymbol{\varepsilon}\right] = \left(\mathbf{I} - \rho\mathbf{W}\right)^{-1} \mathbf{X}\boldsymbol{\beta} \tag{4.81}$$

Secondly, using Equation (2.70), the variance-covariance matrix of the random field is given by:

$$E(\mathbf{y}\mathbf{y}^T) = \mathbf{V}(\sigma^2, \rho) = \sigma^2 \left(\mathbf{I} - \rho\mathbf{W}\right)^{-1} \left(\mathbf{I} - \rho\mathbf{W}\right)^{-T} \tag{4.82}$$

From (4.81) and (4.82) we then derive the Gaussian likelihood of a sample of observations and obtain:

$$\begin{aligned} &L(\sigma^2, \rho, \boldsymbol{\beta}; \mathbf{y}) = \\ &= c(\mathbf{y}) \left|V(\sigma^2, \rho)\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\mathbf{y} - (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]^T \mathbf{V}^{-1}\left[\mathbf{y} - (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]\right\} \end{aligned} \tag{4.83}$$

and, therefore, the log-likelihood can be expressed as:

$$\begin{aligned} &l(\sigma^2, \rho, \boldsymbol{\beta}; \mathbf{y}) = \\ &= c(\mathbf{y}) - \frac{1}{2}\ln\left|\mathbf{V}(\sigma^2, \rho)\right| - \frac{1}{2}\left[\mathbf{y} - (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]^T \mathbf{V}^{-1}\left[\mathbf{y} - (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right] \end{aligned} \tag{4.84}$$

Remembering Equation (4.82), the determinant of the matrix $\mathbf{V}(\sigma^2, \rho)$ can be written as:

$$\left|\mathbf{V}(\sigma^2, \rho)\right| = \left|\sigma^2 \left(\mathbf{I} - \rho\mathbf{W}\right)^{-1} \left(\mathbf{I} - \rho\mathbf{W}\right)^{-T}\right| = \sigma^{2n} \left|\left(\mathbf{I} - \rho\mathbf{W}\right)^{-1} \left(\mathbf{I} - \rho\mathbf{W}\right)^{-T}\right|$$

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

$$\left|\mathbf{V}(\sigma^2,\rho)\right| = \sigma^{2n}\left|(\mathbf{I}-\rho\mathbf{W})\right|^{-2} \tag{4.85}$$

Let us now go back to the log-likelihood and substitute Equations (4.82) and (4.85) into Equation (4.84). We obtain:

$$l(\sigma^2,\rho,\boldsymbol{\beta};\mathbf{y}) =$$

$$= c(\mathbf{y}) - \frac{1}{2}\ln\left(\sigma^{2n}\left|\mathbf{I}-\rho\mathbf{W}\right|^{-2}\right) +$$

$$-\frac{1}{2}\left[\mathbf{y}-(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]^T\left[\sigma^2(\mathbf{I}-\rho\mathbf{W})^{-1}(\mathbf{I}-\rho\mathbf{W})^{-T}\right]^{-1}\left[\mathbf{y}-(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right] =$$

$$= c(\mathbf{y}) - \frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I}-\rho\mathbf{W}\right| +$$

$$-\frac{1}{2\sigma^2}\left[\mathbf{y}-(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]^T(\mathbf{I}-\rho\mathbf{W})^T(\mathbf{I}-\rho\mathbf{W})\left[\mathbf{y}-(\mathbf{I}-\rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}\right]$$

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

$$l(\sigma^2,\rho,\boldsymbol{\beta};\mathbf{y}) =$$

$$= c(\mathbf{y}) - \frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I}-\rho\mathbf{W}\right| - \frac{1}{2\sigma^2}\left[(\mathbf{I}-\rho\mathbf{W})\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right]^T\left[(\mathbf{I}-\rho\mathbf{W})\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right] \tag{4.86}$$

which represents the formal expression of the likelihood of a *spatial lag* linear regression model. (See Anselin, 1988; p. 63, for an alternative derivation).

4.3.5.3 Estimation

Once the log-likelihood of the *spatial lag* linear regression model has been derived, we can maximize it in order to obtain maximum likelihood estimates of the parameters of interest. Unfortunately, this cannot be achieved analytically because Equation (4.86) is highly non linear in the parameters and there exists no exact solution to the maximization problem. Furthermore, a simple univariate simplex optimization algorithm cannot be used because this leaves the problem of finding an estimate of the parameter σ^2 unresolved, as remarked by LeSage (1999).

Anselin (1988) proposed an approximate solution based on a pseudo-likelihood and LeSage (1999) utilizes it to derive the software procedures for its computation.

The procedure is based on the idea of the profile likelihood (PL). This is obtained by reducing the number of unknowns in the problem by replacing in the original likelihood some of the parameters that are not of direct inferential interest (or *nuisance* parameters) with consistent estimates of them (see Pace and Salvan, 1997). Obviously, a profile likelihood is not a genuine likelihood in that it is not deduced from a joint density function. It therefore does not posses all the properties of a true likelihood (for instance, the profile score does not have null expectation). It does, however, have some interesting properties similar to those of a true likelihood.

Following Anselin (1988), LeSage (1999) proposes the following procedure.

In a first step, we build up a partial model by regressing variable **y** only on the variables **X**

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{u}_0 \tag{4.87}$$

and we derive the OLS estimate of $\boldsymbol{\beta}_0$, say $\hat{\boldsymbol{\beta}}_0 = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$.

In a second step, we build up a second partial model by regressing the spatially lagged variable $\mathbf{Wy}$ only on the variables **X**

$$\mathbf{Wy} = \mathbf{X}\boldsymbol{\beta}_L + \mathbf{u}_L \tag{4.88}$$

and we derive the OLS estimate of $\boldsymbol{\beta}_L$, say $\hat{\boldsymbol{\beta}}_L = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Wy}$.

In the third step, we compute the estimated residuals of the two reduced models contained in Equations (4.87) and (4.88) as:

$$\hat{\mathbf{u}}_0 = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_0 \tag{4.89}$$

and

$$\hat{\mathbf{u}}_L = \mathbf{Wy} - \mathbf{X}\hat{\boldsymbol{\beta}}_L \tag{4.90}$$

In the fourth step, we use the estimated residuals obtained in Equations (4.89) and (4.90) to build the partial log-likelihood:

$$l(\rho;\hat{\mathbf{u}}_0,\hat{\mathbf{u}}_L) = c(\hat{\mathbf{u}}_0,\hat{\mathbf{u}}_L) - \frac{n}{2}\ln\left[\frac{1}{n}\left(\hat{\mathbf{u}}_0 - \rho\hat{\mathbf{u}}_L\right)^T\left(\hat{\mathbf{u}}_0 - \rho\hat{\mathbf{u}}_L\right)\right] + \ln\left|\mathbf{I} - \rho\mathbf{W}\right| \tag{4.91}$$

and we maximize it in order to derive an estimate of ρ, say $\hat{\rho}$.

Finally, we use $\hat{\rho}$ to obtain the final estimates of σ^2 and $\boldsymbol{\beta}$, given by:

$$\hat{\sigma}^2 = \frac{1}{n}\left(\hat{\mathbf{u}}_0 - \hat{\rho}\hat{\mathbf{u}}_L\right)^T\left(\hat{\mathbf{u}}_0 - \hat{\rho}\hat{\mathbf{u}}_L\right) \tag{4.92}$$

and

$$\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_0 - \hat{\rho}\hat{\boldsymbol{\beta}}_L) \tag{4.93}$$

respectively.

4.3.5.4 Hypothesis Testing

Once the likelihood of the model has been derived, it is immediate to define a test statistics by specifying the null and the alternative hypotheses. We already know that, under the alternative hypothesis of a linear regression with an additional spatial lag, the log-likelihood assumes the expression:

$$\begin{aligned} &l(\sigma^2, \rho, \boldsymbol{\beta}; \mathbf{y}) = \\ &= c(\mathbf{y}) - \frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I} - \rho\mathbf{W}\right| - \frac{1}{2\sigma^2}\left[(\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]^T\left[(\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right] \end{aligned} \tag{4.94}$$

Under the null hypothesis, on the other hand, we have that $H_0: \rho = 0$ and, hence, the log-likelihood can be expressed as:

$$l_0(\sigma^2, \boldsymbol{\beta}; \mathbf{y}) = c(\mathbf{y}) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]^T\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right] \tag{4.95}$$

As a consequence, the likelihood ratio test statistics is equal to:

$$LRT = -2\left[l\left(\rho, \sigma^2, \boldsymbol{\beta}; \mathbf{y}\right) - l_0\left(\sigma^2, \boldsymbol{\beta}; \mathbf{y}\right)\right] \tag{4.96}$$

By substituting Equations (4.94) and (4.95) into Equation (4.96) we have:

$$\begin{aligned} LRT = -2\Bigg\{&-\frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I} - \rho\mathbf{W}\right| - \frac{1}{2\sigma^2}\left[(\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]^T\left[(\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right] + \\ &+ \frac{n}{2}\ln\sigma + \frac{1}{2\sigma^2}\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]^T\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]\Bigg\} \end{aligned} \tag{4.97}$$

that, after some straightforward algebra, simplifies into

$$LRT = \left\{-2\ln\left|\mathbf{I} - \rho\mathbf{W}\right| + \frac{1}{\sigma^2}\left[(\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]^T\left[(\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right] - \frac{1}{\sigma^2}\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]^T\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]\right\} \tag{4.98}$$

As is known Equation (4.98) is distributed asymptotically as a χ^2 random variable with one degree of freedom and can be used to test the hypothesis of spatial dependence within the framework of the linear regression model treated in this section.

4.3.6 Anselin's General Spatial Model

When discussing possible alternatives to the specifications of the linear regression model presented in the two preceding sections the one suggested by Anselin (1988) (and christened by him as the *general spatial model*) deserves a special mention. Following this specification, the linear regression model can be written as:

$$\text{A1} \qquad y_i = \rho \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij}^{(1)} y_j + \boldsymbol{\beta}^T \mathbf{x}_i + e_i \tag{4.99}$$

where $\boldsymbol{\beta}$ is the usual k-by-1 vector of regressive parameters, $\mathbf{x}_i$ the k-by-1 vector of explicative variables at site i, ρ an autoregressive parameter, $w^{(1)}{}_{ij} \in \mathbf{W}^{(1)}$ the elements of a (possibly row-standardized) weights' matrix and $\mathbf{e}$ is an autoregressive spatial random field such that:

$$\text{A1} \qquad e_i = \lambda \sum_{\substack{j=1 \\ i \neq j}}^{n} w_{ij}^{(2)} e_j + u_i \tag{4.100}$$

defined in terms of the weights' matrix $w_{ij}^{(2)} \in \mathbf{W}^{(2)}$, the autoregressive parameter λ and the Gaussian spatial white noise field:

$$\text{A2} \qquad f_{\mathbf{u}}(\mathbf{u}|\mathbf{X}) \sim \text{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$$

with $\mathbf{I}_n$ an n-by-n identity matrix.

It is important to note that the two weights' matrices $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ do not necessarily have to be the same. Indeed it can be proved that the spatial autoregressive parameters of the model cannot be identified when they are exactly the same. To show this, let us express the A1 assumptions (contained in Equations (4.99) and (4.100)) in a compact matrix form as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho \mathbf{W}^{(1)} \mathbf{y} + \mathbf{e} \tag{4.101}$$

and

$$\mathbf{e} = \lambda \mathbf{W}^{(2)} \mathbf{e} + \mathbf{u} \tag{4.102}$$

Rearranging Equations (4.101) and (4.102), we obtain:

$$\mathbf{y} = \mathbf{X\beta} + \rho\mathbf{W}^{(1)}\mathbf{y} + \lambda\mathbf{W}^{(2)}\mathbf{e} + \mathbf{u} \tag{4.103}$$

and, substituting $\mathbf{e} = \mathbf{y} - \mathbf{X\beta} + \rho\mathbf{W}^{(1)}\mathbf{y}$ in Equation (4.101), we have:

$$\begin{aligned}\mathbf{y} &= \mathbf{X\beta} + \rho\mathbf{W}^{(1)}\mathbf{y} + \lambda\mathbf{W}^{(2)}\left[\mathbf{y} - \mathbf{X\beta} - \rho\mathbf{W}^{(1)}\mathbf{y}\right] + \mathbf{u} = \\ &= \mathbf{X\beta} + \rho\mathbf{W}^{(1)}\mathbf{y} + \lambda\mathbf{W}^{(2)}\mathbf{y} - \lambda\mathbf{W}^{(2)}\mathbf{X\beta} - \lambda\rho\mathbf{W}^{(1)}\mathbf{W}^{(2)}\mathbf{y} + \mathbf{u}\end{aligned} \tag{4.104}$$

Two remarkable instances can be analysed. The first regards the case where $\mathbf{W}^{(1)}$ and $\mathbf{W}^{(2)}$ do not have any element in common (e. g. when they refer to different orders of contiguity). In such a case, we have that $\mathbf{W}^{(1)}\ \mathbf{W}^{(2)} = \mathbf{0}$. Equation (4.104) then becomes:

$$\mathbf{y} = \mathbf{X\beta} + \rho\mathbf{W}^{(1)}\mathbf{y} + \lambda\mathbf{W}^{(2)}\mathbf{y} - \lambda\mathbf{W}^{(2)}\mathbf{X\beta} + \mathbf{u} \tag{4.105}$$

and reduces to a biparametric spatial lag with some additional constraints on the parameters.

Conversely, when $\mathbf{W}^{(1)} = \mathbf{W}^{(2)}$, Equation (4.104) becomes:

$$\mathbf{y} = \mathbf{X\beta} + (\rho + \lambda)\mathbf{W}\mathbf{y} - \lambda\mathbf{W}\mathbf{X\beta} - \lambda\rho\mathbf{W}^2\mathbf{y} + \mathbf{u} \tag{4.106}$$

and the parameters ρ and λ are subject to two (possibly conflicting) constraints. Thus they cannot be identified univocally (Anselin and Bera, 1998; p. 252).

Likewise the *spatial lag model*, Ordinary Least Squares do not provide consistent estimators of the parameters of the model expressed in Equations (4.101) and (4.102) because of the presence of a spatially lagged term amongst the explicatives.

In order to introduce the maximum likelihood estimators, let us now derive the likelihood function associated with the *general spatial model*. First of all, from Equation (4.101) we obtain:

$$(\mathbf{I} - \rho\mathbf{W}^{(1)})\mathbf{y} = \mathbf{X\beta} + \mathbf{e} \tag{4.107}$$

and

$$\mathbf{e} = (\mathbf{I} - \lambda\mathbf{W}^{(2)})^{-1}\mathbf{u} \tag{4.108}$$

From Equations (4.107) and (4.108) we have:

$$(\mathbf{I} - \rho\mathbf{W}^{(1)})\mathbf{y} = \mathbf{X\beta} + (\mathbf{I} - \lambda\mathbf{W}^{(2)})^{-1}\mathbf{u} \tag{4.109}$$

and, finally:

$$\mathbf{u} = (\mathbf{I} - \lambda \mathbf{W}^{(2)})\left[(\mathbf{I} - \rho \mathbf{W}^{(1)})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right] \tag{4.110}$$

We are now in a position to derive the likelihood function of the Gaussian variables **u** given by:

$$L(\sigma^2; \mathbf{u}) = c(\mathbf{u})\left|\mathbf{V}\right|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\mathbf{u}^T\mathbf{V}^{-1}\mathbf{u}\right] \tag{4.111}$$

with $\mathbf{V}$ the variance-covariance matrix of the random field **u.** Since $\mathbf{V} = \sigma^2\mathbf{I}$, we have that $\left|\mathbf{V}\right| = \sigma^{2n}\left|\mathbf{I}\right| = \sigma^{2n}$. Furthermore, the inverse of the variance-covariance matrix is $\mathbf{V}^{-1} = \sigma^{-2}\mathbf{I}$. Hence

$$L(\sigma^2; \mathbf{u}) = c(\mathbf{u})\left(\sigma^{2n}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\mathbf{u}^T\mathbf{u}\right] \tag{4.112}$$

Let us now operate a change of the variable from **u** to **y** by using the transformation expressed in Equation (4.109). The Jacobian of the transformation is:

$$J = \left|\frac{\partial \mathbf{u}}{\partial \mathbf{y}}\right| = \left|\mathbf{I} - \lambda \mathbf{W}^{(2)}\right|\left|\mathbf{I} - \rho \mathbf{W}^{(1)}\right| \tag{4.113}$$

Hence, expressing Equation (4.112) as a function of **y** and multiplying by the Jacobian obtained in (4.113), we obtain the likelihood of **y** as:

$$\begin{aligned} &L(\rho, \lambda, \sigma^2, \boldsymbol{\beta}; \mathbf{y}) = c(\mathbf{y})\left(\sigma^{2n}\right)^{-\frac{1}{2}}\left|\mathbf{I} - \lambda \mathbf{W}^{(2)}\right|\left|\mathbf{I} - \rho \mathbf{W}^{(1)}\right| \\ &\exp\left[-\frac{1}{2\sigma^2}\left\{(\mathbf{I} - \lambda \mathbf{W}^{(2)})\left[(\mathbf{I} - \rho \mathbf{W}^{(1)})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]\right\}^T\left\{(\mathbf{I} - \lambda \mathbf{W}^{(2)})\left[(\mathbf{I} - \rho \mathbf{W}^{(1)})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]\right\}\right] \end{aligned} \tag{4.114}$$

and, consequently, the log-likelihood as:

$$\begin{aligned} &l(\rho, \lambda, \sigma^2, \boldsymbol{\beta}; \mathbf{y}) = c(\mathbf{y}) - \frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I} - \lambda \mathbf{W}^{(2)}\right| + \ln\left|\mathbf{I} - \rho \mathbf{W}^{(1)}\right| \\ &-\frac{1}{2\sigma^2}\left\{(\mathbf{I} - \lambda \mathbf{W}^{(2)})\left[(\mathbf{I} - \rho \mathbf{W}^{(1)})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]\right\}^T\left\{(\mathbf{I} - \lambda \mathbf{W}^{(2)})\left[(\mathbf{I} - \rho \mathbf{W}^{(1)})\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]\right\} \end{aligned} \tag{4.115}$$

As in the case of the *spatial lag model* considered in Section 4.3.5, Equation (4.115) cannot be maximized analytically for estimation purposes, due to the high degree of non linearity in the parameters contained in the expression. As a consequence, once again we have to rely on a profile likelihood approach obtained

by replacing some of the parameters (considered, at least initially, as *nuisance parameters*) with consistent estimates of them.

A particular way of proceeding is described in LeSage (1999) and consists in estimating the parameter's vector $\boldsymbol{\beta}$ with the weighted least squares estimator of the reduced model:

$$\left(\mathbf{I} - \lambda \mathbf{W}^{(2)}\right)\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{4.116}$$

using $\boldsymbol{\Omega} = \left(\mathbf{I} - \rho \mathbf{W}^{(1)}\right)^T \left(\mathbf{I} - \rho \mathbf{W}^{(1)}\right)$ as the weights' matrix, thus obtaining:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \boldsymbol{\Omega} \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{\Omega} \left(\mathbf{I} - \lambda \mathbf{W}^{(2)}\right)\mathbf{y} \tag{4.117}$$

Secondly, we define the estimated residuals of model (4.116) as:

$$\hat{\boldsymbol{\varepsilon}} = \left(\mathbf{I} - \lambda \mathbf{W}^{(2)}\right)\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \tag{4.118}$$

and use them to produce a consistent estimator of the variance of the white noise field given by:

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{n} \tag{4.119}$$

Substituting Equations (4.117) and (4.119) into Equation (4.115) one obtains:

$$\begin{aligned} l(\rho, \lambda, \mathbf{y}) &= c(\mathbf{y}) + \ln\left|\mathbf{I} - \lambda \mathbf{W}^{(2)}\right| + \ln\left|\mathbf{I} - \rho \mathbf{W}^{(1)}\right| + \\ &- \frac{1}{2\hat{\sigma}^2} \left\{\left(\mathbf{I} - \lambda \mathbf{W}^{(2)}\right)\left[\left(\mathbf{I} - \rho \mathbf{W}^{(1)}\right)\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right]\right\}^T \left\{\left(\mathbf{I} - \lambda \mathbf{W}^{(2)}\right)\left[\left(\mathbf{I} - \rho \mathbf{W}^{(1)}\right)\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right]\right\} \end{aligned} \tag{4.120}$$

that can be maximized numerically to obtain a maximum pseudo-likelihood estimate of the parameters λ and ρ, say $\hat{\lambda}$ and $\hat{\rho}$.

Now substituting $\hat{\lambda}$ and $\hat{\rho}$ into Equations (4.117), (4.118) and (4.119), we obtain the estimates for the parameters σ^2 and $\boldsymbol{\beta}$.

As far as the hypothesis testing is concerned we can again exploit the likelihood ratio approach to test the hypothesis of spatial independence in the *general spatial model*'s framework by specifying the null as $H_0: \ \lambda = \rho = 0$ against the alternative $H_1: \ \lambda \neq 0; \rho \neq 0$.

Using this approach, under the null the likelihood becomes:

$$l_0(\sigma^2, \boldsymbol{\beta}; \mathbf{y}) = c(\mathbf{y}) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right]^T \left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right] \tag{4.121}$$

and, by using Equation (4.115) and Equation (4.121), the likelihood ratio test statistics is equal to:

$$\begin{aligned} LRT &= -2\left[l\left(\rho,\sigma^2,\boldsymbol{\beta};\mathbf{y}\right)-l_0\left(\sigma^2,\boldsymbol{\beta};\mathbf{y}\right)\right] \\ &= -2\ln\left|\mathbf{I}-\lambda\mathbf{W}^{(2)}\right|-2\ln\left|\mathbf{I}-\rho\mathbf{W}^{(1)}\right|-\frac{1}{\sigma^2}\left[\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right]^T\left[\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right]+ \\ &+\frac{1}{\sigma^2}\left\{(\mathbf{I}-\lambda\mathbf{W}^{(2)})\left[(\mathbf{I}-\rho\mathbf{W}^{(1)})\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right]\right\}^T\left\{(\mathbf{I}-\lambda\mathbf{W}^{(2)})\left[(\mathbf{I}-\rho\mathbf{W}^{(1)})\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right]\right\} \end{aligned} \tag{4.122}$$

a quantity that, as is known, is asymptotically distributed as a χ^2 with two degree of freedom.

4.4 Violation of the Hypotheses on the Probability Model

4.4.1 Introduction

This section will focus on the violations relating to the probability model that may arise in the basic regression model presented in Section 4.2.

