SPRINGER BRIEFS IN REGIONAL SCIENCE

J. Paul Elhorst

Spatial Econometrics

From Cross-Sectional Data to Spatial Panels

Springer

SpringerBriefs in Regional Science

For further volumes:
http://www.springer.com/series/10096

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact, authored volumes of 50 to 125 pages, the series covers a range of content from professional to academic.

SpringerBriefs in Regional Science showcase emerging theory, empirical research and practical application, lecture notes and reviews in spatial and regional science from a global author community.

J. Paul Elhorst

Spatial Econometrics

From Cross-Sectional Data to Spatial Panels

Springer

J. Paul Elhorst
Faculty of Economics and Business
University of Groningen
Groningen
The Netherlands

ISSN 2192-0427 ISSN 2192-0435 (electronic)
ISBN 978-3-642-40339-2 ISBN 978-3-642-40340-8 (eBook)
DOI 10.1007/978-3-642-40340-8
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013946223

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Abstract

This book provides an overview of three generations of spatial econometric models: models based on cross-sectional data, static models based on spatial panels and dynamic spatial panel data models. The book not only presents different model specifications and their corresponding estimators, but also critically discusses the purposes for which these models can be used and how their results should be interpreted. Special attention is paid to the interpretation of spatial spillover effects. Several of these models are illustrated using well-known datasets. Furthermore, Matlab routines are provided with which the results reported in the book can be replicated and with which researchers can run their own empirical problems.

Keywords Spatial panels · Models · Dynamic effects · Spatial spillover effects · Estimation methods

Chapter 1
Introduction

Abstract This chapter introduces the topics the book will be dealing with.

Keywords Spatial dependence · Cross-sectional data · Spatial panels

1.1 Introduction

Spatial econometrics is a subfield of econometrics dealing with spatial interaction effects among geographical units. Units could be zip codes, cities, municipalities, regions, counties, states, jurisdictions, countries, and so forth depending on the nature of the study. Spatial econometric models can also be used to explain the behavior of economic agents other than geographical units, such as individuals, firms, or governments, if they are related to each other through networks, but this type of research, although growing, is less common.

Whereas the time-series literature focuses on the dependence among observations over time and uses the symbol "$t-1$" to denote variables lagged in time, the spatial econometrics literature is interested in the dependence among observations across space and uses the so-called spatial weights matrix $\boldsymbol{W}$ to describe the spatial arrangement of the geographical units in the sample. It should be stressed here that spatial econometrics is not a straightforward extension of time-series econometrics to two dimensions. One obvious difference is that two geographical units can affect each other mutually, whereas two observations in time cannot. According to Getis (2007), another complicating factor is the wide variety of units of measurement that are eligible for modeling spatial dependence (neighbors, distance, links, etc.) as compared to measuring temporal dependence (time).

In the last decade, the spatial econometrics literature has exhibited a growing interest in the specification and estimation of econometric relationships based on spatial panels. Spatial panels typically refer to data containing time series observations of a number of geographical units. This interest can be explained by the fact that panel data offer researchers extended modeling possibilities to explain

J. P. Elhorst, *Spatial Econometrics*, SpringerBriefs in Regional Science,
DOI: 10.1007/978-3-642-40340-8_1,

causal relationships as compared to the single equation cross-sectional setting, which was the primary focus of the spatial econometrics literature for a long time. Panel data are generally more informative, and they contain more variation and often less collinearity among the variables. The use of panel data results in a greater availability of degrees of freedom, and hence increases efficiency in the estimation. Panel data also allow for the specification of more complicated behavioral hypotheses, including effects that cannot be addressed using pure cross-sectional data (see Sect. 3.2 and Hsiao 2007 for more details).

This book provides an overview of three generations of spatial econometric models. The first generation consists of models based on cross-sectional data. Key contributions in this field are Anselin (1988, 2006), Griffith (1988), Haining (1990), Cressie (1993), Anselin and Bera (1998), Arbia (2006), and LeSage and Pace (2009). These models are discussed in Chap. 2. Since many issues have already been covered in these previous contributions, this chapter mainly focuses on insights that are relatively new and on issues that have led to discussion or confusion.

The second generation comprises non-dynamic models based on spatial panel data. These models might just pool time-series cross-sectional data, but more often they control for fixed or random spatial and/or time-period specific effects. A limited number of studies consider models with one equation for every unit in the sample, where the slope coefficients of the explanatory variables might again be assumed fixed or random. The multi-level model with both fixed and random coefficients can also be classified to this class of models. Other studies consider a set of equations, one for every time period or one for a set of multiple dependent variables, extending each equation to include spatial interaction effects, known as spatial SUR models. This second generation of models is extensively discussed in Chap. 3.