As we stated in Section 4.2.1, the fundamental assumption concerning the probability model is that of normality, all other hypotheses following as mere consequences. This hypothesis can be expressed either in the form of multinormality of the joint distribution (if we use the conditional specification of the linear regression model as in Section 4.2.1) or, alternatively, in the form of the normality of the conditional distribution of the non-systematic component given the set of observed values of the independent variables (if we adopt the standard textbook specification as in Section 4.2.2). Given its importance, it is sensible to analyse the violations of the probability model hypotheses starting from this basic assumption (Section 4.4.2). The two remaining sections of this chapter (4.4.3 and 4.4.4) will deal with the violation of the homoskedasticity hypothesis and the violation of the hypothesis of parameters spatial invariance, respectively.

4.4.2 Normality

4.4.2.1 Generalities

The first important point to note is that, if we relax the normality hypothesis, whilst retaining those of the conditional mean's linearity and of homoskedasticity, the consequences are not dramatic in terms of the model's specification. In fact, the major drawback is that, in this case, the maximum likelihood method and the related testing procedures can only be applied if we are able to fully specify an alternative probability model. It should be remarked that, in the spatial case, this

appears to be the only viable alternative to the maximum likelihood estimators since the OLS estimators cannot be exploited (as is usual in non-spatial regressions) because of the presence of the spatially lagged terms amongst the explicative variables. This section will first consider the problem of testing for normality in a spatial regression (Section 4.2.2.2) and then moves on to consider possible solutions to the violation of this hypothesis (Section 4.2.2.3).

4.4.2.2 Testing for Departures from Normality

The normality hypothesis can be tested in different ways according to the particular specification of the spatial linear model considered. When a model is specified in a conditional form as a multivariate random field (as in Section 4.3.3.5), the normality assumption refers, in fact, to the joint distribution of all the variables involved. Conversely, when we adopt one of the alternative specifications considered in this chapter (e. g. the spatial error model, the spatial lag model or the general spatial model; see Sections from 4.3.4 to 4.3.6), the normality hypothesis refers to the conditional distribution of the non-systematic component of the model given the set of observed values of the independent variables.

Let us start by tackling the problem of testing the non-systematic component's univariate normality. We will then generalize the approach for testing multivariate normality.

In the case of a model specified as in the standard textbook specification, the usual normality tests are available. In this respect, we can distinguish between non-parametric and parametric tests depending on whether the alternative hypothesis is expressed in a parametric or in a non-parametric form. Tests of the first kind are the Kolmogorov-Smirnov test (Kolmogorov, 1933; Feller, 1948) and the Shapiro-Wilk test (Shapiro and Wilk, 1965). The most popular of the tests of the second typology is the Bera-Jarque test (Jarque and Bera, 1980; Bera and Jarque, 1982).

The Kolmogorv-Smirnov test is a general test to measure how well a set of empirical observations fit to a given probability distribution. It is based on the *Kolmogorov statistic* given by:

$$K = \sup_{-\infty<z<+\infty} \left|F_n(z) - F_0(z)\right| \tag{4.123}$$

$F_n(z)$ being the empirical cumulative distribution function based on n observations and $F_0(z)$ the theoretical cumulative distribution function under the null hypothesis. (See Kendall and Stuart, 1979 for details, and Durbin, 1973 for a discussion).

Shapiro and Wilk (1965) introduced a further non-parametric test for normality based on order-statistics. The test is based on the statistics:

$$SW = \frac{\left(\sum_{i=1}^{n} a_i z_{(i)}\right)^2}{\sum_{i=1}^{n} (z_i - \bar{z})^2} \tag{4.124}$$

$z_{(i)}$ being the *i*-th order statistics, $\bar{z}$ the sample mean of the variable Z and a_i a set of tabulated coefficients. Small values of the statistics *SW* are critical in the test procedure (Kendall and Stuart, 1979).

The Bera-Jarque test uses the Pearson family of distributions as the parametric alternative to the white noise Gaussian random field. The Pearson family of density functions (Kendall and Stuart, 1979) is characterized by the expected value, the variance and by the third and the fourth standardized moments defined as $\alpha_3 = \frac{\mu_3}{\sigma^3}$ and $\alpha_4 = \frac{\mu_4}{\sigma^4}$, respectively, $\overline{\mu}_r$ being the central unstandardized *r*-th order moment. Let us now consider the sampling analogue of α_3 and α_4, say $\hat{\alpha}_3$ and $\hat{\alpha}_4$, based on the regression residuals, say $\hat{u}_i$, and defined as:

$$\hat{\alpha}_3 = \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^3}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2\right)^3}} \tag{4.125}$$

and

$$\hat{\alpha}_4 = \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^4}{\left(\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2\right)^2} \tag{4.126}$$

It is known that, under the null hypothesis of normality, $\frac{\sqrt{n}}{6}\hat{\alpha}_3$ and $\sqrt{\frac{n}{24}}(\hat{\alpha}_4 - 3)$ are asymptotically independent and both asymptotically distributed as standardized normal distributions (Kendall and Stuart, 1979). As a consequence, their squared sum is asymptotically distributed, under the null, as a χ^2 with 2 degrees of freedom. This provides Bera-Jarque test with an operational form:

$$BJ = \left[\frac{n}{6}\hat{\alpha}_3^2 + \frac{n}{24}(\hat{\alpha}_4 - 3)^2\right] \tag{4.127}$$

In the case of a model specified as a conditional multivariate field, a test of multivariate normality should, in principle, be considered. Similar tests are suggested in the literature. For example, Mardia (1970, 1974, 1980) proposed a *k*-variate skewness and kurtosis statistics for samples drawn from a *k*-variate distribution. The author suggests using the *k*-variate skewness and kurtosis statistics defined, respectively, as:

$$\alpha_{3,k} = E\left\{\left[\mathbf{z}_i - E(\mathbf{z}_i)\right]^T \mathbf{V}^{-1}\left[\mathbf{z}_i - E(\mathbf{z}_i)\right]\right\}^3 \tag{4.128}$$

and

$$\alpha_{4,k} = E\left\{\left[\mathbf{z}_i - E(\mathbf{z}_i)\right]^T \mathbf{V}^{-1}\left[\mathbf{z}_i - E(\mathbf{z}_i)\right]\right\}^4 \tag{4.129}$$

$\mathbf{z}_i$ being the *n*-dimensional observations of the variable Z_i and $\mathbf{V}$ the spatial variance-covariance matrix.

The author also derives the result that, under the null hypothesis of multinormality, the limiting distribution of $\frac{n}{6}\alpha_{3,k}$ is a χ^2 with $\frac{k(k+1)(k+2)}{6}$ degrees of freedom, while the limiting distribution of $\frac{\sqrt{n}\left[\alpha_{4,k} - k(k+2)\right]}{\sqrt{8k(k+2)}}$ is the standard normal distribution. In the univariate case, these reduce to the usual skewness and kurtosis test (see also Mardia, 1986, for references).

The previous *k*-variate multinormality tests assume that we can dispose of more than one replication of the random variable at each location. In spatial econometrics, however, almost invariably, we have only one single observation at each location in space. Hence the methods described are not applicable and the only alternative for testing normality in a conditionally-specified regression model is to apply one of the above-mentioned tests for univariate normality to the non-systematic component of the model provided by $u_i = Y_i - E(Y_i | \mathbf{X}_i = \mathbf{x}_i)$.

4.4.2.3 Solutions to the Problem of Non-normality

When the hypothesis of normality is rejected on the basis of one of the tests presented in the previous section, only two approaches are possible. The first consists in exploiting one of the common normalising transformations so as to modify the empirical data and reduce them to normality. The second consists in postulating a different distribution for the joint behaviour of **X** and **y** (in the conditional specification) or for the conditional distribution of the disturbance (in the standard textbook specification). This second choice is rarely adopted in the literature due to its greater complexity. These two alternatives will now be considered.

Transformations to Normality

A general expression for a class of data transformations to normality is provided by the Box-Cox transform (Box and Cox, 1964), given by:

$$Z^* = \frac{Z^\delta - 1}{\delta} \tag{4.130}$$

δ being a parameter such that $0 \le \delta \le 1$.

Equation (4.130) can be particularized to the three cases of the reciprocal transformation (δ =-1), the square root transformation (δ =0,5), and the logarithmic transformation (δ = 0). In particular, the logarithmic transformation is one of the most commonly used in the econometric literature and Spanos (1986) lists three major reasons for its popularity. First of all, many random variables characterized by a positive skew (like a gamma, a log-normal or a chi-square) are reduced to normality through this transformation. Secondly, the logarithmic transformation produces a stabilizing effect on the variance that helps in solving certain problems of heteroskedasticity. Finally it has an intuitive appeal in that it quantifies concepts like elasticities and growth rates, as in the case of the β-convergence model presented in Chapter 1.

Alternative Non-normal Specifications

In some instances, the normality hypothesis is rejected and it is not possible simply to transform our data to produce normality. A typical instance occurs when we have dichotomous variables or, more generally, discrete random variables. It is certainly not the aim of this book to discuss all the possible alternative formulations of a regression model when normality cannot be accepted. Simply by way of an example we will now briefly discuss the case of the spatial autologistic model as this represents one of the most common violations of the normality hypothesis encountered in empirical spatial econometric analysis.

Let us suppose that we are considering *n* locations in space and that at each location we observe a binary dependent variable *Y* together with a vector of *k* covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ik})^T$. This is the case that occurs, for example, when we try to explain the presence or absence of an economic agent in one location or the presence of a technological innovation in one region.

In general terms, binary data can be analysed through logistic regression models (Collett, 1991; Cox and Snell, 1989). In this case, the variable Y is assumed to be distributed as a Bernoulli and the regression model is respecified in terms of the logit transform leading to the so-called logistic model (Agresti, 1990):

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \boldsymbol{\beta}^T \mathbf{x}_i \tag{4.131}$$

In Equation (4.131) we define the term p_i as the systematic component of the regression model, expressed in terms of the conditional expectation:

$$p_i = E(Y_i \mid \mathbf{x}_i) = P(Y_i = 1 \mid \mathbf{x}_i) = \frac{\exp(\alpha + \boldsymbol{\beta}^T \mathbf{x}_i)}{[1 + \exp(\alpha + \boldsymbol{\beta}^T \mathbf{x}_i)]} \tag{4.132}$$

Furthermore, the constant α represents the overall intercept and $\boldsymbol{\beta}$ is a vector of k regression parameters.

If in the SM the independent random sampling can be assumed, the estimators of the model's parameters can be derived through the maximum likelihood or by using methods based on the sufficient statistics (see Aitkin et al., 1989; Collett, 1991; McCullagh and Nelder, 1989). The standard framework of the logistic approach, however, is not applicable when the hypothesis of random sampling has to be rejected on empirical bases, as happens to be the case when dealing with spatial data.

The first attempt to extend the logistic model to spatially distributed data was made by Besag (1974) who introduced the so-called auto-logistic random field already discussed in Section 2.4.2.3. Besag's auto-logistic field, however, does not incorporate any explanatory variable.

A further extension was proposed by Arbia and Espa (1996b) when analysing archaeological data and considers both the logistic and the autologistic component. It was exploited by Alfò and Postiglione (2002) in a semi-parametric spatial context.

In this case, the statistical model assumes the form:

$$\Pr(Y_i = y_i \mid Y_j = y_j, j \in N(i); \mathbf{x}_i) = \frac{\exp\left[y_i \left(\alpha + \sum_{j \in N(i)} \rho_{ij} y_j + \boldsymbol{\beta}^T \mathbf{x}_i \right) \right]}{\left[1 + \exp\left(\alpha + \sum_{j \in N(i)} \rho_{ij} y_j + \boldsymbol{\beta}^T \mathbf{x}_i \right) \right]} \tag{4.133}$$

where, in addition to the previous notation, $\rho_{ij} = \rho w_{ij}$, ρ represents an autoregressive parameter and $w_{ij} \in \mathbf{W}$ are the elements of the usual connectivity matrix.

Equation (4.133) can be rewritten in terms of logit transformation as:

$$\text{logit}(p_i) = \ln\left(\frac{\Pr(Y_i = 1 \mid Y_j = y_j, j \in N(i); \mathbf{x}_i}{1 - \Pr(Y_i = 1 \mid Y_j = y_j, j \in N(i); \mathbf{x}_i} \right) = \alpha + \sum_{j \in N(i)} \rho_{ij} y_j + \boldsymbol{\beta}^T \mathbf{x}_i \tag{4.134}$$

The major problem in drawing statistical inference on the parameters contained in Equation (4.134) lies in the fact that, given the dependence postulated in this

model, no closed form is available for the resulting likelihood (Besag, 1974). In order to solve this problem, Besag (1975) suggested estimating model parameters via the maximum pseudo-likelihood procedure presented in Section 3.2.3.

In this case, Strauss and Ikeda (1990) demonstrated that the maximum pseudo-likelihood estimates of the vector of parameters $\theta \equiv (\alpha, \boldsymbol{\beta}, \rho)$ are formally equivalent to the maximum likelihood estimates of a logistic regression parameters where a new spatially lagged variable $L(y_i) = \sum_{j \in N(i)} \rho_{ij} y_j$, with $\rho_{ij} = \rho w_{ij}^*$, is introduced among the independent variables. This formal result is obtained by applying an iteratively reweighted least squares method (see Cressie, 1991). As a consequence, in order to estimate the parameters of model (4.134) in a logistic-autologistic context, one can simply use the logistic regression options available in the standard computer packages.

4.4.3 Spatial Heteroskedasticity

4.4.3.1 Introduction

As already observed in Section 4.2.1.1, the assumption of constancy of the conditional variance (or homoskedasticity) stems directly from the hypothesis of normality considered within the probability model. Thus one cannot simply accept one hypothesis and reject the other without considering the intrinsic links between the two. It is important to remark, preliminarily, that when dealing with spatial samples (and, in particular, with regional data), heteroskedasticity is a common phenomenon due to the nature of data collection. Obvious sources of non-constant variances are linked with the different dimensions of the various regions constituting the study area, the unequal concentration both of population and of economic activity, and the alternance of rural and urban areas.

In terms of parameter estimations classical econometric theory (Davidson and MacKinnon, 1993) suggests that in those circumstances where the PM3 hypothesis is violated, the Ordinary Least Squares estimators of the vector $\boldsymbol{\beta}$, under reasonable assumptions, maintain some desirable properties such as unbiasedness, consistency, and asymptotic normality. However, since the asymptotic variance-covariance matrix of the estimators differs from the usual one, they can be highly inefficient.

In such circumstances, the Generalised Least Squares (GLS) estimator could be derived by minimising the equation:

$$l(\boldsymbol{\beta}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{4.135}$$

where $\mathbf{V}$ is the matrix of variances and spatial auto-covariances of the random field assumed in the probability model.

If we consider, for the sake of simplicity, that the independence hypothesis can still be accepted in the sampling model, such a matrix is diagonal and can be written as:

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & 0 & & \\ 0 & \sigma_2^2 & & \\ & & 0 & \\ & & 0 & \sigma_n^2 \end{bmatrix} = \sigma^2 \boldsymbol{\Lambda} \tag{4.136}$$

From the minimisation of $l(\boldsymbol{\beta})$ we obtain the estimator:

$$\overline{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right) \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} = \left(\mathbf{X}^T \boldsymbol{\Lambda}^{-1} \mathbf{X}\right) \mathbf{X}^T \boldsymbol{\Lambda}^{-1} \mathbf{y} \tag{4.137}$$

and the problem to be solved thus becomes the proper definition of the matrix **V** or, in other words, the specification of a particular form for the non-homoskedasticity of the data generation process.

Indeed, if the matrix **V** were known *a priori*, there would be no need to use estimators other than the OLS (or, equivalently, Maximum Likelihood). In fact, in this case, the matrix of variances and spatial auto-covariances can be broken down using the spectral decomposition:

$$\boldsymbol{\Lambda}^{-1} = \mathbf{H}^T \mathbf{H} \tag{4.138}$$

If we now transform the vector **y** and the matrix **X** according to the following espression

$$\mathbf{y}^* = \mathbf{H}\mathbf{y} \tag{4.139}$$

and

$$\mathbf{X}^* = \mathbf{H}\mathbf{X} \tag{4.140}$$

the statistical model can be respecified as

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{u}^* \tag{4.141}$$

and the vector $\boldsymbol{\beta}$ estimated with the OLS, obtaining:

$$\overline{\boldsymbol{\beta}} = \left(\mathbf{X}^{*T} \mathbf{X}^*\right) \mathbf{X}^{*T} \mathbf{y}^* \tag{4.142}$$

By substituting the Equations (4.139) and (4.140) into Equation (4.142) one obtains Equation (4.137).

On the other hand, if **V** is unknown (as occurs in the majority of empirical cases in economic applications) we need to estimate it on the basis of our sample.

White (1980) proposed a way out of this problem by the suggesting that, for inferential purposes, we don't need an estimate of the matrix $\mathbf{V}$, but rather an estimate of the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$, given by the product $\mathbf{X}^T\mathbf{V}\mathbf{X}$. A consistent estimator of this quantity is provided by:

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2\mathbf{x}_n\mathbf{x}_n^T \tag{4.143}$$

and does not require the skedasticity function to be explicitly specified. This estimator is called Heteroskedasticity-Consistent Covariance Matrix Estimator (HCCME) and can be proved to be asymptotically justifiable (see Davidson and MacKinnon, 1993 for details).

4.4.3.2 Testing for Spatial Heteroskedasticity

When dealing with spatial regressions, the situation is much more complicated than the one described above. In fact, when dealing with spatial data, it is purely academic to discuss the case of the violation of the condition of homoskedasticity separately from that of the independent sampling hypothesis since both are likely to occur in practice (for the links between spatial dependence and spatial heteroskedasticity, see Kelejian and Robinson, 2004). Anselin (1987a) showed that, when the spatial independence hypothesis is violated, all traditional homoskedasticity tests are distorted in favour to the null or to the alternative hypothesis and their power is significantly reduced.

As a matter of fact, it is not conceivable to test homoskedasticity separately from spatial independence. If we relax the hypothesis of independent sampling, all the statistical tests proposed in the econometric literature concerning the PM3 hypothesis of homoskedasticity can be adapted to the case of spatial econometric models. Some of them will now be reviewed here.

Historically, the first test for heteroskedasticity was the one proposed by Goldfeld and Quandt (1965). The authors suggested starting by ordering the data according to the values of some variable that can constitute the basis of the heteroskedastic behaviour. In a spatial context, obvious choices for these ordering variables are the dimensions of the region (in terms of its surface, its population or other dimensional indicators) where the variables are observed, as well as their spatial co-ordinates. In a second step, the procedure develops by dividing the full sample into three portions of dimensions, respectively, n_1, n_2 and n_3 ($n = n_1 + n_2 + n_3$) and by estimating the regression model by using only the first and the last sets of observations. Finally, the regression residuals are computed and, based on these, the test statistic *GQ* is calculated as the ratio:

$$GQ = \frac{RSS_3}{RSS_1}\frac{n_1 - k}{n_3 - k} \tag{4.144}$$

with RSS_1 and RSS_3 denoting the residual sum of squares based on the first and the last portion of the sample, respectively. The *GQ* statistics under the null hypothesis of homoskedasticity is distributed as an $F(n_3 - k; n_1 - k)$. This distribution is exact in the case we can accept normality and holds asymptotically in the case normality has to be rejected.

A second, possible testing procedure is the White test (White, 1980). This author suggests that we can use the consistent estimator of the variance-covariance matrix of the estimates (reported in Equation (4.143)) to construct a test of departure from homoskedasticity. Although the White test is consistent with respect to a wide range of parametric alternatives, it may not be very powerful in finite samples (Davidson and MacKinnon, 1993).

A third option to test heteroskedasticity in linear regression is offered by artificial regressions. Let us introduce this approach by referring to the standard textbook specification of the linear regression model presented in Section 4.2.2 where we have:

$$y_i = \boldsymbol{\beta}^T \mathbf{x}_i + u_i$$

and

$$f_u(\mathrm{u_i} | \mathbf{X}_i) \sim \mathrm{N}(0, \sigma^2) \tag{4.145}$$

or, equivalently,

$$E(\mathrm{u_i}^2 | \mathbf{X}) = \sigma^2 \tag{4.146}$$

Let us further consider Equation (4.146) as our null hypothesis and let us contrast it with the alternative implicitly expressed as:

$$E(\mathrm{u_i}^2 | \mathbf{X}) = h\left(\alpha + \boldsymbol{\tau}^T \mathbf{Q}_i\right) \tag{4.147}$$

where $\mathbf{Q}_i$ is a vector of exogenously predetermined variables (e. g. the regional dimension or the spatial coordinates) and α and $\boldsymbol{\tau}$ unknown parameters. The hypothesis of homoskedasticity can then be tested by using the parametric hypothesis that $\boldsymbol{\tau} = 0$ through an ordinary *F*-statistic. For instance an explicit alternative hypothesis is provided by (Davidson and MacKinnon, 1993):

$$E(\mathrm{u_i}^2 | \mathbf{X}) = \alpha\left(\boldsymbol{\beta}^T \mathbf{X}_i\right)^r \tag{4.148}$$

On a similar basis, it is also possible to derive a Lagrange multiplier test under the assumption of normality by following the approach originally introduced by Godfrey (1978) and by Breusch and Pagan (1979).

In particular, Breusch and Pagan (1979) proposed a generic form of homoskedasticity expressed by the following equation

$$E(\mathrm{u_i}^2|\mathbf{X}) = \alpha_1 X_{1i} + \alpha_2 X_{2i} + ... + \alpha_k X_{ki} \tag{4.149}$$

with $\boldsymbol{\alpha}^{\mathrm{T}} \equiv (\alpha_1, \alpha_{2,...} \alpha_k)$ a set of constants, X_1 the constant term of the regression and $X_2,...,X_k$ the regressors. In the case of homoskedasticity, quite obviously, we have:

$$H_0 : \alpha_2 = ... = \alpha_k = 0 \tag{4.150}$$

so that, under H_0 we have $\sigma^2_{\mathrm{i}} = \alpha_1 =$ constant.

In these circumstances, the Breusch-Pagan test statistics can be derived using the general expression of the Lagrange multiplier test (Section 3.4). Following Anselin (1988), this could be expressed as:

$$BP = \frac{1}{2}\left(\sum_{i=1}^{n} \mathbf{x}_i f_i\right)^T \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right)\left(\sum_{i=1}^{n} \mathbf{x}_i f_i\right) \tag{4.151}$$

with $f_i = \left(\frac{\hat{u}_i}{\hat{\sigma}} - 1\right); \hat{u}_i = (y_i - \hat{\boldsymbol{\beta}}^T \mathbf{x}_i)$ and $\hat{\sigma}^2 = \sum_{i=1}^{n} \hat{u}_i^2$.

Under the null hypothesis of constant conditional variance, the test-statistic contained in Equation (4.151) is distributed as a χ^2 with *k-1* degrees of freedom. Anselin (1987b) nevertheless demonstrated that, in the presence of a positive spatial dependence in the sampling model, the Breusch and Pagan test (4.151) presents a distortion towards the alternative hypothesis of heteroskedasticity, while the White's test presented above shows a distortion in favour of the null hypothesis of homoskedasticity.

On these bases Anselin (1988) proposed to consider a joint test statistic that considers both the homoskedasticity and the independence hypothesis simultaneously. As we have already seen (Equation (4.54)), the Lagrange multiplier test for the spatial independence hypothesis assumes the expression:

$$LMT = \left[\frac{\left(\mathbf{y} - \boldsymbol{\beta}^T \mathbf{X}\right)\mathbf{W}\left(\mathbf{y} - \boldsymbol{\beta}^T \mathbf{X}\right)}{\sigma^2 \sqrt{J}}\right] \tag{4.152}$$

with $J = tr\left[\left(\mathbf{W} + \mathbf{W}^T\right)\mathbf{W}\right]$ and is distributed asymptotically as a χ^2 (1). The procedure suggested by Anselin (1987b) considers the sum of the two tests statistics contained in Equations (4.151) and (4.152) leading to the *joint spatial heteroskedasticity and independence* test expressed as:

$$SHI = \frac{1}{2}\left[\sum_{i=1}^{n} \mathbf{x}_i f_i\right]^T \left[\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T\right]^{-1} \left[\sum_{i=1}^{n} \mathbf{x}_i f_i\right] + \left[\frac{\left(\mathbf{y} - \boldsymbol{\beta}^T \mathbf{x}\right)^T \mathbf{W}\left(\mathbf{y} - \boldsymbol{\beta}^T \mathbf{x}\right)}{\sigma^2 \sqrt{J}}\right]^2 \tag{4.153}$$

If *BP* and *LMT* can be assumed to be independently distributed, then *SHI* is a sum of two independent χ^2 and, thus, is also distributed as a χ^2 with k degrees of freedom. The strategy suggested by Anselin (1988) is that, once the joint null hypotheses of no spatial independence and no heteroskedasticity are rejected, one could test the two hypotheses separately.

There are many other test statistics proposed in the literature for testing the heteroskedasticity hypothesis, including the celebrated Glejser test (Glejser, 1969) and those proposed by Szroeter (1978), Harrison and McCabe (1979), Ali and Giacotto (1984) Evans and King (1988) and Newey and Powell (1987). In principle, all of these can be adapted to test spatial independence and spatial homoskedasticity jointly.

4.4.3.3 Solution to the Problem of Spatial Heteroskedasticity

When the hypothesis of non-constant conditional variance is rejected on the basis of one of the foregoing testing procedures, the problem to be solved becomes that of finding alternatives to the regression model in order to accommodate for this situation.

Due to the links between the two problems, the standard solutions to the heteroskedasticity problem are closely related to those of non-normality already discussed. In particular, one could consider a couple of alternatives.

The first alternative consists in applying an appropriate normalising transformation of the kind discussed in Section 4.4.2.3. As previously noted, this has the side-effect of stabilising the variance. The inverse and the logarithmic transformations are examples of this kind.

A second alternative is that of postulating a different, non-normal, random field in the probability model that better captures the peculiarities of the empirical data. If we follow this strategy, the next step is to derive an explicit formulation for the conditional expected value $E(Y_i \mid \mathbf{X}_i = \mathbf{x}_i)$ and the conditional variance $Var(Y_i \mid \mathbf{X}_i = \mathbf{x}_i)$ and substitute these expressions in the PM2 and PM3 assumptions in the probability model. It must be observed, however, that the results in this direction so far have beed very limited in the literature because of the complexity of specifying multivariate random fields (already discussed in Chapter 2) and because of the problems in deriving explicit expressions for the conditional mean and variance when we move from the normality assumption.

4.4.4 Spatial Invariance of the Parameters

4.4.4.1 Testing Parameters' Spatial Invariance

An important assumption that stays behind the linear regression model is that the vector of the parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2)$ is constant with respect to the observations. In a spatial context, this implies constancy of the parameters when the location changes. Clearly, this hypothesis occurs only rarely in practical instances: typically different sites in space react in a different way to similar stimuli.

The present discussion will be limited to a particular form of violation of this assumption most likely to occur in empirical cases: i. e. a shift in the parameters due to *structural changes*. For other forms of violation of the parameter spatial invariance hypothesis see Anselin (1988).

One situation that is particularly common in the econometric analysis of spatial data is the case in which the model assumes different regimes in different regions of the space. In particular, for the sake of simplicity, we will consider the case where only one structural change takes place along a boundary line delimiting two regions of the space S, say S_1 and S_2, such that $\mathsf{S}_1 \cap \mathsf{S}_2 = \mathsf{S}$. Extensions to more than two regimes are straightforward. Let us further assume that the total number of locations considered is n and that the two regions are made up of n_1 and n_2 locations respectively so that $n_1 + n_2 = n$.

In this case, the statistical model (GM) can be specified through the equations:

$$y_i = \boldsymbol{\beta}_1^T \mathbf{X}_i + u_{1,i} \quad \mathrm{i} \in \mathsf{S}_1 \tag{4.154}$$

for the first region, and

$$y_i = \boldsymbol{\beta}_2^T \mathbf{X}_i + u_{2,i} \quad \mathrm{i} \in \mathsf{S}_2 \tag{4.155}$$

for the second. In Equations (4.154) and (4.155) the sets S are defined as $\mathsf{S}_1 = \{1,2,...,n_1\}$, and $\mathsf{S}_2 = \{n_1+1, n_1+2,...,n\}$ and the vectors of the parameters of interest as $\boldsymbol{\theta}_1 \equiv (\boldsymbol{\beta}_1, \sigma_1^2)$ and $\boldsymbol{\theta}_2 \equiv (\boldsymbol{\beta}_2, \sigma_2^2)$. In a more compact matrix notation, Equations (4.154) and (4.155) can be re-written as:

$$\mathbf{y}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{u}_1 \quad \mathrm{i} \in \mathsf{S}_1 \tag{4.156}$$

and

$$\mathbf{y}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{u}_2 \quad \mathrm{i} \in \mathsf{S}_2 \tag{4.157}$$

In this case we can postulate the following probability model:

$$\begin{pmatrix} y_1 | \mathbf{X}_1 \\ y_2 | \mathbf{X}_2 \end{pmatrix} \sim \mathrm{MVN}\left[\begin{pmatrix} \mathbf{X}_1 \boldsymbol{\beta}_1 \\ \mathbf{X}_2 \boldsymbol{\beta}_2 \end{pmatrix}; \begin{pmatrix} \sigma_1^2 \mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0}^T & \sigma_2^2 \mathbf{I}_{n_2} \end{pmatrix} \right] \tag{4.158}$$

where, $\mathbf{I}_{n_i}$ is a unitary matrix of dimension n_i - by - n_i $(i = 1,2)$, and $\mathbf{0}$ an n_1 - by - n_2 matrix of zeroes. In such a framework, the hypothesis of no structural changes can be specified by the system of hypotheses:

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$$

against

$$H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$$

In the time series analysis context Chow (1960) proposed a test based on the statistic:

$$CH = \frac{SQRT - SQR1 - SQR2}{SQR1 + SQR2} \frac{n-2k}{k} \tag{4.159}$$

where $SQRT$, $SQR1$ and $SQR2$ represent the total sum of squares, the sum of squares referring only to $i \in S_1$, and the sum of squares referring to $i \in S_2$ respectively. This quantity is distributed under H_0 as a Fisher-*F* with *k* and (*n-2k*) degrees of freedom.

Consigliere (1981) and Corsi *et al.* (1982) showed that, if the hypothesis of independence is rejected in the sampling model, the Chow test can lead to invalid results. In this situation, Anselin (1988) proposed using the following test statistics based on Wald's asymptotic test:

$$AT = \mathbf{e}_0{}^T \mathbf{V}^{-1} \mathbf{e}_0 - \mathbf{e}_1{}^T \mathbf{V}^{-1} \mathbf{e}_1 \tag{4.160}$$

where $\hat{\mathbf{e}}_0$ and $\hat{\mathbf{e}}_1$ are the maximum likelihood estimates of the residuals under the null hypothesis and under the alternative hypothesis respectively, and $\mathbf{V}$ the variance- covariance matrix of the field. The *AT* statistic is distributed as a χ^2 with *k* degrees of freedom, *k* being the dimension of the vector $\boldsymbol{\theta}$.

In particular, if we consider the approach based on autocorrelation (Section 4.3.4) and the additional hypothesis that the non-systematic component follows a SAR field, we know that the matrix $\mathbf{V}^{-1}$ assumes the form:

$$\mathbf{V}^{-1} = \hat{\sigma}^2 (\mathbf{I} - \mathbf{B})^T (\mathbf{I} - \mathbf{B}) \tag{4.161}$$

(see Section 2.4.3.1) and, consequently, the test statistic can be written as:

$$\frac{\left[\mathbf{e}_0^T (\mathbf{I} - \mathbf{B})^T (\mathbf{I} - \mathbf{B}) \mathbf{e}_0\right] - \left[\mathbf{e}_1^T (\mathbf{I} - \mathbf{B})^T (\mathbf{I} - \mathbf{B}) \mathbf{e}_1\right]}{\hat{\sigma}^2} \tag{4.162}$$

with $\hat{\sigma}^2$ the maximum likelihood estimate of the variance of field **u**.

4.4.4.2 Estimation in the Presence of Structural Changes

By analogy with the time series analysis, to estimate structural changes we can use the test procedure proposed by Quandt (1958).

Let us first consider the formulation presented in the previous section and let us introduce the following notation:

$$\mathbf{y}^* = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}; \mathbf{X}^* = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \boldsymbol{\beta}^* = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}; \mathbf{e}^* = \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{bmatrix} \tag{4.163}$$

With these definitions, we can write the statistical model compactly as:

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{u}^* \tag{4.164}$$

If we consider the approach based on the residual spatial autocorrelation modelling, then the model is complemented by the equation:

$$\mathbf{e}^* = \mathbf{B}^*\mathbf{e}^* + \mathbf{u} \tag{4.165}$$

with $\mathbf{u}$ a spatial white noise with its variance-covariance matrix given by:

$$\boldsymbol{\Sigma}_u = \begin{vmatrix} \sigma_1^2\mathbf{I}_{n_1} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2\mathbf{I}_{n_2} \end{vmatrix} \tag{4.166}$$

$\mathbf{B}^*$ being a connectivity matrix which can also assume a two-regime structure. For instance, we can assume

$$\mathbf{B}^* = \begin{bmatrix} \rho_1\mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & \rho_2\mathbf{W}_2 \end{bmatrix} \tag{4.167}$$

with $\mathbf{W}_1$ and $\mathbf{W}_2$ the contiguity matrices referring, to the set of locations included in S_1 e S_2 respectively.

The log-likelihood function of the model thus specified assumes the form

$$l(\sigma_1^2, \sigma_2^2, \rho_1, \rho_2, \boldsymbol{\beta}^*; \mathbf{X}) = \tag{4.168}$$

$$= -\frac{n_1}{2}\ln\sigma_1^2 - \frac{n_2}{2}\ln\sigma_2^2 + \ln\left|\mathbf{I} - \mathbf{B}^*\right| - \frac{1}{2}(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}^*)^T\left(\mathbf{I} - \mathbf{B}^*\right)^T\boldsymbol{\Sigma}_u^{-1}(\mathbf{I} - \mathbf{B}^*)(\mathbf{y}^* - \mathbf{X}^*\boldsymbol{\beta}^*)$$

Equation (4.168) can be employed to derive the maximum likelihood estimates following the usual procedures.

5 Italian and European β-convergence Models Revisited

5.1 Introduction

In Section 1.3 we presented two European regional datasets that were used informally to illustrate, on a mere intuitive basis, some of the problems that may arise in linear regression modelling of spatial economic data. The whole range of problems emerging in spatial economic modelling has now been discussed at length in the preceding chapters and a series of techniques has been introduced that define the basic elements on which a formal spatial econometric analysis may be founded. At this stage, therefore, it is useful to go back to the same set of data and re-analyse them in the light of the approach presented here. This will enable us to see the spatial econometric techniques at work and discuss, albeit briefly, some of the problems of interpretation that may arise in this context. Such is the aim of the present chapter.

It must be noted at the outset that the standard statistical and econometric packages do not contain the routines needed to implement all the methodologies described in Chapter 4. This fact strongly limits the possibility of applying spatial econometric techniques in practice and the interested researcher can only rely on the few existing softwares that are dedicated to this specific topic. As a consequence, this chapter will present examples of applications that are limited to the model estimation and hypothesis testing procedures supported by existing software. In particular, the examples given in this chapter are developed by making use of the SpaceStat package (Anselin, 1992a; 1992b). We will discuss aspects connected with the software for spatial econometric analysis more thoroughly in the Appendix.

5.2 A Spatial Econometric Analysis of the Italian Provinces β-convergence Model

5.2.1 Violation of the Hypotheses on the Sampling Model

Let us start revisiting the results of the β-convergence analysis we performed in Section 1.3.2 by examining the long-run growth dynamics for the 92 Italian provinces during the years 1950 – 1999. An analysis of the regression diagnostic led us to accept the hypotheses of normality and homoskedasticity, but we questioned the

hypothesis of independence of the OLS regression residuals on the basis of the visual inspection of Figure 1.7. Indeed, we noticed that the residuals were arranged in that map with a sort of spatial continuity and displayed a gradual decline from large positive to negative residuals. We also noted that the model had the tendency to overestimate the growth rates in some definite portions of the geographical space and to underestimate them in others, thus indicating that some explicative factor of geographical differentiation was missing from the analysis.

We are now in a position to face the problem in a more formal way and to test the null hypothesis of spatial independence. To do so, we will employ some of the tests of spatial dependence presented in Section 4.3 when discussing violations of the sampling model assumption. In particular, three different tests for spatial dependence will be considered in the present context. The first is the general-purpose Moran's *I* test which (as noted in Section 4.3.2) does not admit an explicit alternative hypothesis to contrast the null. The other two are Lagrange multiplier tests (LMT) that consider the "spatial lag" and the "spatial error" models as alternatives to the hypothesis of spatial independence. In computing these tests, and in all the subsequent analyses, we have assumed that the topology of the area can be described by a connectivity matrix built according to the simple contiguity-based neighbourhood structure (see Definition 4). Other definitions of neighbourhood were also tested and the results did not change in substance.

The results of the computations on the spatial dependence tests are displayed in Table 5.1. All tests lead to the rejection of the null hypothesis. Both the LMT (tested against the spatial error alternative) and the LMT (tested against the spatial lag alternative) present very high values and are both highly significant.

These results highlight the fact that the original model, which has been the workhorse of much previous empirical research, suffers (at least in the particular instance examined here) from a serious misspecification due to omitted spatial dependence.

It would be useful to attempt some alternative specifications in order to remove the problem of spatial dependence. Unfortunately, the available software (in our case the SpaceStat package) only allows the *spatial error* model and the *spatial lag* model as alternative specifications to the classical linear regression model. The parameters are estimated via the maximum likelihood methods using the pseudo-likelihood definition and the approximate non-linear maximization method discussed in Sections 4.3.4.2 and 4.3.5.1.

Table 5.1. Spatial dependence tests for the OLS residuals of the β-convergence in the 92 Italian provinces (1950-1999) *(figures in brackets refer to the p-values).*

Moran's I	7.226 (0.000)
LMT (spatial error model as alternative hypothesis)	45.866 (0.000)
LMT (spatial lag model as alternative hypothesis)	15.959 (0.000)

The results of the estimation and of the hypothesis testing procedures are reported in Table 5.2. All parameters in both specifications are highly significant. The table reveals a big change in the estimated value of the parameter β and, consequently, in the *speed of convergence* and *half-life* evaluation. In fact, in the classical linear model considered in Section 1.3, the speed of convergence was estimated at 0.047 (see Table 1.2) whereas now it is 0.0187 in the *spatial error* model and 0.0181 in the *spatial lag* model. This, in turn, results in a longer *half-life* time. In Table 1.2 it was estimated at 14.74 years and it is now equal to 36 years in the *spatial error*

Table 5.2. Convergence of per-capita income in the 92 Italian provinces (1950-1999)– Spatial Dependence Models – ML Estimates *(numbers in brackets refer to the p-values).*

	Spatial error model	**Spatial lag model**
α (Constant)	0.122 (0.046)	0.110 (0.003)
β	-0.608 (0.000)	-0.5955 (0.000)
Speed of convergence (*)	0.0187	00181
Half-life (**)	36.87	37.95
ρ (see Equation (4.30))	0.437 (0.000)	
ρ (see Equation (4.78))		0.265 (0.001)
Goodness of fit		
Schwartz Criterion	-40.326	-47.061
Regression Diagnostics		
1. Spatial heteroskedasticy		
Breusch-Pagan heteroskedasticity test	1.107 (0.575)	0.682 (0.711)
2. Spatial dependence		
LRT (Spatial error model vs. OLS)	10.973 (0.000)	
LMT (spatial lag model as alternative hypothesis)	0.307 (0.578)	
LRT (Spatial lag model vs. OLS)		8.759 (0.003)
LMT (spatial error model as alternative hypothesis)		2.167 (0.141)

(*) Speed of Convergence $b = -\frac{\ln(1+\beta)}{T}$ *; (**) Half-life* $= t_{half-life} = \frac{\ln(2)}{b}$.

model and to 37 years in the *spatial lag* model. It should be noted, however, that in this new specification we cannot attach to the *speed* and *half-life* parameters exactly the same meaning they had in the original Barro and Sala-i-Martin specification. On the interpretation of the parameters in different regional convergence models see Arbia *et al.* (2005).

A second feature emerging from Table 5.2 is that the new specifications do not induce problems from a heteroskedasticity point of view: the Breusch-Pagan tests are non-significative in both cases.

From a spatial dependence point of view, we can observe a general reduction of the problem. In fact, according to the Lagrange multiplier tests, the residuals no longer display a spatial pattern (the p-values are now 0.578 and 0.141, respectively). However, the Likelihood ratio test still reports significant values, thus showing that there could still be a component of spatial dependence in the empirical residuals of the two regression models.

5.2.2 Violation of the Hypotheses on the Probability Model

The regression results reported in the previous section have assumed homogeneity of variances in space and constancy of the parameters' value in the various geographical regions. In this section, we will consider the violation of the hypothesis of constant parameters and we will test whether, in the study area as a whole it is possible to distinguish differentials in the speed of convergence that may lead to the identification of spatial regimes. The issue of spatial regimes is a very important aspect and one often neglected in the empirical studies, it is although strongly emphasised in the "convergence club" literature (Quah, 1996a, 1996b; 1997; Baumont *et al.*, 2003). The idea of convergence clubs is that regions within a country or integrated area (such as the European Union) might experience not so much a global convergence process, but instead a convergence by "clubs", having the initial conditions in common. The initial conditions, in turn, can be strongly correlated to geographical (e. g. a center-periphery or a North-South dichotomy) or socio-economic peculiarities (like human capital, unemployment rate, public infrastructure, R&D activity or financial deepening). The convergence club hypothesis has an unequivocal implication in terms of the distribution of per-capita GDP: if the parameters characterising each regime are different, a threshold process should be consistent with a bimodal distribution of the per-capita incomes. It is known, from previous studies, that the data relating to the per-capita GDP of the Italian provinces do exhibit such bimodality (see Arbia *et al.* 2002, 2003; Arbia and Basile, 2005 amongst the others). This implies that the assumption of a fixed relationship between regional growth rate and initial per-capita incomes for the dataset as a whole is untenable. Rather, heterogeneity may be present, in the form of different intercepts and/or slopes in the regression equation relating to data subsets.

Empirical tests of β–convergence between Italian provinces often take the possibility of different intercepts into account by including dummy variables in the regression specification, typically, variables which indicate whether the region or the province belongs to the South or to the North-Centre (see e. g. Fabiani and Pellegrini, 1997). Generally speaking, however, they do not consider any of the spatial regime specifications discussed in Section 4.4.4.

Here we wish to test the hypothesis of the existence of two distinct spatial regimes: the first corresponding to the Centre-North area and the second to the southern part of the country. The provinces falling within to the two sub-areas are displayed in Figure 5.1. They refer to the specific policy programme (known in Italy as *Cassa del Mezzogiorno*) that provided the southern Italian regions with additional financial resources during the period considered.

In order to test the hypothesis of the existence of two different speeds of convergence in the 2 sub-areas, we considered a Chow test of spatial regimes (see Equation (4.159) in Section 4.4.4.1). The results of such test are reported in Table 5.3. The tests are highly significant as regards both the *spatial error* and the *spatial lag* specification. They lead, in both cases, to a rejection of the hypothesis of constant parameters' values.

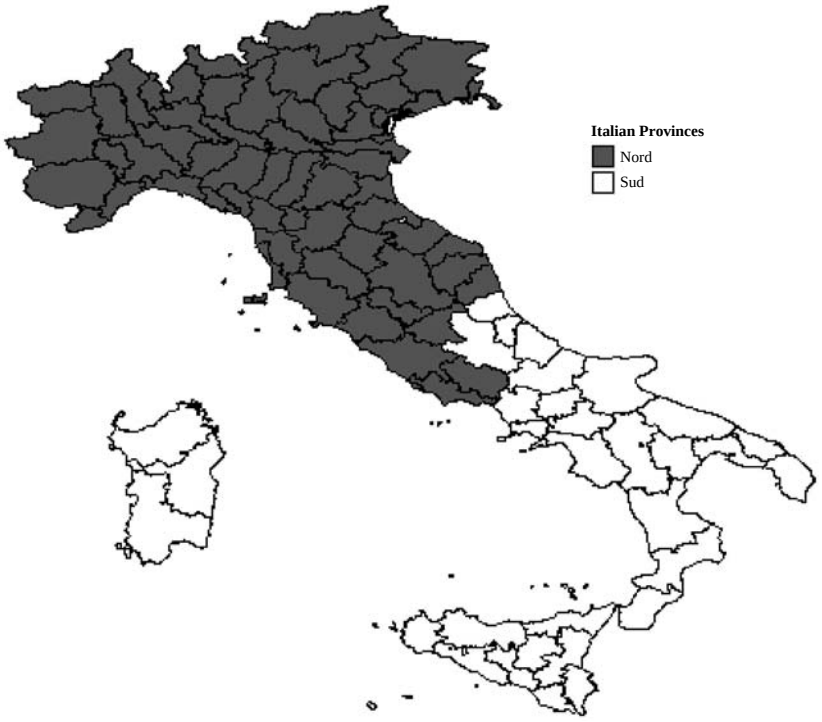


Fig. 5.1. Classification of the 92 Italian provinces within the two geographical regimes.

Table 5.3. Convergence of per-capita income in the 92 Italian provinces (1950-1999). Spatial invariance of parameters.

	Spatial error model	**Spatial lag model**
Chow test	42.623 (0.000)	71.273 (0.000)

Table 5.4. Convergence of per-capita income in the 92 Italian provinces (1950-1999). Spatial Dependence Models with spatial regimes (ML Estimates) *(numbers in brackets refer to the p-values).*

	Spatial error model	**Spatial lag model**
α (Constant) North-Centre	0.146 (0.014)	0.111 (0.003)
β North-Centre	-0.6172 (0.000)	-0.6004 (0.000)
Speed of convergence (*)	0.019248	0.01834
Half-life (**)	35.84	37.61
α (Constant) Mezzogiorno	-0.551 (0.000)	-0.563 (0.000)
β Mezzogiorno	-0.5899 (0.000)	-0.585 (0.000)
Speed of convergence (*)	0.01783	0.01759
Half-life (**)	38.69	39.22
ρ (see Equation (4.30))	0.421 (0.000)	
ρ (see Equation (4.78))		0.259 (0.002)
Goodness of fit		
Schwartz Criterion	-42.156	-51.479
Regression Diagnostics		
1. Spatial heteroskedasticy		
Breusch-Pagan heteroskedasticity test	0.977 (0.322)	0.504 (0.477)
2. Spatial dependence		
LRT (Spatial error model vs. OLS)	12.961 (0.000)	
LMT (spatial lag model as alternative hypothesis)	0.249 (0.617)	
LRT test (Spatial lag model vs. OLS)		8.159 (0.004)
LMT (spatial error model as alternative hypothesis)		2.183 (0.139)

(*) Speed of Convergence $b = -\frac{\ln(1+\beta)}{T}$; *(**) Half-life* $= t_{half-life} = \frac{\ln(2)}{b}$.

Having ascertained the existence of multiple regimes, we can now estimate the two alternative specifications of the regression model by considering two distinct regimes in the two partitions displayed in Figure 5.1. We will allow the parameters α and β to change in the two regimes, as postulated in Equation (4.164). In contrast, we will consider the spatial effects incorporated by the two alternative models in the two sub-regions as constant and we will therefore assume $\rho_1 = \rho_2$ in Equation (4.167). The results of the analysis are displayed in Table 5.4.

We can see from the table that, in the *spatial error* model, the speed of convergence (originally estimated with a value of 0.0187) is now estimated at 0.0192 in the Centre-North and at 0.0178 in the *Mezzogiorno*. Consequently, the *half-life* time (originally 36.87 years for the whole country) is now estimated at 35.84 years in the Centre-North and at 38.69 years in the South. It is therefore patent that convergence was more rapid in the Centre-North than in the South during the period considered.

The same kind of result is evident when exploring the effects of the two regimes in the *spatial lag* model. Here we see that the speed of convergence (originally estimated at 0.0181) is now estimated at 0.0183 in the Centre-North and at 0.0175 in the *Mezzogiorno*. Consequently, the *half-life* time (37.95 years for the whole country), is now estimated at 37.61 years in the Centre-North and at 39.22 years in the South.

In conclusion, the results reported in this section provide strong evidence of spatial effects in the convergence model. These effects have some important implications for the estimated convergence speed. In particular, our results clearly suggest that, in the presence of a positive spatial autocorrelation in the OLS residuals, convergences rates estimated via the traditional non-spatial regression model, may be strongly biased due to the fact that regional spill-over effects allow regions to grow faster or slower than one would expect.

5.3 A Spatial Econometric Analysis of the European Regions β-convergence Model[1]

5.3.1 Violation of the Hypotheses on the Sampling Model

Let us now reconsider the β-convergence process of the 129 NUTS-2 European regions already considered in Section 1.3.3 for the period 1980-1996. As in the case of the Italian provinces discussed above, when we considered the European NUTS-2 regions in Section 1.3.3, we noticed some indications of spatial dependence among the residuals of the OLS a-spatial regression (see Figure 1.12). In this section we will therefore try to provide more formal grounds for the visual impression reported there. The results of the spatial dependence analysis of the OLS

[1] The empirical results presented here are partly based on Postiglione *et. al.* (2002).

Table 5.5. Spatial dependence tests for the OLS residuals of the β-convergence in the 129 European NUTS-2 regions (1950 – 1999). *(Figures in brackets refer to the p-values).*

Moran's I	5.055 (0.000)
LMT (spatial error model as alternative hypothesis)	22.235 (0.000)
LMT (spatial lag model as alternative hypothesis)	14.757 (0.000)

residuals are reported in Table 5.5. Here, again, we examine the Moran's *I* test together with two versions of the Lagrange multiplier test, considering the *spatial lag* and the *spatial error* model as alternatives to the hypothesis of spatial independence. In the computation of the test (and in all subsequent analysis) we considered the spatial data topology incorporated into a simple contiguity-based definition of weights. Again, more sophisticated definitions of neighbourhood did not lead to significant differences.

As is patent from Table 5.5, all three tests are highly significant and lead to the rejection of the hypothesis of independence among the residuals. This conclusion leads us to try and eliminate the disturbing effects caused by spatial dependence using one of the spatial regression models discussed in Chapter 4.

The results of the estimation procedures are given in Table 5.6. Parameters are estimated via the maximum pseudo-likelihood procedure. All parameters are highly significant in both specifications. They lead to a convergence speed of 0.015 and 0.025, respectively, in the two models (it was 0.019 in the a-spatial model reported in Section 1.3.3) and, consequently, to a half-life time of 44 and 27 years respectively in the two models (it was 34 years in the a-spatial model). Thus the introduction of spatial effects leads to a faster speed of convergence amongst the regions.

As far as the model's diagnostics, we observe that the Breusch-Pagan test leads to accept the hypothesis of constant variances.

Regarding the problem of spatial dependence amongst residuals in the spatial lag model, both the Likelihood ratio test and the Lagrange multiplier tests lead us to reject the hypothesis of spatial independence while, in the case of the spatial error model, there is a contrast between the results of the likelihood ratio test (that leads us to reject spatial independence) and those provided by the Lagrange multiplier test (that, conversely, lead to acceptance).

The general conclusions are thus not very different from those of the previous analysis referring to the Italian provinces. When a high degree of positive spatial autocorrelation is detected amongst residuals, the convergence's speed estimated through the a-spatial linear regression are strongly biased with respect to those obtained through the spatial regressions. In this case, however, the model is not entirely satisfactory because it leaves a high degree of residual spatial dependence.

Table 5.6. Convergence of per-capita income in the 129 European NUTS-2 regions (1950-1999)– Spatial Dependence Models – Maximum Likelihood Estimates *(numbers in brackets refer to the p-values).*

	Spatial error model	**Spatial lag model**
α (Constant)	2.539 (0.000)	3.623 (0.000)
β	-0.222 (0.002)	-0.302 (0.002)
Speed of convergence (*)	0.01568	0.0251
Half-life (**)	44.20	27.60
ρ (see Equation (4.30))	0.501 (0.000)	**---**
ρ (see Equation (4.78))	**---**	0.385 (0.000)
Goodness of fit		
Schwartz Criterion	-168.613	-177.472
Regression Diagnostics		
1. Spatial heteroskedasticy		
Breusch-Pagan heteroskedasticity test	0.776 (0.378)	0.692 (0.452)
2. Spatial dependence		
LRT test (Spatial error model vs. OLS)	21.259 (0.000)	**---**
LMT (spatial lag model as alternative hypothesis)	0.074 (0.785)	**---**
LRT test (Spatial lag model vs. OLS)	**---**	14.400 (0.000)
LMT (spatial error model as alternative hypothesis)	**---**	5.598 (0.018)

(*) Speed of Convergence $b = -\frac{\ln(1+\beta)}{T}$; *(**) Half-life* $= t_{half-life} = \frac{\ln(2)}{b}$.

5.3.2 Violation of the Hypotheses on the Probability Model

The model specifications chosen in the previous section are not satisfactory in that we can observe that the spatial dependence in the residuals has not been eliminated. This provides further scope for deriving alternative specifications. In this section, we will suggest some alternatives by considering possible violations of the probability model.

In order to improve our models, we will consider the study area divided into two zones each characterised by different regimes. Let us again observe Figure 1.9 where we reported the distribution of natural logarithm of per-capita GDP in EU regions at the NUTS 2 level both in 1980 (the initial period of observation) and in 1996 (the final period of observation). The graphical analysis suggests the existence of at least two different clubs with a clear geographical distinction: the richer regions located at the core of the continent and the poorer regions located in the periphery.

In order to test the hypothesis of the existence of two regimes characterized by two different speeds of convergence, we classify the EU region "poor" if the level of per-capita GDP is below the European median and "rich" if it exceeds it. The classification can be based on the initial year 1980 (Figure 1.9a) or, alternatively, on the final year 1996 (Figure 1.9b). In this section we attempt both specifications. The results of the Chow test of parameters's invariance (see Equation (5.159) in Section 4.4.4.1) are reported in Table 5.7 for the four specifications considered. The Chow tests results are highly significant for three of the four models analyzed, the only exception being the spatial lag model that adopts the 1980 classification.

Table 5.7. Convergence of per-capita income in the 129 European NUTS-2 regions (1980-1996) – Spatial invariance of parameters.

	Spatial error model (1980 classification)	**Spatial error model (1996 classification)**	**Spatial lag model (1980 classification)**	**Spatial lag model (1996 classification)**
Chow test	5.990 (0.050)	65.175 (0.000)	1.344 (0.511)	64.475 (0.000)

Having found sufficient evidence of the existence of two regimes, we can now proceed to an estimation of the parameters and the diagnostic checking of the various models. Table 5.8 shows the results of the maximum likelihood estimation of the spatial lag and spatial error parameters for the two spatial regimes.

The parameters estimates are always significant and the estimates of β are of the expected sign (see Table 5.8). The speed of convergence ranges from 1.489% to 5.089% when employing the four different models. In agreement with the theoretical convergence hypothesis, the structural change model identifies a higher speed (5.089%) for "poor" regions and a lower speed (1.489%) for the "rich" regions.

The Breusch-Pagan tests are not significant, so heteroskedasticity is rejected. Based on the value of AIC, the spatial error model with the 1996 classification is indicated as the one that achieves the best fit.

Table 5.8. Convergence of per-capita income in the 119 European NUTS-2 regions (1950-1999). Spatial Dependence Models with spatial regimes (ML Estimates) *(numbers in brackets refer to the p-values).*

	Spatial error model (1980 classification)	**Spatial error model (1996 classification)**	**Spatial lag model (1980 classification)**	**Spatial lag model (1996 classification)**
α (Constant) Rich	2.815 (0.000)	4.380 (0.000)	2.475 (0.002)	4.313 (0.000)
β Rich	-0.213 (0.004)	-0.378 (0.000)	-0.212 (0.015)	-0.395 (0.000)
Speed of convergence (*)	0.011497	0.02968	0.011489	0.03141
Half-life (**)	46,30	23,35	46.83	22.06
α (Constant) Poor	5.015 (0.000)	5.708 (0.000)	3.046 (0.000)	4.577 (0.000)
β Poor	-0.468 (0.000)	-0.557 (0.000)	-0.280 (0.000)	-0.446 (0.000)
Speed of convergence (*)	3.944%	5.089%	2.053%	3.691%
Half-life (**)	17.57	13.62	33.76	18.77
ρ (see Equation (4.30))	0.369 (0.000)	0.205 (0.023)		
ρ (see Equation (4.78))			0.369 (0.000)	0.205 (0.023)
Goodness of fit				
Schwartz Criterion	-176.792	-219.507	-163.447	-213.801
Regression diagnostics				
1. Spatial heteroskedasticity				
Breusch-Pagan heteroskedasticity test	0.059 (0.808)	1.551 (0.213)	0.091 (0.763)	0.262 (0.609)
2. Spatial dependence				
LRT (Spatial error model vs. OLS)	23.844 (0.000)	8.294 (0.003)		
LMT (spatial lag model as alternative hypothesis)	0.115 (0.735)	0.051 (0.821)		
LRT (Spatial lag model vs. OLS)			12.499 (0.000)	4.589 (0.032)
LMT (spatial error model as alternative hypothesis)			8.276 (0.004)	0.010 (0.920)

(*) Speed of Convergence $b = -\frac{\ln(1+\beta)}{T}$; *(**) Half-life* $= t_{half-life} = \frac{\ln(2)}{b}$.

Finally, notice that the alternative two-regime specifications considered also remove the problem of residual spatial dependence in three of the specifications considered if we base our conclusions on the Lagrange multiplier test, although some contrasting results are provided by the likelihood ratio test.

6 Looking Ahead: A Review of More Advanced Topics in Spatial Econometrics[1]

6.1 Introduction

As remarked in the first chapter, the aim of this monograph is to present a statistically based introduction to the field of spatial econometrics. Consequently, in the present context, we have only considered the basic techniques and we have omitted many important, more advanced, topics.

However we do not want the reader to remain totally unaware of the many other possibilities offered by spatial econometrics. The aim of this chapter therefore is to provide a short review of some of the advances proposed in the literature that have, over the last decades, substantially improved the basic spatial econometric toolbox presented so far. It is obviously not possible to give a full and detailed account of all the developments registered in this field and, as we said, this is not the purpose of the book anyway. The aim of this chapter is simply to provide the interested reader with the necessary references to enable him to deepen his knowledge in this direction if he so wishes. Furthermore, since the rest of the book has focussed solely on regional convergence of income as an example of possible applications, we also wish to take this opportunity to present a series of emerging fields, in which spatial econometric methods are potentially useful. This to open the reader's mind to other possible applications. More thorough reviews of problems and methods can be found in Anselin *et al.* (2004), Anselin and Bera (1998), LeSage (1999), Anselin (2001b; 2002) and Florax and van de Vlist (2003).

This chapter is divided into three sections. The first section is devoted to models that offer an alternative to the basic simple regression framework considered in Chapter 4. The second is a review of diagnostic tools, alternative to those presented in the previous chapters, for testing the various regression hypotheses. Finally, the third section focuses on some estimation methods, other than to the GLS and maximum likelihood methods, that can be used in a spatial econometric context to improve the estimates accuracy.

---

[1] This chapter is written jointly with Gianfranco Piras.

6.2 Alternative Models

6.2.1 Panel Data Models

A typology of models that is potentially of great interest in spatial econometrics is that based on panel data. As is well known, this kind of data allows the contemporaneous study of both the dynamic and the individual variation of economic phenomena. Baltagi (2001) lists some of the benefits and limitations of using such data (see also Hsiao, 1986; Klevmarken, 1989; Solon, 1989). First of all they allow to control for individuals heterogeneity. Furthermore, they are more informative than pure time series or cross-sectional data, they present more variability, less collinearity among the variables and more degrees of freedom. The other side of the coin is that they have a number of drawbacks. To start with, design and data-collection problems are more complicated then in the case of pure time series or cross-sectional data. Measurement errors may also arise and may produce inferential distortions. In many instances, the time dimension is too short to allow a proper dynamic modelling due to the heavy costs associated with data collection. Finally, there are major problems associated with selectivity of the sample arising in the various forms of self-selectivity, non-response, attrition or new entry.

Notwithstanding these problems, the diffusion of panel data has been supported by the increasing data availability. Until only a few years ago, the diffusion of panel data sets was restricted to the United States, the only country in which panel data were collected on a regular basis. Nowadays, many of the European countries have their own longitudinal surveys (e. g. the Italian Survey on Household Income and Wealth run by the Bank of Italy), and the European Community Household Panel (ECHP) is a precious source of information for empirical economic studies. Spatially-referenced panel data are also common in economics. The already quoted REGIO database (see Section 1.3.1) represents an example of a spatial panel dataset that is acquiring increasing importance in regional economic studies.

Recently, there has been a wide diffusion of contributions to the statistical methods designed to analyse panel data. However, only a few papers in the literature deal with spatial panel data. The contribution made by Paul Elhorst (2001, 2003) has a particular relevance in this respect. The author exhaustively examines the specification of a series of models, developed from the classical framework of traditional panel data specification and conjugated with the typical techniques for modelling spatial dependence discussed in this book. In particular, Elhorst elaborates the specification and estimation strategies for spatial panel data models that include spatial error autocorrelation and a spatially lagged dependent variable. The author starts from the classical literature on panel data and adapts what can be learned from the econometric literature by discussing four models: the spatial fixed-effect model, the spatial random-effect model and both the fixed and random coefficient spatial error models. He also derives the relative likelihood for each model and discusses the asymptotic properties and the estimation procedures. Possible problems arising from the spatial version of these four models are also dis-

cussed in detail. Another interesting aspect is the derivation of the likelihood function of a fixed-effect dynamic panel data model extended to include spatial error autocorrelation or spatially lagged dependent variables. We will return to these aspects when discussing the estimation techniques in Section 6.4 below.

Some advances in considering prediction in panel data regression models have been made by accounting for spatial autocorrelation among states and regions. Baltagi and Li (1999) derive the best linear unbiased predictor for the random error component model with spatial correlation. They compare the performances of several predictors of a simple demand equation for cigarettes based on a panel of 46 states over the period 1963-1992. The estimators they compare in the forecasting exercise are the OLS with fixed effect (both accounting for and disregarding spatial correlation effects) and the GLS estimator for random effect (again, both ignoring and considering spatial correlation effects). The main result obtained is that it is important to take spatial correlation and heterogeneity across states into account because their consideration markedly improve performance in terms of RMSE of the forecasts. Baltagi *et al.* (2003) provide further results and extended the previous findings.

More thorough reviews of spatial panel data may be found in Anselin (2001a) and Anselin *et al.* (2004b).

6.2.2 Regional Convergence Models

As is clear from the examples in this book, growth theory and economic convergence are certainly the fields in which spatial econometrics has been applied most frequently over the last decades. Within this broad field, many different models departing substantially from the simple spatial linear regression model of the previous chapters have been proposed in the literature and necessitate appropriate spatial econometric treatment.

An approach that can be considered an important step forward as regards the neoclassical growth convergence modelling framework is the one based on the concept of club convergence developed by Durlauf and Johnson (1995) and Quah (1993a; 1993b; 1996a; 1996b). See Baumont *et al.* (2003) for a review. Such a concept can help to explain why we observe economic polarization and persistence of poverty in empirical studies. The idea is based on endogenous growth models leading to a situation of multiple steady-state equilibria like the one described in Azariadis and Drazen (1990). According to this theory, some economies may converge, but only if their initial conditions fall within the basin of attraction of the same steady-state. Galor (1996) demonstrates that such a concept of convergence is, in fact, consistent with the standard neoclassical growth models, if we allow for individual heterogeneity.

Along the same lines, Quah (1993a) introduced an original approach based on Markov chain transition matrices. Quah's approach rests on the idea that convergence is such a complicated issue that it cannot be studied just by looking at the

linear correlation between growth rates and initial GDP levels, as in standard β-convergence analysis. Rather, we have to look at the whole bivariate distribution of the two variables involved (estimated with kernel densities) and also at their temporal developments to be able to fully understand all the driving forces and the dynamics. By developing such an approach, Quah (1993a; 1993b) reveals the existence of a polarization into two clubs of rich and poor countries in the European regional income distribution (see Chapter 5.3). He calls these "twin peaks".

Baumont *et al*. (2003) approach the same problem in the form of a structural instability across spatial convergence clubs in the estimation of the β-convergence process among 138 European regions over the period 1980-1995. The estimation of the appropriate spatial regime error models shows that the convergence process is different across the regimes. Furthermore, the authors also estimate a strongly significant spatial spillover effect: the average growth rate of per-capita GDP in a given region appears to be positively affected by the average growth rate of neighbouring regions. Other particular specifications of club convergence are discussed by Rey (2001, 2003) and deal, in particular, with spatial Markov models and models for spatial inequality.

As already pointed out, most of the empirical analysis uses traditionally cross-sectional econometric techniques for testing convergence hypotheses. However, as Bernard and Durlauf (1995, 1996) suggested, these procedures pose several problems. The authors propose a new definition of convergence based on the unit-root concept developed in the context of time series analysis. If technological progress (which drives long-run economic growth) contains a stochastic trend, then convergence implies that permanent components of GDP are the same across regions. In this context, convergence is presented as a "catching up over a certain time period" (Bernard and Durlauf, 1996). Stochastic convergence can be tested by performing panel unit-root tests (see Evans and Karras, 1996a,b; Bernard and Jones, 1996; Fleissig and Strauss, 2001; Arbia and Costantini, 2004). So far however, no consideration has been given, in this field, to problems connected with the spatial nature of observations that would require explicit spatial econometric modelling. It is noticeable, however, that some of the concepts in time series analysis pertaining to unit roots and cointegration have been investigated in the context of spatial econometrics (Fingleton, 1999; Mur and Trìvez, 2003). Getis and Griffith (2002), notice that this important topic is still missing in the treatment of other spatial problems, the only noticeable exception being the works by Griffith and Tiefelsdford (2002) and Getis and Aldstadt (2004).

In the context of stochastic convergence, it is interesting to consider the approach recently proposed by Pesaran (2004b). This is based on the computation of convergence measures derived from a consideration of all possible pairs of (log) per-capita output gaps across, say, N economies.

A further innovative modelling framework recently proposed in the regional convergence literature, departing from the standard conceptualisation, is that developed by Arbia and Paelinck (2003a; 2003b). Starting from the consideration that traditional convergence analysis provides indication about the convergence of

regions towards common steady-states, but not about the path they follow to reach such convergence, the authors consider a continuous-time framework based on the classical Lotka-Volterra predator-prey system of two equations (Lotka, 1956 and Volterra, reprinted in Chapman, 1931) first proposed in an economic context by Samuelson (1971). (On continuous time econometric modelling see Bergstrom, 1990 and Gandolfo, 1990). They extend such a modelling framework to the case of more than two regions and explicitly introduce a modellization of the spatial dependence displayed by neighbouring regions. The model thus obtained can be seen as a generalized version of the β-convergence model, in which a system of regions, under some conditions, moves towards a mathematically stable point of convergence. They also consider statistical inference and introduce a discrete approximate solution based on Simultaneous Dynamic Least Squares (Paelinck, 1996) for estimating the model's parameters. Thus, by generalizing the traditional predator-prey model to a multiregional system, they show that each region may follow its own trajectory, leading to a series of distinct convergence paths. The authors illustrate their approach by comparing the empirical results generated by a traditional convergence equation and by a spatially-conditioned convergence equation with those of the Lotka-Volterra model for 119 European regions over the period 1980-1994.

Like Arbia and Paelinck (2003), Harvey and Carvalho (2002) are also interested in the dynamics of convergence rather than its occurrence within a certain time period. They propose a second-order error correction mechanism embedded within a stochastic convergence framework that provides an informative breakdown into trend, cycle and convergence components. They also show that time series tests of economic convergence can be formulated within this framework. No attention is paid in the quoted paper to spatial effects, however.

Fingleton (2004) proposes a new economic growth model that goes beyond the neoclassical scheme by incorporating (i) increasing (rather than constant) returns to scale, (ii) the diffusion of technological innovations, (iii) catch up and (iv) spatial externalities. The model is an extension of Verdoorn's law (Verdoorn, 1949) employed in economic analysis by Kaldor (1957; 1970) and augmented to include spatial lag, spatial errors and other spatial effects.

We have already treated the topic of panel data modelling in Section 6.2.1. Applications of such methodologies can be found in the study of regional convergence starting from the work of Islam (1995, 1998). For a recent application see, e. g. Arbia and Piras (2005) and Arbia *et al.* (2005).

Finally, Arbia *et al*. (2003) and Arbia and Basile (2005) propose a non parametric framework to study growth and convergence in the EU.

6.2.3 Space-Time Models

Often, one of the most critical points in spatial econometrics analysis is the need to consider simultaneously both spatial and temporal dependence present in the observations under examination. In the literature only few articles consider the

simultaneity of these occurrences. Here follows a short survey of some of the recent works in this direction. The basis for space-time modelling was founded in the seventies by Bennett (1979) and Pfeiffer and Deutsch (1980) who introduced the class of space-time autoregressive and moving average processes (STARMA). These processes still represent the point of departure for more complicated conceptualisations.

Pace *et al.* (1998) observed that there is no obvious optimal way of incorporating both spatial and temporal dependencies into empirically feasible pricing models. To better capture the effect of both spatial and temporal information on real estate prices, by overcoming the problems associated with indicator variable models, they introduced a spatio-temporal model which uses information from nearby recently sold properties when predicting the value of a given property. In other words, instead of assuming that each region has its own effects modelled by separate parameters, the STARMA formulation assumes that nearby properties have the same relation to observations across the entire sample. Using data on housing prices, they show the substantial benefits obtained by modelling the data's spatial as well as temporal dependence. In more detail, the spatio-temporal autoregression significatively reduced the median absolute error with reference to an indicator-based model. The improved performance of their specification is confirmed by the analysis of one-step-ahead forecast.

Giacomini and Granger (2003) compare the relative efficiency of different methods for forecasting the aggregation of spatially correlated time series. Using asymptotic approximations and the results of some simulation studies they show that forecasting performance can be improved by imposing a priori constraints on the amount of spatial correlation in the system. One way of doing so is to aggregate forecasts from a Space-Time Autoregressive model, as the latter offers a solution to the *curse of dimensionality* problem arising when forecasting using the VAR methodology. The importance of their paper lies in its proof that ignoring spatial correlation, even if weak, leads to highly inaccurate forecasts. It is important to stress, however, that the results are based on a very small sample simulation (based on a maximum of 16 observations laid on a 4x4 regular lattice) and are thus affected by strong edge effects. Furthermore the authors restrict their simulations to the case of positive spatial autocorrelation and do not say anything about the forecast's performance in the case of negative spatial autocorrelation.

Some advances have recently been recorded regarding the building of spatial and spatio-temporal ARCH models derived directly from the time series analogue. Developments in this field require a specification and a full understanding of the notion of "spatial risk" (Arbia, 2003) that has important bearing in fields like economic inequality and poverty analysis. Some of the ideas and developments in this field are reported in Florax *et al.* (2004). For a recent review of space-time modelling see Arbia (2004). On the relationship between the temporal and the spatial components of space-time model see Arbia (1992).

6.2.4 Discrete Variables

The basic spatial correlation model developed by Cliff and Ord (1981) and Anselin (1988) allows for spatial dependence in the dependent variable or in the error component referring to quantitative variables. However, many empirical studies have an explicit interest in modelling spatial dependence in cases where categorical variables are involved. Fleming (2004) include the concept of spatial correlation in models that involve limited dependent variables in a discrete choice context. Along similar lines the spatial probit model has been investigated in some recent works by Pinkse and Slade (1998), LeSage (2000), Beron *et al.* (2003), Murdoch *et al.* (2003) and Beron and Vijverberg (2004).

Garrett *et al.* (2003) adopt a framework similar to the standard spatial econometric techniques, but their specification is modified to account for the discrete nature of the dependent variable and the data's panel structure. They use a spatial probit to model a state's choice of branch banking and interstate banking regimes as a function of the regime choices made by other states. They extend the basic model by allowing spatial correlation to vary in different geographical regions.

6.2.5 Spatial Externalities

The modelling of spatial externalities is one of the most frequently addressed questions in the recent literature (Anselin, 2002; Lee, 2002; Dubin, 2003; Wall, 2004). Spatial externalities play a central role in many social sciences. For instance, increasing attention is being devoted in economics to the modelling of social interaction that introduces dependence among agents in a system. Furthermore, the theoretical focus on imperfect competition and increasing returns to scale led to a growing interest in the identification and measurement of spatial externalities (Anselin, 2003a). The empirical testing of such effects requires the formal specification of adequate spatial models. Anselin (2001a) outline a taxonomy of spatial econometric models that incorporate spatial externalities in various ways. The point of departure is a reduced form in which local or global spillovers are expressed as spatial multipliers. From this starting point, a range of familiar and less familiar specifications are derived for the structural form of a spatial regression. This work allowed some of more familiar models' limitations (in terms of their interpretation as models for spatial externalities) to become apparent. In a similar fashion, the work by Anselin *et al.* (1997) proposes a new approach to formalizing spatial externalities by combining spatial dependence and spatial heterogeneity in the form of spatial regimes.

6.2.6 Bayesian Models

Another aspect that merits consideration is the use of Bayesian models in spatial econometrics. LeSage (1997; 2000) has contributed the most to the diffusion of Bayesian techniques in spatial econometrics. See, also, LeSage (2004) for an up-

dated review. This author formulates a Bayesian probit model with individual effects that exhibits spatial dependencies. Since probit models are often used to explain variation in individuals' choices, these models may well explain spatial interaction effects due to the varying spatial location of the decision makers. The model proposed by LeSage allows for a parameter vector of spatial interaction effects that takes the form of a spatial autoregression. This model is an extension of spatial probit/logit models presented in LeSage (2000), and was applied to the 1996 presidential election results for US counties. In LeSage (2001), the author argued that the use of Bayesian methods in the estimation of geographical regressions can help solve the problems that may arise with the classical estimation, producing remarkable advantages over Ordinary Least Squares estimation in geographically-weighted regression methods. Finally, in LeSage (2004), the author proposes a family of locally evaluated regression models (termed Geographically Weighted Regression, GWR) that are based on kernel smoothing obtained through spatial distance decay functions. He also extends this framework to the so-called Bayesian Geographically Weighted Regression (BGWR) using robust estimators and parameter smoothing to deal with spatial outliers and spatial heterogeneity, respectively. Finally, he makes use of MCMC methods for estimation purposes.

Model comparison and selection is a central point in econometrics. Bayesian theory provides a comprehensive framework for such model choice. Hepple (2003) develops this Bayesian framework for the family of spatial econometric models. He derives the Bayes factor and the marginal likelihood for each of the main spatial specifications, and builds up the relative computational form. These framework is then applied to two different data-sets to illustrate the methods' advantages. An application of Bayesian methods is also provided by Baumont *et al.* (2003) who use Bayesian spatial econometric techniques to control for spatial autocorrelation, spatial heterogeneity and outliers in the empirical analysis of employment and population density in the area of Dijon.

Spatial priors for space-time modelling are also used by Dowde and LeSage (1997) and LeSage and Krivelyova (1999) amongst the others, as well as by Holloway *et al.* (2002) and LeSage (1997, 2000) in the spatial probit analysis.

6.2.7 Non-parametric Techniques

Developments of non-parametric and semi-parametric estimation methods as applied to spatial problems constitute a major departure from the classical spatial linear regression models considered in this book. One of the first contributions to this field was made by Conley (1999) who started observing that the traditional spatial ARMA modelling captures the dependence among regions by assuming that economic distances between them are measured with no measurement error. In order to avoid such an assumption (and to avoid, also, the complications of introducing this further source of uncertainty into the likelihood) the author suggests concentrating on moment-based estimation and assuming a non-parametric ap-

proach to modelling spatial dependence. In the paper cited, the author proposes the use of a Generalized Method of Moment estimator (GMM; see Hansen, 1982) that he proves to be consistent, as in the case of time series data. However, since the distribution theory of the estimators is different, he further proposes estimating the covariance matrix non-parametrically, allowing for spatial dependence by using methods that are analogous to those employed by, e. g. Newey and West (1987), Andrews (1991) White (1984) and White and Domowitz (1984) within the time series context. He suggests a covariance estimator that is based on a sequence of weighted averages of sample auto-covariances computed for subsets of observation pairs that fall within a given distance threshold in a fashion similar to that in other fields of spatial statistics such as geostatistics (see Cressie, 1991; p. 69). Finally, the author proves that the covariance matrices thus obtained are consistent estimators of the true one even when it is assumed that the true distances are distorted by measurement errors.

In a similar fashion, Driscoll and Kraay (1998) propose a consistent covariance estimation method for spatially dependent panel data, and Keleijan and Prucha (1999) discuss the use of GMM in a spatial model. Examples of semi-parametric estimation of spatial dependence may be found in Chen and Conley (2001) and Pace and LeSage (2002) .

Another important contribution in this area has been made by Gress (2003) who starts from the simple consideration that spatial models are not a straightforward extension of time series models, even if they show many shared characteristics. The paper cited combines non-parametric estimators of the mean with the usual parametric estimators of the spatial-lag parameters. This is done for the three models of primary interest in the literature: the Spatial Autoregressive Error model, the Spatial Autoregressive with exogenous variables model, and the Spatial Autoregressive with exogenous variables and spatially Autoregressive Errors model. Small sample properties with Monte Carlo simulations are then analysed and some empirical analyses are carried out in order to compare the result obtained using the usual econometric methods.

Parsimonious regression models using spatial data often yield non-normal, heteroskedastic and spatially dependent residuals. Pace *et al.* (2004) develop a model which simultaneously performs spatial and functional form transformations to mitigate the effects of this problem. They show that a better specification of the functional form could reduce spatial autocorrelation of errors given spatial clustering of similar observations and may also simultaneously reduce heteroskedasticity and residuals' non-normality. The authors apply this framework to housing prices' market data and obtain a good fit of the model with a pattern of residuals that is significantly improved with respect to the standard modelling framework.

McMillen and McDonald (2004) propose a non-parametric method for dealing with spatial heterogeneity within a probit model framework. Performance is evaluated through Monte Carlo experiments.

6.3 Alternative Tests

After the properties and characteristics of the dataset in use have been exploited, the focus of empirical works in a spatial regression context is mainly concentrated on misspecification testing. Since the work of Anselin (1988), the development of misspecification tests within a maximum likelihood framework and the derivation of these test's asymptotic properties and associated small sample properties have been of interest. The range of tests available in such a context has significantly improved in recent decades, particularly in relation to the simpler cases considered in the main body of this book.

In the context of testing the null hypothesis of spatial independence, many authors have devoted their attention to studying the properties of Moran's spatial correlation test under varying conditions and to extending its possible applications. Many tests for spatial dependence are, in fact, based on the Moran statistic or can be written in a like form (e. g. the Lagrange multiplier test (LMT); see Burridge, 1980). Some authors have explored the use of LMT in the context of spatial regression models (see Anselin and Rey, 1991; Anselin and Florax, 1995; Anselin, *et al.*, 1996). However, Anselin and Rey (1991) and Anselin and Florax (1995) used simulations to prove that Moran's *I* has a slightly better power than the LMT in small samples, even if the differences disappear when dealing with medium and large sized samples.

Among the new tests available, we recall the variant of Moran's *I* conceived by Kelejian and Prucha (2001) and a large sample test derived by Kelejian and Robinson (1992, 1997). A great number of unidirectional, multidirectional and robust Lagrange multiplier (LMT) tests have also been developed in the literature (Anselin and Griffith, 1988; Anselin and Florax, 1995; Anselin *et al.*, 1996; de Graaff *et al.*, 2001; Anselin and Moreno, 2003, Florax and de Graaf, 2004 and Saavedra, 2003). The vast literature on testing for spatial dependence includes work by Anselin and Kelejian (1997), and Kelejian and Robinson (1992). These tests focus on detecting spatially correlated residuals. Thus, they may be used to test well-defined misspecifications, such as a spatial autoregressive error process or an omitted spatially lagged dependent variable.

Tiefelsdorf and Boots (1995) have shown that, rather than assuming normality to derive approximate distribution of the Moran test, it is possible to construct an exact test based on numerical integration (see also Tiefelsdorf, 2002). Anselin and Keleijan (1997) have extended the application of Moran's *I* to test the hypothesis of dependence among the residuals of the Spatial Two Stages Least Squares (S2STLS) procedure (see Anselin, 1980; 1988), whilst Pinkse (1998) extends it to generalized residuals in a probit model. Of particular relevance in this context are two works by Pinkse (1999, 2004). In his first paper (Pinkse, 1999), the author establishes weak conditions under which the Moran test for spatial correlation (or a cross-correlation variant of it) has a limiting normal distribution under the null hypothesis of independence. For both tests, a result based on nuisance parameter is

provided which allows the test to be applied to proxies for the variables whose independence is to be tested. Pinkse uses the test statistic to determine whether three Lagrange multiplier tests for spatial correlation are valid in a probit model context. In his second paper (Pinkse, 2004), the author illustrates the general conditions under which Moran-flavoured tests for spatial correlation and spatial cross-correlation have a limiting normal distribution in the presence of a nuisance parameter in six frequently-encountered spatial models. The conditions which have to be imposed are weaker than those considered in a related work by the author (Pinkse, 1999) and the class of nuisance parameter problems allowed is far broader.

Generally speaking, the diagnostics for spatial error dependence can be classified as tests against an unspecified alternative to spatial correlation and tests against specific spatial processes. In the second case, the typical alternative to spatial autocorrelation is expressed in the form of a spatial autoregressive process. Many specification tests and estimation methods have been proposed for this model in the literature. Kelejian and Robinson (1993) suggested a different type of spatial process that combines a location-specific or local error component with a regional or spillover component in what they refer to as a spatial error component process. Anselin and Moreno (2003) proposed an interesting alternative based on the following formulation of the model. They considered a number of specification tests against this alternative, based on both a maximum likelihood framework and on a generalized methods of moments estimation approach. Furthermore, they compared the performance of these tests in a series of Monte Carlo simulation experiments for a range of different spatial layouts and under a number of different error distributions, and found that the new statistics perform better in terms of power, particularly with reference to those cases that are not covered by the normality assumption. Similarly, the variant of the Kelejian-Robinson statistics, which was suggested in this paper to account for second order neighbours, also performs well. Thus, the degree of generality of the paper's results is limited by the design taken into account in the simulation experiments.

In a recent work, Baltagi and Li (1999) observed that none of the LM tests for spatial dependence available in the literature have been computed by running an artificial regression. The main purpose of their paper is to show that simple LMT for both spatial lag dependence and spatial error dependence can be obtained by using the Double Length Regression (DLR) proposed by Davidson and MacKinnon (1993). Double length regressions are useful econometric tools for deriving LM and equivalent test statistics. Moreover, they can be applied to models that are more general than some of the other artificial regression counterparts and have better finite sample properties than their outer product gradient regression counterparts. Baltagi and Li (1999) derive the DLR tests for both spatial lag dependence and spatial error dependence by only using Least Squares residuals of the restricted model. They use two simple examples to illustrate these tests: one is based on the simple crime relationship considered by Anselin (1988) and the other uses the Irish data considered by Ord (1975). In addition, Monte Carlo experiments are

performed to study the small sample performance of these tests and, as expected, they have a performance similar to that of their corresponding LMT counterparts.

Pesaran (2004a) is interested in testing cross-sectional dependence in dynamic panels. He starts from the consideration that the traditional way of testing dependence amongst spatial cross-sectional units in panel data is too dependent on the subjective choice of a connectivity matrix. Furthermore, a purely geographical connectivity is not appropriate in many economic applications where economic and sociopolitical factors could better explain the pattern of dependence. As an alternative, he proposes a test based on a simple average of all pair-wise correlation coefficients of the OLS residuals from the single regressions contained in the panel. For the cross-section dependence test thus obtained, asymptotic and small sample properties are proved and compared with the standard LMT test by using Monte Carlo experiments. A proper spatial correlation test is derived as a generalization to cases where we have an *a-priori* spatial order for the cross-sectional units.

More recent contributions on hypothesis testing in the spatial context may be found in Kelejian and Robinson (2004) where the authors expand their previous work to deal with test statistics for multiple sources of misspecification in linear regression models.

Finally, some tools have been introduced to test spatial dependence non-parametrically. A non-parametric test for spatial independence can be found in Brett and Pinkse (1997) based on a similar test for serial independence introduced by Pinkse (1998)

In the context of testing for heteroskedasticity in a spatial model, Baltagi *et al.* (2003) extended the Breusch and Pagan LM test to the case of a spatial error component model and derived several Lagrange multiplier tests (LMT) for the panel data regression model with spatial error correlation. The starting point is to allow for both spatial error correlation and regional effects in panel data regression models and to test their joint significance. The authors use some Monte Carlo experiments to show clearly that the spatial econometric literature should not ignore heterogeneity across cross-sectional units when testing for the presence of spatial error correlation. Similarly, the panel data econometric literature should not ignore the spatial error correlation when testing for the presence of random regional effects. Baltagi *et al.* (2003) derive joint and conditional LM tests that are easy to implement and more powerful than the one-dimensional LM test. The sample size used in the Monte Carlo analysis is small, as is typical in micro panels.

Generally speaking, the specification search in spatial econometrics has focused on the detection of spatial autocorrelation and spatial heteroskedasticity. The joint occurrence of autocorrelation and heteroskedasticity has not been addressed, even if it is a central concept in spatial econometrics, as already remarked in Chapter 4.4.3. Anselin and Griffith (1988) suggest that the traditional spatial autocorrelation tests may have some power against heteroskedasticity, but Kelejian and Robinson (1992) reach the opposite conclusions. No theoretical result yet

exists to show how heterogeneity affects spatial correlation tests. For a recent review and new results in this field, see Keleijan and Robinson (2004).

The specification testing procedures available in the literature, consist in expanding a spatial linear regression model with spatially lagged dependent variables, conditional upon the results of a misspecification test. Florax *et al.* (1998) bring together a number of new specification search strategies. In particular, they investigate a Hendry-like specification strategy, starting from the spatial common-factor model and subsequently reducing the number of spatially lagged variables on the basis of significance tests. Their experimental simulation pertains to different samples of varying sample sizes and the spatial field is modelled on regular lattice surfaces. They conclude that the classical forward stepwise approach outperforms the Hendry strategy in terms of finding the true data generating process as well as in the observed accuracy of the estimators for spatial and non-spatial parameters. It also dominates the concurrent stepwise approach suggested in the literature. Florax and Rey (1995) investigated the small sample performances of a sequence of unidirectional and multidirectional tests and tests for local misspecification, both in terms of the probability of finding the real data-generating process, and in terms of the mean squared error of the estimated parameters. They established the existence of a dominant strategy that is highly relevant for practitioners of spatial regression modelling.

Finally de Graaff *et al.* (2001) suggested a very particular test. Their paper is concerned with a methodological and empirical analysis of chaos in spatial systems. Their aim is to create a link between the classical diagnostic tools developed in spatial econometrics and the non-linearity tests for empirical data series, with particular regard to the so-called BDS test (Brock *et al.,* 1987). They developed a spatial variant of this test and subsequently applied it to the case of a shift-share model for Dutch regional labour markets over the period 1987-1992.

6.4 Alternative Estimation Methods

Until a few years ago, the prevalent estimation methods were basically founded on GLS and ML estimators for the spatial lag and spatial error model, as discussed in earlier chapters. Recently, significant advances have been made in developing alternatives.

A major step forward in the direction of providing a viable alternative to the traditional GLS and ML estimators, has been made by Kelejian and Robinson (1997), Kelejian and Prucha (1998; 1999) and the already-cited contribution of Conley (1999) who introduced the use of GMM estimators in a spatial econometric context. In particular, Kelejian and Prucha (1999) develop a set of moment conditions that yield estimation equations for the parameter of a SAR error model and suggest the use of non-linear least squares to derive a consistent generalized method of moments estimator. Conley (1999) studies the properties of GMM estimators in detail and provides the formal conditions for their consistency and as-

ymptotic normality. These results enable estimation, inference and hypothesis testing with dependent cross-sectional data.

Pinkse and Slade (1998) use moment conditions in their probit model with SAR errors already quoted in Section 6.2.4, however not explicitly dealing with the structure of the spatial covariance matrix to be estimated.

Pinkse *et al.* (2003) provide a very interesting example of such moment-based procedures. They make use of a one-step GMM estimator that allows for generic spatial and time series dependence. This estimator is, moreover, consistent and asymptotically normal under weak conditions. They use this new estimation procedure to estimate a dynamic spatial probit model with fixed effects which, in turn, enables operational decisions in a real-options context. The main result of the empirical application is that the data are more supportive of a mean/variance-utility model than of a real-option model. Kelejian and Prucha (2004) offer another example of the use of GMM estimators for the autoregressive parameter of a spatial model.

A further alternative estimator is the spatial two-stage least squares method (S2SLS) originally proposed by Anselin (1980; 1988) and developed by Land and Leane (1992), Kelejian and Robinson (1993) and Kelejian and Prucha (1998). It deals with the endogeneity of spatially lagged dependent variables by making use of instrumental variables. Under a set of conditions that are often satisfied when dealing with connectivity matrices based on pure contiguity, the method is proved to be consistent and produces asymptotically normal estimators. Extension to a three stages least squares is discussed in Anselin (1988).

Computationally intensive techniques have recently been introduced to estimate the parameters of spatial models. A recent example is the recursive importance sampling estimator (RIS) proposed by Vijverberg (1997) and applied by Beron and Vijverberg (2004) to solve spatial problems.

Pinkse *et al.* (2002) studied spatial price competition among firms producing differentiated products and competing in U.S. wholesale gasoline markets. They found that, in this market, competition is highly localized. In this interesting and innovative work, the authors make use of an instrumental variables estimator for the matrix of cross-price response coefficients. They also demonstrate that the estimator is consistent and derive its asymptotic distribution. They make use of a semi-parametric approach that allows discriminating between models of global competition (in which all product compete with all others) and local competition (in which products compete only with their neighbours).

The already cited work by Elhorst (2001) highlights an interesting aspect of the progress made with regards to estimation methods. The author discusses the estimation methods of a fixed-effects dynamic panel data model extended either to include spatial error autocorrelation or a spatially lagged dependent variable. It is well known that the traditional Least Squares dummy variable estimator leads to inconsistent estimates of the parameters of interest. In fact, no straightforward estimation procedure is yet available. The reason for this has to be sought in the fact that existing methods developed for dynamic (but non-spatial) methods and for

spatial (but non-dynamic) panel data models might produce biased estimates when these methods are considered jointly. In order to overcome this problem, the models are first-differenced to eliminate the fixed effects and then the unconditional likelihood function is derived by taking account of the density function of the first-differenced observations on each spatial unit. This procedure yields a consistent estimator both of the response parameters and of the spatial autocorrelation coefficient when the cross-sectional dimension tends to infinity and this is regardless of the time dimension size. The rows and the columns of the spatial weight matrix do not diverge to infinity at a rate equal to (or faster than) the rate of the sample size in the cross-section domain. The only problem still to be solved lies in the estimation of the fixed-effect. The latter cannot be estimated consistently, since the number of these coefficients increases as the number of observations increases (the *curse of dimensionality* problem). Furthermore, there is evidence of the fact that, when exogenous variables are omitted, the exact likelihood function can be specified, but, when, in contrast, exogenous variables are included, the pre-sample values of this variable (and thus the likelihood function) can only be approximated. The work cited considers two cases for modeling the pre-sample values of the exogenous variables for the first-differenced observations on each spatial unit: the Bhargava and Sargan (1983) approximation and the Balestra and Nerlove one (Balestra and Nerlove, 1966). The decision to exclude exogenous explanatory variables is based on the fact that the presence of such variables further complicates the analysis although different approaches have been suggested in the econometric literature to deal with the pre-sample values of these variables in a dynamic context.

An important problem arising in spatial dependence estimation procedures relates to computational aspects. This problem, already discussed in Chapter 3.2.4, was predominant in the past when computer capabilities were more limited. Chapter 3.2.4 examined the algorithm proposed by Ord (1975) involving an eigenvalues evaluation which made spatial estimation practical for small to moderate-sized datasets. Along the same line of research, Anselin (1988), Haining (1990), Anselin and Hudak (1992), Griffith (2004) and others have worked on the implementation of spatial estimation procedures by writing codes that reduce the computational burden and make the estimation feasible in practical situations.

Despite the dramatic improvements in computer capabilities, some operations such as determinants, eigenvalues, and inverse evaluation remain unsolved problems when dealing with very large spatial databases. Pace and Barry (1997a) provide a way of quickly computing parameters' estimations when the dependent variable follows a spatial autoregressive scheme. They start from the evidence that, in some instances, only a few observations are influential on the neighbouring sites, whereas others have a negligible effect and so the spatial weight matrix may become sparse. In this way, by rearranging the weights' matrix, the estimates of the parameters of interest may be computed at a very low cost, even in the presence of large datasets. In demonstration of the technique's accuracy, the work cited provides Monte Carlo evidences of the short time a computer needs to produce estimates in the presence of very big datasets.

Further methods for reducing the computational burden of increasingly time-demanding spatial models have been proposed by Smirnov and Anselin (2001) (using a characteristic polynomial approach) by Griffith (2000, 2004) (using polynomial functions approximations) and by Pace and LeSage (2004) and Pace and Zou (2000) (using Chebyshev approximations). Exact methods based mainly on decomposition techniques for sparse matrices, (e. g. Choleski or Lu decomposition. See Gentle, 1998; Press *et al.*, 1992), have also been exploited by Pace (1997) and Pace and Berry (1997b, 1997c) using an idea originally contained in Arbia (1986)

6.5 Exploratory Tools

For the sake of completeness, it is necessary to refer to one final emerging field in spatial econometrics. This regards the set of techniques developed as a preliminary way of looking at data prior to a more rigorous formalization within an explanatory spatial econometric model. Such tools are known in the literature as Exploratory Spatial Data Analysis (ESDA) and can be seen as a natural extension of the statistical methods known as Exploratory Data Analysis (EDA; see Tuckey, 1977). A general introduction to ESDA within a spatial statistical application can be found in Haining (1990; 2003). ESDA methods are devised to visualize data, describe spatial variability and identify spatial outliers. They usually take the form of plots, graphs and global or local spatial association measures. We can classify instruments such as Moran's scatterplot (proposed by Anselin, 1996), the Getis-Ord local statistics (Getis and Ord, 1992; Ord and Getis, 1995) and the class of Local Indicators of Spatial Association (or LISA, Anselin, 1995b) within this category. See Haining (1990), Bailey and Gatrell (1995) and Anselin (1996) for reviews. See, also, Ertur and Le Gallo (2003) for a review and applications to European regional disparities.

Appendix: A Review of the Available Software for Spatial Econometric Analysis[1]

A.1 Introduction

The diffusion of statistical software certainly plays a fundamental role in empirical studies. As we have seen in Chapter 4, spatial econometric analysis often requires *ad hoc* routines in order to implement estimation and hypothesis-testing techniques. Thus the difficulties associated with programming such routines in the absence of a dedicated software has certainly been one of the factors contributing in the past to the slow diffusion of empirical studies in spatial econometric analysis.

In recent years, the landscape has changed dramatically and a number of options for applying spatial econometric methodologies in real cases are currently on offer to the interested researcher.

The purpose of this appendix is twofold. It seeks both briefly to review the software currently available for implementing a spatial econometric approach to data and to describe the main functions supported by the various programmes. In writing this section, we are conscious that this Appendix is doomed to become rapidly obsolete, given the speed with which the situation is developing in this field. The reader is referred to the websites quoted here for a continuous update on the situation, in the authors' hope that the website addresses will not also change as rapidly as their content.

Until a few years ago (when there was already a widespread diffusion of software for dynamic econometric analysis), the SpaceStat package (Anselin, 1992a; 1992b) represented the only opportunity available for researchers engaged in spatial data analysis. Indeed, until 1995, this was the only freestanding software and, even now, it is the most complete in terms of the wealth of facilities it provides. As a consequence, we will start our review in Section A.2 with this programme. In Sections A.3 and A.4 we will consider two alternatives that are currently available: the GeoDa package and a series of toolboxes written in various programming languages. Our aim is to give the reader some indications and help him choose from the various alternatives by describing the main functions and utilities they support.

---

[1] This Appendix has been written jointly with Gianfranco Piras.

A.2 The SpaceStat Programme

Historically, SpaceStat was the first programme written for spatial econometric analysis and it is still the one most widely used by researchers today. Before SpaceStat, no other software was available to run specific spatial econometric or spatial statistic analysis. The first release dates back to 1991 and it has been upgraded several times since then. Obviously, it is not the aim of this section to give a detailed description of all the functions the programme supports. We will simply summarize some of the main features, whilst referring the interested reader to Anselin (1992a, 1992b, 1995a) and the website[2] for further information.

One important utility offered by SpaceStat regards the possibility of imputing and manipulating spatial weights matrices. The programme also presents some features relating to explanatory spatial data analysis (see Chapter 6.5), including techniques for describing and visualizing spatial distributions, identifying patterns of spatial association (spatial clusters) and suggesting different spatial regimes. The software's most powerful tools, however, are those relating to the estimation and hypothesis-testing of spatial regressions: they allow spatial dependence and spatial heterogeneity to be incorporated within the modelling framework. In particular, the programme allows the ML estimation of the parameters in the *spatial lag* and the *spatial error* models as well as the testing of the various misspecification hypotheses, including normality (the Jarque-Bera test), heteroscedasticity (the Breusch-Pagan test) and spatial dependence (the Lagrange multiplier and Likelihood ratio test).

SpaceStat's main drawback is its interface. The programme is written in the "Gauss" language and presents an unpleasant, old-fashioned, black-and-white Dos-like interface, characterized by command lines, no mouse and no windows-assisted commands. An extension to the programme allows data interchange with ArcView (the Geographical Information System produced by ESRI) and provides facilities that are particularly useful when building connectivity matrices, running exploratory data analysis and visualizing output maps.

A.3 GeoDa

GeoDa is a very recent programme designed to implement techniques for exploratory data analysis on spatial data in the form of points or polygons in a geographical space. It represents a dramatic evolution with respect to SpaceStat and, like its ancestor, was developed by Luc Anselin and his co-workers (Anselin, 2003b, 2004; Anselin *et al.*, 2004c). One of its main advantages over SpaceStat is the fact that it provides a user-friendly graphical interface based on a windows environment.

---

[2] http://www.terraseer.com/products/spacestat.html.

The programme is undergoing a very rapid phase of evolution. The first version was released in February 2003 and was followed by a second in June 2003. The most recent edition became available in January 2004. At the moment of writing, the package may be downloaded from its website[3] free of charge.

So far, the programme has mainly focused on graphical tools and simple descriptive spatial analysis such as spatial autocorrelation statistics, the analysis of spatial outliers and a wide range of functions relating to explanatory spatial data. In this respect, it allows the evaluation of global and local spatial autocorrelation by means of the Moran's *I* spatial autocorrelation statistic and the graphical tool known as the *Moran Scatter Plot* (Anselin, 1996; see also Chapter 6.5).

The spatial econometric regression routines, on the other hand, are still very limited in their range and currently only allow estimation of the classical a-spatial linear regression via OLS and Maximum-Likelihood estimators of the parameters associated with the *spatial error* and the *spatial lag* models. The basic diagnostic for spatial dependence, spatial heteroskedasticity and normality are available for the standard OLS regression residuals. Asymptotic inference is based on the Likelihood Ratio Test and on an estimate of the asymptotic covariance matrix using the algorithm developed by Smirnov (2003).

As regards spatial weights matrices, the programme offers the interesting possibility of building matrices based on different criteria simply by reading a digitalized map (through a *shape file*). However, the estimation procedures only admit symmetric structures for the spatial weight (e. g. contiguity or distance-based weights) and they cannot be performed on more sophisticated structures such as a *k*-nearest neighbours weighting scheme, for example.

GeoDa only contains a few functions dedicated to mapping and geo-visualizing data that are not comparable with those available in the more sophisticated Geographical Information Systems. The functions implemented, however, are by and large sufficient for the needs of econometricians whose main interest is in data description and preliminary explanatory analysis. On the other side of the coin, the programme is definitely much easier to use by non experts than the commercially available GIS.

A.4 Toolboxes

SpaceStat and GeoDa are the only dedicated products available for spatial data analysis. Recently, however, an increasing number of initiatives set up by individual researchers or groups have made specific routines available. These have been written in various programming languages and can be invaluable to those working in the field of applied spatial econometrics.

---

[3] Its installation routine is available at the internet address: http://sal.geoda.uiuc.edu/default.Php and contains all the required files and libraries.

Perhaps the best known of the toolboxes for spatial econometric analysis are those developed by Pace[4] (see Pace and Barry, 1998) and by LeSage[5]. Both of them make use of Matlab routines. Kelly Pace's toolbox is more oriented toward estimating spatial models for large data samples (e. g. those found in environmental and physical applications). James LeSage's toolbox, on the other hand, is more oriented toward economic modelling, with a particular attention to the Bayesian spatial methods (including Gibbs sampler routines) discussed in Chapter 6.2.7. In addition, it provides routines for the classical hypothesis-testing and estimation procedures relating to the *spatial lag* model, the *spatial error* model and *Anselin's general spatial model* (see Chapter 4.3.7.1) as well as the *spatial probit*, *logit* and *tobit* models (see Chapter 6.2.4) and their robust versions. Furthermore, some routines are also available for quickly estimating spatial models using the GMM technique (see Section 6.2.8) and testing accuracy via Monte Carlo simulations. The library also features specialized sparse matrix procedures for handling large data sets. All in all, LeSage's Econometric toolbox contains around 50 Matlab functions and the spatial econometric functions constitute just one part.

The Stata toolbox displays similar features. This library contains regression diagnostics, maximum-likelihood estimation and routines for implementing the Conley GMM estimator (Conley, 1999). See Pisati (2001)[6], for a detailed description.

There exist many web communities that make programmes and routines for spatial econometrics available. One that deserves particular attention is that based on the R programming language[7] and linked to the R-Geo initiative. This library has a lot of new functions for analysing spatial data, including descriptive spatial autocorrelation statistics and a complete set of spatial econometric functions. Bivand's SPDEP (Spatial Analysis Tools)[8] is also written in the R language and provides programmes for spatial autocorrelation and regression analysis (see Bivand and Gebhardt, 2000; Bivand, 2002, and Bivand and Portnov, 2004).

Last, but not least, the S+SpatialStats extension to the S-PLUS statistical package (Kaluzny et al., 1997) includes some spatial regression routines[9] and the Geobugs extension to the Winbugs[10] programme contains routines specifically devoted to Gibbs sampler and Monte Carlo Markov Chain (MCMC) spatial model estimation.

---

4 Available at the website http://www.spatial-statistics.com/.

5 LeSage's econometrics toolbox is downloadable without charge from the web site: http://www.spatial-econometrics.com/.

6 For Stata toolbox see http://www.faculty.econ.nwu.edu/faculty/conley/statacode.html web site.

7 See the website http://sal.uiuc.edu/csiss/Rgeo/.

8 See the website http://cran.r-project.org/src/contrib/packages.html.

9 See the website http://www.insightful.com/products/splus/default.asp.

10 See http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml.

References

Agresti A., 1990. *Categorical Data Analysis*, Wiley, New York.

Aitkin M., Anderson D.A., Francis B.J. and Hinde J.P., 1989. *Statistical Modelling in GLIM*, Clarendon Press, Oxford.

Aitchinson J. and Silvey S.D., 1958. Maximum likelihood estimation of parameters subject to restraints, *Annals of Mathematical Statistics,* 29: 813-828.

Aitchinson J. and Silvey S.D., 1960. Maximum likelihood estimation procedures and associated tests of significance, *Journal of the Royal Statistical Society, Series B,* 22: 154-171.

Alfò M. and Postiglione P., 2002. Semi-parametric modelling of spatial binary observations, *Statistical Modelling*, Arnold Publishers, 2: 123-137.

Ali M.M. and Giacotto C., 1984. A study of several new and existing tests for heteroscedasticity in the general linear model, *Journal of the American Statistical Association*, 80: 730-735.

Andrews D.W.K., 1991. Heteroscedasticity and autocorrelation-consistent covariance matrix estimation, *Econometrica*, 59: 817-858.

Anselin L., 1980. *Estimation Methods for Spatial Autoregressive Structures*, Ithaca, NY, Cornell University, Regional Science Dissertation and Monograph series, no. 8.

Anselin L., 1987a. Model validation in spatial econometrics: a review and evaluation of alternative approaches, *International Regional Science Review*, 11:279-316.

Anselin L., 1987b. Spatial dependence and spatial heterogeneity: a closer look at alternative modelling approaches, *Working Paper*, Department of Geography, University of California, Santa Barbara, CA .

Anselin L., 1988. *Spatial Econometrics, Methods and Models*, Kluwer Academic, Boston.

Anselin L., 1992a. *SpaceStat: A Program for the Analysis of Spatial Data*, National Centre for Geographic Information and Analysis, University of California, Santa Barbara, CA.

Anselin L., 1992b. *SpaceStat Tutorial: A Workbook for Using SpaceStat in the Analysis of Spatial Data*, National Centre/er* for Geographic Information and Analysis, University of California, Santa Barbara, CA.

Anselin L., 1995a. *SpaceStat Version 1.80 User's Guide*, Regional Research Institute, West Virginia University, Morgantown,WV.

Anselin L., 1995b. Local indicators of spatial association – LISA, *Geographical Analysis,* 27(2): 93-115.

Anselin L., 1996. The Moran scatterplot as an exploratory spatial data analysis tool to assess local instability in spatial association, in M.M. Fischer, H.J. Scholten and D.Unwin (eds.), *Spatial Analytical Perspective on GIS*, Taylor & Francis, London, 111-125

Anselin L.,2001a. Spatial econometrics, in B.H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Basil Blackwell, Oxford: 310-330.

Anselin L.,2001b. Rao's score test in spatial econometrics, *Journal of Statistical Planning and Inference* 97(1): 113-139.

Anselin L., 2002. Under the hood: issues in the specification and interpretation of spatial regression models, *Agricultural Economics*, 17: 247-267.

Anselin L.,2003a. Spatial externalities, spatial multipliers and spatial econometrics, *International Regional Science Review*, 26: 147-152.

Anselin L.,2003b. *GeoDa 0.9 User's Guide*, Spatial Analysis Laboratory (SAL). Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, IL.

Anselin L., 2004. *GeoDa 0.95 Release Notes*, Spatial Analysis Laboratory (SAL). Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, IL.

Anselin L. and Bera A.K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics, in A. Ullah and D. Giles (eds.), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York: 237-289.

Anselin L. and Florax R., 1995. *New Directions in Spatial Econometrics*, Springer-Verlag, New York.

Anselin L. and Griffith D.A., 1988. Do spatial effects really matter in regression analysis? *Paper of the Regional Science Association*, 65: 11-34

Anselin L. and Hudak S., 1992. Spatial econometrics in practice: a review of software options, *Regional Science and Urban Economics,* 22: 509-536.

Anselin L. and Kelejian H.H., 1997. Testing for spatial error autocorrelation in the presence of endogenous regressors, *International Regional Science Review*, 20: 153-180.

Anselin L. and Moreno R., 2003. Properties of tests for spatial error components, *Regional Science and Urban Economics*, 33, 5: 595-618.

Anselin L. and Rey S., 1991. Properties of tests for spatial dependence in linear regression models, *Geographical Analysis*, 23: 110-131.

Anselin L., Bera A.K., Florax R.J.G.M. and Yoon M.J., 1996. Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics*, 26: 77-104.

Anselin L., Florax R.J.G.M. and Rey S., 2004. *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, New York.

Anselin L., Le Gallo J. and Jayet H., 2004b. Spatial panel econometrics, in L. Matyas and P. Sevestre (eds), *The Econometrics of Panel Data*, Third edition, Kluwer Academic Publishers, Dordrecht.

Anselin L., Syabri I. and Kho Y., 2004c. GeoDa: an Introduction to spatial data analysis, *Geographical Analysis.* Forthcoming.

Anselin L., Varga A. and Acs Z., 1997. Local geographical spillovers between university research and high technology innovations, *Journal of Urban Economics*, 42: 422-448.

Arbia G., 1986. On the simulation of two-dimensional stationary stochastic processes: a comparison of different approaches, in F. De Antoni, C. Lauro and A. Rizzi (eds), *Proceedings of the Seventh Symposium in Computational Statistics, Short Communications and Posters,* Physica-Verlag, Rome: 17 – 18.

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

Arbia G., 1992. Sulla verifica empirica di teorie economiche: serie storiche o serie spaziali, *Rassegna di Statistiche del Lavoro*, Centro Studi Confindustria, 3: 97-105, Rome.

Arbia G., 1993. Recenti sviluppi nella modellistica spaziale, in S. Zani (ed) *Metodi statistici per le analisi territoriali*, Franco Angeli Editore, Milan: 193-217.

Arbia G., 1995a. Updating existing sampling designs in repeated environmental surveys. *Working Paper, Padua*, 6.

Arbia G., 1995b. Le relazioni degli aggregati economici nel tempo e nello spazio, in G. Alvaro (ed), *Contabilita' Nazionale e Statistica Economica*, Cacucci editore, Second edition, Bari: 883-958.

Arbia G., 2001a. Modelling the geography of economic activities on a continuous space, *Papers in Regional Sciences,* 80: 411-424.

Arbia G., 2001b. The role of spatial effects in the empirical analysis of regional concentration*, Journal of Geographical Systems,* 3: 271-281.

Arbia G., 2003. Bivariate Value-at-risk, *Statistica*, 2: 231-247.

Arbia G., 2004. Alternative approaches to regional convergence exploiting both spatial and temporal information. To appear in *Estudios de Economia Aplicada,* 22(2): 429-448.

Arbia, G. and Basile, R. 2005. Spatial Dependance and Non-Linearities in regional growth behaviour in Italy, *Statistica*, Bologna (in press).

Arbia G. and Costantini M., 2004. Testing the stochastic convergence of Italian regions using panel data. Paper presented at the *"Computational Management Science"* Conference and Workshop on *"Computational Econometrics and Statistics"*, April, Neuchâtel, Switzerland. To appear in *Applied Economics Letters*, Routledge.

Arbia G. and Espa G., 1996a. *Statistica Economica Territoriale*, CEDAM, Padua.

Arbia G. and Espa G., 1996b. Forecasting statistical models of archaeological site location, *Archeologia e Calcolatori*, Rome.

Arbia G. and Lafratta G., 1997. Evaluating and updating the sample design: the case of the concentration of SO_2 in Padua, *Journal of Agricultural, Biological and Environmental Statistics*, 2(4): 451-466.

Arbia G. and Lafratta G., 2001. Spatial sampling designs optimized under unisotropic superpopulation models, *Journal of the Royal Statistical Society, Series C – Applied Statistics*, 51, 2: 223-241.

Arbia G. and Piras G., 2005. Convergence in per-capita GDP across EU-NUTS2 regions using panel data models extended to spatial autocorrelation effects, *T03-05 REAL Working Paper*, University of Illinois at Urbana-Champain.

Arbia G. and Paelinck J.H.P., 2003a. Spatial econometric modelling of regional convergence in continuous time, *International Regional Science Review,* 26: 342-362.

Arbia G. and Paelinck J.H.P., 2003b. Spatial econometric modelling of regional convergence in continuous time, *Journal of Geographical Systems*, 5: 291-314.

Arbia G. and Switzer P., 1994. Optimal stepwise spatial sampling designs, *Working Paper*, Padua University.

Arbia G., Basile R. and Piras G., 2005. Using spatial panel data in modelling regional growth and convergence, *ISAE Working Paper*, no. 51-2005.

Arbia G., Basile R and Salvatore M., 2002. Regional convergence in Italy 1951-1999: a spatial econometric perspective, *ISAE Working Paper*, 29/02.

Arbia G., Basile R. and Salvatore M., 2003. Spatial effects on regional growth. A parametric and a non-parametric approach. Paper presented at the conference on *Analytical Frontiers in Spatial Aspects of Economic Development,* May, WIDER, Helsinki.

Arbia G., Benedetti R. and Espa G., 1996. Effects of MAUP on image classification, *Geographical Systems*, 3: 123-141.

Arbia, G, Le Gallo, J. and Piras, G. 2005. A meta-analysis of Regional Convergence of the NUTS-2 European regions, 1977-2002, Paper presented at the 45th Congress of the ERSA, Vrije Universitat, Amsterdam, August, 2005, submitted to *Spatial Economic Analysis*, Routledge.

Aten B., 1996. Evidence of spatial autocorrelation in international prices, *Review of Income and Wealth*, 42: 149-163

Aten B., 1997. Does space matter? International comparison of the prices of tradable and non-tradables, *International Regional Science Review,* 20: 35-52.

Azariadis C. and Drazen A., 1990. Threshold externalities in economic development, *Quarterly Journal of Economics*, 105: 501-526.

Azzalini A., 1996. *Statistical Inference: based on Likelihood*, Chapman and Hall, London.

Bailey T. and Gatrell A.C., 1995. *Interactive Spatial Data Analysis*, Longman, Harlow.

Balestra P. and Nerlove M., 1966. Pooling cross-section and times-series data in the estimation of a dynamic model: the demand for natural gas. *Econometrica*, 34:3.

Baltagi B.H., 1999. *Econometrics*, (third edition), Springer Verlag, Berlin.

Baltagi B.H., 2001. *Econometric Analysis of Panel Data*, (second edition), John Wiley and Sons, Chichester, England.

Baltagi B.H. and Li D., 1999. Double-length artificial regressions for testing spatial dependance, *Econometric Review*, 20: 31-40.

Baltagi B.H., Song S.H. and Koh W., 2003. Testing panel data regression models with spatial error correlation, *Journal of Econometrics*, 117: 123-150.

Banerjee S., Carlin B.p. and Gelfand A.E., 2004. *Hierarchical modelling and analysis for spatial data*, Chapman & Hall/CRC, Boca Raton.

Barro R.J. and Sala-i-Martin X., 1995. *Economic Growth*, McGraw-Hill.

Bhargava A. and Sargan J.D., 1983. Estimating dynamic random effects models from panel data covering short time periods, *Econometrica*, Econometric Society, vol. 51(6), 1635-59.

Bartels C., 1979. Operational statistical methods for analysing spatial data. In C. Bartels and R. Ketellapper (eds.), *Exploratory and Explanatory Analysis of Spatial Data,* Martinus Nijhoff, Boston: 5-50.

Bartlett M.S., 1963. The spectral analysis of point processes, *Journal of the Royal Statistical Society, B,* 25: 264-281.

Bartlett M.S., 1971. Physical nearest-neighbour models and non-linear time-series, *Journal of Applied Probability*, 8: 222-232.

Bartlett M.S., 1975. The statistical analysis of point pattern, *Advances in Applied Probability*, 6: 336-358.

Bartlett M.S. and Besag J., 1969. Correlation properties of some nearest-neighbour models, *Bulletin of the International Statistical Institute*, 43(2): 191-193.

Basu S. and Thibodeau T.G., 1998. Analysis of spatial autocorrelation in housing prices, *The Journal of Real Estate Finance and Economics*, 17: 61-85.

Bates C. and White H., 1985. A unified theory of consistent estimation for parametric models, *Econometric Theory*, 1: 151-78.

Baumont C., Ertur C. and Le Gallo J., 2003. Spatial convergence clubs and the European regional growth process, 1980-1995, in B. Fingleton (ed.), *European Regional Growth,* Springer-Verlag, Berlin.

Baumont C., Ertur C. and Le Gallo J., 2004. Spatial analysis of employment and population density: the case of the agglomeration of Dijon, 1999, *Geographical Analysis*, 36, 2, 146-176.

Bennett R.J., 1979. *Spatial Time-series Analysis, Forecasting and Control*, Pion, London.

Bera A.K. and Jarque C.M., 1982. Model specification tests: a simultaneous approach, *Journal of Econometrics*, 20: 59-82.

Bergstrom A.R., 1990. *Continuous Time Econometric Modelling*, Oxford University Press, Oxford.

Bernard A.B. and Durlauf S.N., 1995. Convergence in international output, *Journal of Applied Econometrics*, 10: 97-108.

Bernard A.B. and Durlauf S.N., 1996. Interpreting tests of the convergence hypothesis, *Journal of Econometrics,* 71: 161-173.

Bernard A.B. and Jones C., 1996. Technology and convergence, *Economic Journal,* 106.

Berndt E.R., 1991. *The Practice of Econometrics, Classic and Contemporary*, Addison-Wesley Publishing Company, Reading.

Beron K.J. and Vijverberg W.P., 2004. in L. Anselin, R.J.G.M. Florax and S. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, Berlin.

Beron K.J., Murdoch J.C. and Vijverberg W.P., 2003. Why cooperate? Public goods, economic power and the Montreal protocol, *The Review of Economics and Statistics*, 85: 286-297.

Besag J., 1972. On the statistical analysis of nearest-neighbour systems, *Proceedings of the European Meeting of Statisticians*, Budapest.

Besag J., 1974. Spatial interaction and the statistical analysis on lattice systems, *Journal of Royal Statistical Society*, *Series B*, 36: 192-236.

Besag J., 1975. Statistical analysis of non-lattice data, *The Statistician*, 24: 179-195.

Besag J., 1986. On the statistical analysis of dirty pictures, *Journal of the Royal Statistical Society, Series B,* 36: 192-236.

Besag, J., York J. and Mollié A., 1991. Bayesian image restoration with two applications in spatial statistics, *Ann. Inst. Statistics Mathematics*, 43: 1-21

Bivand R.S., 2002. Spatial econometrics' functions in R: classes and methods, *Journal of Geographical Systems*, 4: 405–421.

Bivand R.S. and Gebhardt A., 2000. Implementing functions for spatial statistical analysis using the R language, *Journal of Geographical Systems*, 2: 307–317.

Bivand R.S. and Portnov B.A., 2004. Exploring spatial data analysis techniques using R: the case of observations with no neighbours, in L. Anselin, R.J.G.M. Florax and S. Rey, 2004, *Advances in Spatial Econometrics*, Springer-Verlag, Berlin.

Box G.E.P. and Cox D.R., 1964. An analysis of transformation, *Journal of the Royal Statistical Society, Series B*, 26: 211-243.

Box G.E.P. and Jenkins G.M., 1970. *Time-series Analysis, Forecasting and Control*, Wiley.

Box G.E.P. and Pierce D.A., 1970. The distribution of residual autocorrelation in autoregressive integrated moving average time-series models, *Journal of the American Statistical Association,* 65: 1509-1526.

Box G.E.P. and Cox D.R., 1964. An analysis of transformation, *Journal of the Royal Statistical Society, Series B*, 26, 211-43.

Brett C. and Pinkse J., 1997. Those taxes all over the map! A test for spatial independence of municipality tax rates in British Columbia, *International Regional Science Review*, 20, 131-151.

Breusch T.S. and Pagan A.R., 1979. A simple test for heteroscedasticity and random coefficient variation, *Econometrica,* 47: 1287-1294.

Brock W.A., Dechert W.D. and Scheinkman J., 1987. A test for independence based on the correlation dimension, SSRI *Working Paper 8702,* Madison: Department of Economics, University of Wisconsin,

Brook D., 1964. On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbours systems, *Biometrika*, 51, 481-3.

Burridge P., 1980. On the Cliff-Ord test for spatial autocorrelation, *Journal of the Royal Statistical Society*, *Series B*, 42: 107-108.

Case A.C., Rosen H. and Hines J.R., 1993. Budget spillovers and fiscal policy interdependence: evidence from the States, *Journal of Public Economy,* 52: 285-307.

Chapman R.N., 1931. *Animal Ecology with Particular Reference to Insects,* McGraw-Hill, New York.

Chen X. and Conley T.G., 2001. A new semi-parametric spatial model for panel time series, *Journal of Econometrics*, 105: 59-80.

Chow G., 1960. Tests of equality between sets of coefficients in two linear regressions, *Econometrica*, 28: 591-605.

Clayton D. and Kaldor J., 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping, *Biometrics*, 43: 671-681.

Cliff A.D. and Ord J.K., 1972. Testing for spatial autocorrelation among regression residuals, *Geographical Analysis*, 4: 267-84.

Cliff A.D. and Ord J.K., 1973. *Spatial Autocorrelation*, Pion, London.

Cliff A.D. and Ord J.K., 1981. *Spatial Processes: Models and Applications*, Pion, London.

Cochrane D. and Orcutt G.H., 1949. Application of least-squares regression to relationships containing autocorrelated error terms, *Journal of the American Statistical Association*, 44: 32-61

Collett D., 1991. *Modelling Binary Data*, Chapman and Hall, New York.

Conley T.G., 1999. GMM estimation with cross-sectional dependence, *Journal of Econometrics*, 92: 1-45.

Conley T.G. and Topa G., 2002. Socio-economic distance and spatial patterns in unemployment, *Journal of Applied Econometrics*, 17: 303-327.

Consigliere I., 1981. The Chow test with serially correlated errors, *Rivista Internazionale di Scienze Sociali*, 89: 125-37.

Corsi P., Pollock R. and Prakken J., 1982. The Chow test in the presence of serially correlated errors, in Chow and Corsi (eds.), *Evaluating the Reliability of Macro-economic Models*, New York, Wiley: 163-87.

Cox D.R., 1970. *Analysis of Binary Data*, Methuen, London.

Cox D.R. and Snell E.J., 1989. *Analysis of Binary Data* (second edition), London: Chapman and Hall.

Cressie N., 1991. *Statistics for Spatial Data*, Wiley, New York.

Cross G.R. and Jain A.K., 1983. Markov random fields texture models*, IEEE Transaction on Pattern Analysis and Machine Intelligence, PAMI,* 1-5, 1: 25-39.

Christakos G., 1992. *Random Field Models in Earth Sciences,* Academic Press, San Diego.

Davidson J., 2000. *Econometric Theory*. Blackwell Publishers.

Davidson R. and MacKinnon J.G., 1993. *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

de Graaff T., Florax R.J.G.M., Nijkamp P. and Reggiani A., 2001. A general misspecification test for spatial regression models: Heteroskedasticity, dependence and non-linearity, *Journal of Regional Sciences*, 41: 255-276.

DeLong J. and Summers L., 1991. Equipment investment and economic growth, *The Quarterly Journal of Economics,* 106: 445-502.

Dobrushin R.L., 1968. Description of a random field by means of conditional probabilities and the conditions governing its regularity, *Theory of Probability and its Applications*, 13: 197-224.

Doreian P., 1980. Linear models with spatially distributed data, spatial disturbance or spatial effects?, *Sociological Methods and Research,* 9: 29-60.

Dougherty M., 2002. *Introduction to Econometrics (second edition).*

Douglas P.H., 1934. *Theory of Wages*, MacMillan, New York.

Dowde M.R. and LeSage J.P., 1997. Analysis of spatial contiguity influences on state price formation*, International Journal of Forecasting*, 13: 245-253.

Driscoll J.C. and Kraay A.C., 1998. Consistent covariance matrix estimation with spatially dependent panel data, *The Review of Economics and Statistics*, 80: 549-560.

Dubin R., 1988. Estimation of regression coefficients in the presence of spatially autocorrelated error terms, *Review of Economics and Statistics*, 70: 466-474.

Dubin R., 1992. Spatial autocorrelation and neighbourhood quality, *Regional Sciences and Urban Economics*, 22: 433-452.

Dubin R., 2003. Robustness of spatial autocorrelation specification, some Monte Carlo evidence, *Journal of Regional Sciences*, 43: 221-248.

Durbin J., 1960. Estimation of parameters in time-series regression models, *Journal of the Royal Statistical Society, Series B,* 22: 559-566.

Durbin J., 1973. Distribution theory for tests based on the sample distribution function, *S.I.A.M.*, Philadelphia.

Durbin J. and Watson G.S., 1950. Testing for serial correlation in least-squares regression I, *Biometrika,* 37: 409-428.

Durbin J. and Watson G.S., 1951. Testing for serial correlation in least-squares regression II, *Biometrika,* 38: 159-178.

Durlauf S.N. and Johnson P.A., 1995. Multiple regimes and cross-country growth behaviour, *Journal of Applied Econometrics,* 10: 365-384.

Durlauf S.N. and Quah D., 1999. The new empirics of economic growth, in: J. Taylor and M. Woodford (eds), *Handbook of Macroeconomics,* Elsevier Science, North Holland, New York and Oxford: 235-308.

Edgard G.A., 1995. *Measure, Topology and Fractal Geometry*, Springer-Verlag, Berlin.

Elhorst J.P., 2001. Dynamic models in space and time, *Geographical Analysis*, 33: 119-140.

Elhorst J.P., 2003. Specification and estimation of spatial panel data models, *International Regional Science Review*, 26: 244-268.

Engle R.F., Hendry D.F and Richard J.-F., 1983. Exogeneity, *Econometrica*, 51: 277-304.

European Union. 2005. Third progress report on cohesion: towards a new partnership for growth jobs and cohesion, *Communication from the Commission of the European Community,* SEC (2005), 632, May, Luxemburg.

Ertur C. and Le Gallo J., 2003. An exploratory spatial-data analysis of European regional disparities, 1980-1995, in B. Fingleton (ed), *European Regional Growth*, Sprinter, Berlin.

Eurostat, 2002. *European Regional Statistics, Reference Guide,* European Community, Luxemburg.

Evans M.A. and King M.L., 1988. A further class of tests for heteroskedasticity, *Journal of Econometrics,* 27: 163-178.

Evans P. and Karras G., 1996a. Do economies converge? Evidence from a panel of U.S. states, *Review of Economics and Statistics*, : 384-388.

Evans P. and Karras G., 1996b. Convergence Revisited, *Journal of Monetary Economics***,** 37: 249-265.

Fabiani S. and Pellegrini G., 1997. Education, infrastructure, geography and growth: an empirical analysis of the development of Italian provinces, *Temi di discussione,* Bank of Italy, Rome.

Feller W., 1948. On Kolmogorov-Smirnov limit theorems for empirical distributions, *Annals of Mathematical Statistics*, 30: 381.

Fingleton B., 1999. Spurious spatial regression: some Monte Carlo results with a spatial unit root and spatial cointegration, *Journal of Regional Sciences*, 39: 1-19.

Fingleton B., 2003. *European Regional Growth*, Springer-Verlag , Berlin.

Fingleton B., 2004. Regional economic growth and convergence: insights from a spatial econometric perspective, in L. Anselin, R.J.G.M. Florax and S. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, New York.

Fleissig A. and Strauss J., 2001. Panel unit root tests of OECD stochastic convergence, *Review of International Economics*, 9: 153-162.

Fleming M.M., 2004. Techniques for estimating spatially dependent discrete choice models, in L. Anselin, R.J.G.M. Florax and S. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, New York: 145-168.

Florax R.J.G.M. and de Graaff T., 2004. The performance of diagnostic tests for spatial dependence in linear-regression models: A meta-analysis of simulation studies, in L. Anselin, R.J.G.M. Florax and S. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, New York.

Florax R.J.G.M and Rey S., 1995. The impact of misspecified spatial structure in linear-regression models, in L. Anselin, and R.J.G.M. Florax (eds.), *New Directions in Spatial Econometrics*, Springer-Verlag, New York: 111-135.

Florax R.J.G.M and van de Vlist A., 2003. Spatial econometric data analysis: moving beyond traditional models, *International Regional Science Review*, 26: 233-243.

Florax R.J.G.M, De Dominicis L. and De Graaff T., 2004. Spatial ARCH processes: an LM-test, experimental simulation results and empirical illustrations. Paper presented at the 50th North American *Conference of the Regional Science Association International* (RSAI), Philadelphia, November, 2003.

Florax R.J.G.M., Folmer H. and Rey S.J., 1998. The relevance of Hendry's methodology: experimental simulation results for linear spatial models, *Working Paper* 98, 125/4 Tinberger Institute, Amsterdam, The Netherlands.

Fujita M., Krugman P. and Venables A.J., 1999. *The Spatial Economy: Cities, Regions and International Trade*, MIT Press, New York.

Galambos J., 1988. Truncation methods in probability theory, in S. Kotz and N.L. Johnson (eds.), *Encyclopedia of Statistical Sciences*, Wiley, New York: 355-357.

Galor O., 1996. Convergence ? Inference from theoretical models*, Economic Journal*, 106: 1056-1069.

Gandolfo G., 1981. *Qualitative Analysis and Econometric Estimation of Continuous- time Dynamic Models*, North Holland.

Garrett T.A., Wagner G.A. and Wheelock D.C., 2003. A spatial probit analysis of state banking regulation, *Working Paper 044A*, Federal Reserve Bank of St. Louis, St. Louis, MO.

Geman S. and Graffigne C., 1987. Markov random fields and their applications to computer vision, in A. Gleason (ed), *Proceedings of the International Congress of Mathematics*, American Mathematical Society, Providence.

Gentle J.E., 1998. Cholesky Factorization. §3.2.2 in *Numerical Linear Algebra for Applications in Statistics.* Springer-Verlag, Berlin: 93-95.

Getis A. and Aldstadt J., 2004. On the specification of the spatial weights matrix. *Geographical Analysis*, 35.

Getis A. and Griffith D.A., 2002. Comparative spatial filtering in regression analysis, *Geographical Analysis*, 34: 130-140.

Getis A. and Ord J.K., 1992. The analysis of spatial association by distance statistics, *Geographical Analysis*, 24: 189-206.

Giacomini R. and Granger C.W.J., 2003. Aggregation of space-time processes, *Journal of Econometrics*.

Glejser H., 1969. A new test for heteroscedasticity, *Journal of the American Statistical Association*, 64: 316-323.

Godfrey L.G., 1978. Testing for multiplicative heteroscedasticity, *Journal of Econometrics*, 8: 227-236.

Goldberg A.S., 1998. *Introductory Econometrics* , Harvard edition world.

Goldfeld S.M. and Quandt R.E., 1965. Some tests for homoscedasticity, *Journal of the American Statistical Association*, 60, 539-547.

Gourieroux C. and Montfort A., 1995. *Statistics and Econometric Models,* Vols. 1 and 2, Cambridge University Press.

Granger C.W.J., 1969. Spatial data and time-series analysis, in A. Scott (ed), *Studies in Regional Science*, Pion, London.

Granger C.W.J., 1974. Aspects of the analysis and interpretation of temporal and spatial data, *The Statistician*, 24: 197-210.

Greene W.H., 2003. *Econometric Analysis (fifth edition)*, New York, Macmillan.

Gress B., 2003. Semi-parametric spatial auto-covariance models, *PhD Thesis*.

Griffith D.A., 1988. *Advanced Spatial Statistics*, Kluwer Academic Publishers, Dordrecht.

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., 2003. *Spatial autocorrelation and spatial filtering*, Springer-Verlag, Berlin.

Griffith D.A. 2004. Extreme eigenfunctions of adjacency matrices for planar graphs employed in spatial analyses, *Linear Algebra and its Applications*, 388: 201-219.

Griffith D.A. and Tiefelsdorf M., 2002. Semi-parametric filtering of spatial autocorrelation: The eigenvector approach. Paper presented at the *North American Meetings of the Regional Sciences Association International*, November, San Juan, Puerto Rico.

Grimmett G.R., 1973. A theorem about random fields, *Bulletin of the London Mathematical Society,* 79: 5.

Grimmett G.R. and Stirzaker D.R., 1994. *Probability and Random Processes*. Clarendon Press, Oxford, United Kingdom.

Gujarati D., 2003. *Basic Econometrics*, (fourth edition), McGraw-Hill.

Guyon X., 1995. *Random Fields on a Network*, Springer-Verlag, Berlin.

Haining R.P., 1978. The moving average model for spatial interaction, *Transaction Institute of British Geographers,* N.S. 3: 202-225.

Haining R.P., 1990. *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge.

Haining R.P., 2003. *Spatial Data Analysis*, Cambridge University Press, Cambridge.

Hamilton J.D., 1994. Time-series analysis, Princeton University Press, Princeton, New Jersey.

Hammersey J.M. and Clifford P., 1971. Markov fields on finite graphs lattices, *unpublished manuscript*.

Hansen L.P., 1982. Large sample properties of generalized method-of-moments estimation, *Econometrica*, 1029-1054.

Harrison M.J. and McCabe B.P., 1979. A test for heteroscedasticity based on Ordinary Least Squares residuals, *Journal of the American Statistical Association*, 74: 494-499.

Harvey A. and Carvalho V., 2002. Models for converging economies, *Cambridge Working Paper*, March.

Hausdorff F., 1914. *Gundzüge der Mengenlehre*, Von Vent, Leipzig, Germany. Reprinted as *Set theory,* (second edition), Chelsea, New York, 1962.

Hayashi F., 2000. *Econometrics*, Princeton University Press.

Heijmans R. and Magnus J., 1986. Consistent maximum likelihood estimation with dependent observations: the general (non-normal) case and the normal case, *Journal of Econometrics*, 32: 253-285.

Hendry D.F. and Morgan M.S., 1995. *The Foundations of Econometric Analysis*. Cambridge University Press, Cambridge.

Hepple L.W., 2003. Bayesian model choice in spatial econometrics, *Paper for LSU, Spatial Econometrics Conference*, Baton Rouge, November.

Holloway G., Shankar B. and Rahman S., 2002. Bayesian spatial probit estimation: a primer and an application to HYV rice adoption, *Agricultural Economics*, 27: 383-402.

Holly S. and Weale M., 2000. *Econometric Modelling*, Cambridge University Press, Cambridge.

Hordijk L., 1974. Spatial correlation in the disturbance of linear interregional models, *Regional Science and Urban Economics*, 4: 117-40.

Hsiao C., 1986. *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Huang J.S., 1984. The autoregressive moving average model for spatial analysis, *Australian Journal of Statistics,* 26: 169-178.

Isard W., 1960. *Methods of Regional Analysis*, Technology press of MIT and Wiley, New York.

Islam N., 1995. Growth empirics: a panel data approach, *Quarterly Journal of Economics*, 110: 1127-1170.

Islam N., 1998. Growth empirics: a panel data approach. A reply, *Quarterly Journal of Economics*, 113(1): 325-329.

Ising E., 1925. Beitray sur theorie des ferromagnetismus, *Zeitschrift Physic*, 31: 253-258.

Jarque C.M. and Bera A.K., 1980. Efficient tests for normality, heteroscedasticity and serial independence of regression residuals, *Economic Letters*, 6: 255-59.

Joe H., 1997. *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.

Johnston J., 1991. *Econometric Methods*, McGraw Hill, New York.

Johnston J. and Dinardo N., 1997. *Econometric Methods*, (fourth edition), McGraw-Hill, New York.

Judge G.G., Griffiths W.E, Hill R.E., Luetkepohl H. and Lee T-C., 1988. *Introduction to the Theory and Practice of Econometrics,* John Wiley and Sons, New York.

Kaiser M.S. and Cressie N., 1997. Modelling Poisson variables with positive spatial dependence, *Statistics and Probability Letters*, 35: 423-432.

Kaldor N., 1957. A model of economic growth, *Economic Journal*, 67: 591-624

Kaldor N., 1970. The case for regional policies, *Scottish Journal of Political Economy*, 17: 37-48.

Kaluzny S., Vega S., Cardoso T. and Shelly A., 1997. (*S+)SpatialStats User's Manual*, Springer-Verlag, New York, NY.

Kelejian H.H. and Prucha L., 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 L., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review*, 40: 509-533.

Kelejian H.H. and Prucha L., 2001. On the asymptotic distribution of the Moran I test statistic with applications, *Journal of Econometrics*, 104: 219-257.

Kelejian H.H. and Prucha L., 2003. Estimation of simultaneous systems of spatially interrelated cross-sectional equations, *Journal of Econometrics*, 118: 27-50.

Kelejian H.H. and Robinson D.P., 1992. Spatial autocorrelation: a new computationally simple test with an application to per capita county policy expenditures, *Regional Science and Urban Economics*, 22: 317-331.

Kelejian H.H. and Robinson D.P., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model, *Papers in Regional Science*, 72: 297-312.

Keleijan H.H. and Robinson D.P., 1995. Spatial correlation: a suggested alternative to the autoregressive model, in L. Anselin and R.J.G.M. Florax (eds.), *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin: 5-95.

Keleijan H.H. and Robinson D.P., 1997. Infrastructure productivity estimation and its underlying econometric specifications: a sensitivity analysis, *Papers in Regional Science*, 76: 115-131.

Kelejian H.H. and Robinson D.P., 2004. The influence of spatially correlated heteroscedasticity on tests for spatial correlation, in L. Anselin, R.J.G.M. Florax and S. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, New York.

Kendall M. and Stuart A., 1979. *The Advanced Theory of Statistics*, Vol. 2 (fourth edition), Charles Griffin & Co, London.

Kennedy P., 2003. *A Guide to Econometrics,* (fifth edition), Blackwell Publishers.

Khinchin A., 1949. *Mathematical Foundations of Statistical Mechanics*, Dover, New York.

Kim H., Sun D. and Tsutaka R.K., 2001. A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model, *Journal of the American Statistical association*, 96: 1506-1521.

Klaassen L.H., Paelinck J.H.P. and Wagenaar S., 1979. *Spatial Systems,* Saxon House, Farnborough.

Klevmarken N.A., 1989. Panel studies: what can we learn from them? An introduction, *European Economic Review*, 33: 523-529.

Kmenta J., 1997. *Elements of Econometrics*, (second edition), Macmillan, New York.

Kobayashi M., 1991. Testing for autocorrelated disturbances in non-linear regression analysis, *Econometria,* 59: 1153-1159.

Kolmogorov A., 1933. Sulla determinazione empirica di una legge di distribuzione, *Giornale dell'Istituto Italiano Attuari*, 4: 83.

Krugman P., 1991. *Geography and Trade*, Leuven University Press, Leuven and MIT Press, Cambridge, Massachussets.

Krugman P. and Venables A.J., 1995. Globalization and the inequality of nations, *Quarterly Journal of Economics,* 11:857-880.

Künsch H., 1981. Thermodynamics and statistical analysis of Gaussian random fields, *Whar. Ver. Geb.,* 58: 401-421.

Land K. and Leane G., 1992. On large-sample estimation of regression models with spatial network effects terms: a two-stage least-squares approach, in P. Marsden (ed.), *Sociological Methodology*, Jossey–Bass, San Francisco: 221-248.

Lee L.F., 2002. Consistency and efficiency of least-squares estimation for mixed regressive spatial autoregressive models, *Econometric Theory*, 18: 252-277.

LeSage J., 1997. Bayesian estimation of spatial autoregressive models, *International Regional Science Review*, 20: 113-129.

LeSage J., 1999. *Spatial Econometrics*: *The Web Book of Regional Science*, Regional Research Institute, West Virginia University, Morgantown, WV.

LeSage J., 2000. Bayesian estimation of limited dependent variable spatial autoregressive models, *Geographical Analysis*, 32: 19-35.

LeSage J., 2004. A family of geographically weighted regression models, in L. Anselin, R.J.G.M. Florax and S. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, New York.

LeSage J. and Krivelyova A., 1999. A spatial prior for Bayesian vector autoregressive models, *Journal of Regional Science*, 39: 297-317.

Ljung G.M. and Box G.E.P., 1978. On a measure of lack of fit in time-series models, *Biometrika,* 65: 297-303.

Lotka A.J., 1956. *Elements of Mathematical Biology,* Dover, New York (reprint of the original 1924 book).

Maddala G., 2001. *Econometrics,* McGraw-Hill, New York.

Mankiw N.G., 1995. The growth of nations, *Brooking Papers on Economic Activity*, 1: 275-326.

Mankiw N.G., Romer D. and Weil D., 1992. A contribution to the empirics of economic growth, *Quarterly Journal of Economics*, 107: 407-437.

Mardia K.V., 1970. Measures of multivariate skewness and kurtosis with applications, *Biometrika*, 57: 519-530.

Mardia K.V., 1974. Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies, *Sankhya*, B36: 115-128.

Mardia K.V., 1980. Tests of univariate and multivariate normality, in P.R. Krishnaiah (ed.), *Handbook of Statistics*, Vol.1, North Holland, New York: 279-320.

Mardia K.V., 1986 Mardia's test of multinormality, in N.L. Johnson, S. Kotz and C.Read (eds.), *Encyclopaedia of Statistics*, 5, Wiley, New York: 217-221.

Mardia K.V., 1988. Multi-dimensional multivariate Gaussian Markov random fields with application to image processing*, Journal of Multivariate Analysis*, 24: 265-284.

Mardia K.V., 1990. Maximum likelihood estimation for spatial models, in *Spatial Statistics: Past, Present and Future*, D.A. Griffith (ed.), Monograph N. 12, Institute of Mathematical Geography.

Mardia K.V. and Watkins A.J., 1989. On multimodality of the likelihood for the spatial linear model, *Biometrika*, 76: 289-295.

Matern B., 1960. *Spatial Variation*, Meddelanden Fråns Statens Skogsforskninstitut, 49: 1-144.

Matern B., 1986. *Spatial Variation*, (second edition), Springer-Verlag, Berlin.

McCullagh P. and Nedler J.A., 1989. *Generalized Linear Models*, Chapman and Hall, London.

McLeish D., 1975. A maximal inequality and dependent strong law, *The Annals of Probability*, 3: 827-39.

McMillen D.P. 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*, Springer-Verlag, New York.

Mead R., 1967. A mathematical model for the estimation of interplant competition, *Biometrics*, 30: 295-307.

Molliè A., 1996. Bayesian mapping of diseases, in W. Gilks, S. Richardson and S.D.J. Spiegelhalter (eds.), *Markov Chain Monte Carlo in Practice*, Chapman and Hall.

Mood A.M., Graybill F.A. and Boes D.C., 1974. *Introduction to the Theory of Statistics*, McGraw-Hill, New York.

Moran P.A.P., 1950. The interpretation of statistical maps, *Journal of the Royal Statistical Society, Series B*, 10: 243-51.

Mur J. and Trìvez F.J., 2003. Unit roots and deterministic trends in spatial econometric models, *International Regional Science Review*, 26: 289-312.

Murdoch J.C., Sandler T. and Sargent K., 1997. A tale of two collectives: Sulphur versus nitrogen oxide emission reduction in Europe, *Public Finance Quarterly*, 21: 334-350.

Murdoch J.C. Sandler T. and Vijverberg W.P., 2003. The participation decision versus the level of participation in an environmental strategy: a spatial probit analysis, *Journal of Public Economy*, 87: 337-362.

Newey W.K. and Powell J.L., 1987. Symmetric least-squares estimation and testing, *Econometrica,* 55: 819-847.

Newey W. and West K., 1987. A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica*, 55: 703-708.

Oberhofer J. and Kmenta J., 1974. A general procedure for obtaining maximum likelihood estimates in generalised regression models, *Econometrica*, 42: 579-90.

Olmo J.C., 1995. Spatial estimation of housing prices and locational rents*, Urban Studies*, 32: 1331-1344.

Openshaw S. and Taylor P.J., 1981. The modifiable areal unit problem, *Catmog*, no. 38.

Ord J.K., 1975. Estimation methods for models of spatial interaction, *Journal of the American Statistical Association*, 70: 120-126.

Ord J.K. and Getis A., 1995. Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis,* 27, 4: 286-306.

Ottaviano G.M.I.P. and Puga D., 1998. Agglomeration in the global economy: a survey of the 'new economic geography', *World Economy*, 21 (6): 707-731.

Pace L. and Salvan A., 1997. *Principles of Statistical Inference*, World Scientific.

Pace R.K., 1997. Performing large spatial regressions and autoregressions, *Economic Letters*, 54, 283-291.

Pace R.K. and Barry R., 1997a. Fast CarS, *Journal of Computation and Simulation*, 59: 123-147.

Pace R.K. and Barry R., 1997b. Quick computation of spatial autoregressive estimation, *Geographical Analysis*, 29: 232-246.

Pace R.K. and Barry R., 1997c. Sparse spatial autoregressions, *Statistics and Probability Letters*, 33: 291-297.

Pace R.K. and Barry R., 1998. *Spatial Statistics Toolbox 1.0*, Real Estate Research Institute, Louisiana State University, Baton Rouge, LA.

Pace R.K. and LeSage J.P., 2002. Semi-parametric maximum likelihood estimates of spatial dependence, *Geographical Analysis*, 34: 76-90.

Pace R.K. and LeSage J.P., 2004. Chebyshev approximation of log-determinants of spatial weights matrices, *Computational Statistics and Data Analysis*, 45: 179-196.

Pace R.K. and Zou D., 2000. Closed-form maximum likelihood estimates of nearest-neighbour spatial dependence, *Geographical Analysis*, 32: 154-172.

Pace R.K., Barry R., Clapp J.M. and Rodiquez M.,1998. Spatio-temporal autoregressive models of neighbourhood effects, *Journal of Real Estate Finance and Economics*, 17: 15-33.

Pace R.K., Barry R., Slawson V.C. and Sirman C.F., 2004. Simultaneous spatial and functional form transformations, in L.Anselin, R.J.G.M. Florax and S.J. Rey (eds), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, Berlin.

Paelinck J.H.P., 1967. *L'efficacité de la politique économique régionale,* Namur, Faculté des Sciences Economiques.

Paelinck J.H.P., 1983 *Formal Spatial Economics,* Gower.

Paelinck J.H.P., 1996. Studies in spatial econometrics, *Research Papers, George Mason University*, The Institute of Public Policies, Center for Regional Analysis, Fairfax, VA, USA.

Paelinck J.H.P. and Klaassen L.H., 1979. *Spatial Econometrics*, Gower, Westmead, Farnborough.

Paelinck J.H.P. and Nijkamp P., 1975. *Operational Theory and Methods in Spatial Economics,* Saxon House and Lexington, Farnborough and Massachusetts.

Pesaran M.H., 2004a. General diagnostic tests for cross-sectional dependence in panel data, *Cambridge University Working Paper*, June.

Pesaran M.H., 2004b. A pair-wise approach to testing for output and growth convergence, *Mimeo, Cambridge University*, August.

Pfeiffer P.E. and Deutsch J.S., 1980. A STARIMA model-building procedure with application to description and regional forecasting, *Transaction of the Institute of British Geographers*, 70: 330-349.

Pinkse J., 1998. A consistent non-parametric test for series independence, *Journal of Econometrics*, 84: 205-231.

Pinkse J., 1999. Asymptotics of the Moran test and a test for spatial correlation in probit models, *Department of Economics Working Paper*, University of British Columbia, Vancouver, BC.

Pinkse J., 2004. Moran-flavoured tests with nuisance parameters: examples in L. Anselin, R.J.G.M. Florax and S. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, New York.

Pinkse J. and Slade M.E., 1998. Contacting in space: an application of spatial statistics to discrete-choice models, *Journal of Econometrics*, 85: 125-154.

Pinkse J., Shen L. and Slade M., 2003. Dynamic spatial probit with fixed effects using one-step GMM: an application to mine operating decisions. Under submission to the *Journal of Econometrics*.

Pinkse J., Slade M.E. and Brett C., 2002. Spatial price competition: a semiparametric approach, *Econometrica*, 70: 1111-1153.

Pirandello L., 1921. *Sei personaggi in cerca d'autore*, R. & F. Bemporad, Florence.

Pisati M., 2001. Tools for spatial data analysis, *Stata Technical Bulletin*, 60: 21–37.

Postiglione P., Arbia G. and Salvatore M., 2002. Empirics for regional convergence: the role of spatial dependence. *Paper presented at the national meeting of the Italian Statistical Society*, Milan, June.

Press, W.H., Flannery B.P., Teukolsky S.A. and Vetterling, W.T., 1992. LU Decomposition and Its Applications. §2.3 in *Numerical Recipes in FORTRAN: The Art of Scientific Computing, (second edition),* Cambridge University Press, Cambridge, England: 34-42.

Preston C.J., 1973. Generalized Gibbs states and Markov random fields, *Advances in Applied Probability,* 5: 242-261.

Puga D. and Venables A.J., 1997. Preferential trading and national location, *Journal of International Economics,* 43 (3-4): 347-368.

Puga D. and Venables A.J., 1999. Agglomeration and economic development, *Economic Journal*, 109 (455): 292-311.

Quah D., 1993a. Empirical cross-section dynamics in economic growth, *European Economic Review,* 37:426-434.

Quah D., 1993b. Galton's fallacy and test of the convergence hypothesis, *Scandinavian Journal of Economics,* 95: 427-443.

Quah D., 1996a. Twin peaks: growth and convergence in models of distributed dynamics, *Economic Journal,* 106: 1045-1055.

Quah D., 1996b. Regional convergence clusters across Europe, *European Economic Review,* 40: 951-958.

Quah D., 1997. Empirics for growth and distribution: stratification, polarization and convergence clubs, *Journal of Economic Growth,* 2: 27-59.

Quandt R., 1958. The estimation of parameters of a linear-regression system obeying two separate regimes, *Journal of the American Statistical Association*, 53: 873-880.

Rao C.R., 1948. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Proceedings of the Cambridge Philosophical Society,* 44: 50-57.

Rey S.J., 2001. Spatial empirics for economic growth and convergence, *Geographical Analysis*, 33: 195-214.

Rey S.J., 2003. Spatial analysis of regional economic growth, inequality and change, in M. Goodchild and D. Janelle (eds.), *Spatially Integrated Social Science*, Oxford University Press, New York, NY.

Rey S.J. and Montouri B.D., 1998. U.S. regional income convergence: a spatial econometric perspective, *Regional Studies,* 33: 143-156.

Ripley B.D., 1981. *Spatial Statistics*, Wiley, New York.

Ripley B.D., 1988. *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge.

Ripley B.D., 1990. Gibbsian interaction models, in D. Griffith (ed), *Spatial Statistics: Past, Present and Future*, Monograph no. 12, Institute of Mathematical Geography.

Ruud P.A., 2000. *An Introduction to Classical Econometric Theory*, Oxford University Press.

Saavedra L.A., 2003. Test for spatial-lag dependence based on method-of-moments estimation, *Regional Science and Urban Economics*, 33: 27-58.

Sala-i-Martin X., 1996. The classical approach to convergence analysis, *Economic Journal*, 106, 437: 1019-1036.

Samuelson P., 1971. Generalized predator-prey oscillations in ecological and economic equilibrium, *Proceedings of the National Academy of Sciences*, 68: 980-3.

Sargan J.D., 1964. Wages and prices in the UK: a study in econometric methodology, in P.E. Hart, G. Mills and J.K. Whitaker (eds), *Economic Analysis for (the) National Economic Planning*, Butterworths, London. Reprinted in K.F. Wallis and D.F. Hendry (eds), *Quantitative Economics and Econometric Analysis*, 1984, Basil Blackwell, Oxford.

Seddighi H.R., Lawler K.A. and Katos A.V., 2000. *Econometrics*: *A Practical Approach*, Routledge, Taylor & Francis Group.

Serfling R., 1980. *Approximation Theorems of Mathematical Statistics*, New York, John Wiley.

Shapiro S.S. and Wilk M.B., 1965. An analysis of variance tests for normality (complete samples), *Biometrika*, 52: 591.

Silvey S.D., 1959. The Lagrangian multiplier test, *Annals of Mathematical Statistics,* 30, 389-407.

Smirnov O., 2003. Computation of the information matrix for models of spatial interaction. *Technical Report*, Regional Economics Applications Laboratory (REAL), University of Illinois, Urbana-Champaigne, IL.

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: 301-319.

Smith T., 1980. A central limit theorem for spatial samples, *Geographical Analysis*, 12: 299-324.

Solon G.S., 1989. The value of panel data in economic research, in D. Kasprzyk, G.J. Duncan, G. Calton and M.P. Singh (eds), *Panel Surveys*, John Wiley, New York: 486-496.

Solow R.M., 1956. A contribution to the theory of economic growth, *Quarterly Journal of Economics,* 70, 1: 65-94.

Spanos A., 1986. *Statistical Foundations of Econometric Modelling*, Cambridge University Press, Cambridge.

Stewart J. and Gill M., 1998. *Econometrics, (second edition)*, Prentice-Hall.

Stock J.H. and Watson M.W., 2003 *Introduction to Econometrics,* Addison Wesley.

Strauss, D.J. 1977. Clustering of coloured lattices, *Journal of Applied Probability*, 14: 135-43.

Strauss D.J. and Ikeda M., 1990. Pseudo-likelihood estimation for social networks, *Journal of the American Statistical Association*, 85: 204-212.

Swan T.W., 1956. Economic growth and capital accumulation, *Economic Records,* 32: 334-361.

Szroeter J., 1978. A class of parametric tests for heteroscedasticity in linear econometric models, *Econometrica*, 46: 1311-1327.

Temple J., 1999. The new growth evidence, *Journal of Economic Literature,* 37, 112-156.

Theil H., 1954. *Linear Aggregation in Economic Relations*, North-Holland Publ. Company, Amsterdam.

Thomas L., 1997. *Modern Econometrics: An Introduction*, Pearson Education.

Tiefelsdorf M., 2002. The saddlepoint approximation of Moran's *I*'s and local Moran's I_i's reference distributions and their numerical evaluation, *Geographical Analysis*, 34: 187-206.

Tiefelsdorf M. and Boots B.N., 1995. The exact distribution of Moran's *I*. *Environment and Planning A*, 27: 985-99.

Tobler W., 1970. A computer movie simulating urban growth in the Detroit region, *Economic Geography Supplement*, 46: 234-40.

Tuckey J.W., 1977. *Exploratory Data Analysis*, Reading, MA, Addison-Wesley.

Upton G. and Fingleton B., 1985. *Spatial Data Analysis by Example: Point Pattern and Quantitative Data*, Volume 1, Wiley, New York.

Upton G. and Fingleton B., 1989. *Spatial Data Analysis by Example: Categorical and Directional Data*, Volume 2, Wiley, New York.

Verbeek M., 2000. *A Guide to Modern Econometrics*. John Wiley and Sons, New York.

Verdoorn P., 1949. Fattori che regolano lo sviluppo della produttività del lavoro, *L'industria*, 1: 3-10.

Vijverberg W.P., 1997. Monte Carlo evaluation of multivariate normal probabilities, *Journal of Econometrics,* 76: 281-307.

Wall M.M., 2004. A close look at the spatial structure implied by the CAR and SAR models, *Journal of Statistical Planning and Inference*, 121: 311-324.

Wallis K.F., 1972. Testing for fourth-order autocorrelation in quarterly regression equations, *Econometrics,* 40: 617-636.

White H., 1980. A heteroscedastic-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, 48: 817-38.

White H., 1984. *Asymptotic Theory for Econometricians*, Academic Press, Orlando.

White H. and Domowitz I., 1984. Non-linear regression with dependent observations, *Econometrica*, 52: 143-161.

Whittle P.J., 1954. On stationary processes in the plane, *Biometrika*, 41: 434-49.

Whittle P.J., 1962. Topographic correlation, power-law covariance functions and diffusion, *Biometrika*, 49: 305-314.

Whittle P.J., 1963. Stochastic processes in several dimensions, *Bulletin of the International Statistical Institute,* 40(2): 975-994.

Woolridge J.M., 2002a. *Econometrics: A Modern Approach (second edition).*

Woolridge J.M., 2002b. *Econometric Analysis of Cross-section and Panel Data,* the MIT Press, Cambridge, Massachusetts.

Woolridge J.M and White H., 1988. Some invariance principles and central limit theorems for dependent heterogeneous processes, *Econometric Theory*, 4: 210-230.

Yaglom A.M., 1957. Some classes of random fields in n-dimensional space, related to stationary random processes, *Theor. Probability Appl*., 2: 273-320.

Yaglom A.M., 1961. Second-order homogeneous random fields, *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, University of California Press, Berkeley, 2: 593-622.

Yaglom A.M. 1962. *An Introduction to the Theory of Stationary Random Functions*, Prenctice-Hall, Englewood Cliffs, New Jersey.

Yule U. and Kendall M.S., 1950. *An Introduction to the Theory of Statistics*, Charles Griffin, London.

List of Tables

List of Figures

Name Index

D

E

F

G

H

I

J

K

W

Y

Z

Subject Index

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

Printing and Binding: Strauss GmbH, Mörlenbach