The third generation of spatial econometric models encompasses dynamic spatial panel data models. At the beginning of this century, there was no straightforward estimation method for this type of models. This was because methods developed for dynamic but non-spatial and for spatial but non-dynamic panel data models produced biased estimators when these methods/models were put together. Chapter. 4 provides an overview of the main methodological studies that have attempted to solve this shortcoming.

Since this book is partly written from a practitioner's point of view, it not only presents different model specifications but also draws attention to and sometimes critically discusses the purposes for which these specifications can be used. This is important because the spatial econometric literature has also been criticized. Recent examples are Partridge et al. (2012), Gibbons and Overman (2012), McMillen (2012) and Corrado and Fingleton (2012) in a special issue of the Journal of Regional Science. By illustrating and comparing the results of different model specifications, it is shown that these critical studies have a point; some models are indeed more promising than others.

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Anselin L (2006) Spatial econometrics. In: Mills TC, Patterson K (eds) Palgrave handbook of econometrics, vol 1. Basingstoke, Palgrave, pp 901–969

Anselin L, Bera A (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles D (eds) Handbook of applied economics statistics. Marcel Dekker, New York, pp 237–289

Arbia G (2006) Spatial econometrics: statistical foundations and applications to regional convergence. Springer, Berlin

Corrado L, Fingleton B (2012) Where is the economics in spatial econometrics? J Reg Sci 52(2):210–239

Cressie NAC (1993) Statistics for spatial data. Wiley, New York

Gibbons S, Overman HG (2012) Mostly pointless spatial econometrics? J Reg Sci 52(2):172–191

Getis A (2007) Reflections on spatial autocorrelation. Reg Sci Urban Econ 37:491–496

Griffith DA (1988) Advanced spatial statistics. Kluwer, Dordrecht

Haining R (1990) Spatial data analysis in the social and environmental sciences. Cambridge University Press, Cambridge

Hsiao C (2007) Panel data analysis—advantages and challenges. Test 16:1–22

LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press Taylor & Francis Group, Boca Raton

McMillen DP (2012) Perspectives on spatial econometrics: Linear smoothing with structured models. J Reg Sci 52(2):192–209

Partridge MD, Boarnet M, Brakman S, Ottaviano G (2012) Introduction: whither spatial econometrics? J Reg Sci 52(2):167–171

Chapter 2
Linear Spatial Dependence Models for Cross-Section Data

Abstract This chapter gives an overview of all linear spatial econometric models with different combinations of interaction effects that can be considered, as well as the relationships between them. It also provides a detailed overview of the direct and indirect effects estimates that can be derived from these models. In addition, it critically discusses the stationarity conditions that need to be imposed on the spatial interaction parameters and the spatial weights matrix, as well as the row-normalization procedure of the spatial weights matrix. The well-known cross-sectional dataset of Anselin (1988), explaining the crime rate by household income and housing values in 49 Columbus, Ohio neighborhoods, is used for illustration purposes.

Keywords Interaction effects · Model overview · Stationarity conditions · Normalization · Spatial spillover effects · Software · Crime

2.1 Introduction

Starting with a standard linear regression model, three different types of interaction effects in a spatial econometric model can be distinguished: endogenous interaction effects among the dependent variable (Y), exogenous interaction effects among the independent variables (X), and interaction effects among the error terms (ε). Originally, the central focus of spatial econometrics has been the spatial lag model, also known as the spatial autoregressive (SAR) model, and the spatial error model (SEM), both with one type of interaction effect.[1] The first model contains endogenous interaction effects, and the second model interaction effects among the error terms. The seminal book by Anselin (1988) and the testing procedure for a spatial lag or a spatial error model based on the robust Lagrange Multiplier tests

[1] In this book, we use the acronyms most commonly used in the spatial econometrics literature to refer to the model specifications (see e.g., LeSage and Pace 2009).

J. P. Elhorst, *Spatial Econometrics*, SpringerBriefs in Regional Science,
DOI: 10.1007/978-3-642-40340-8_2,

developed by Anselin et al. (1996) may be considered as the main pillar behind this way of thinking.

In 2007 the interest for models containing more than one spatial interaction effect increased. In his keynote speech at the first World Conference of the Spatial Econometrics Association in 2007, Harry Kelejian advocated models that include both endogenous interaction effects and interaction effects among the error terms (based on Kelejian and Prucha 1998 and related work). This model is denoted by the term SAC in LeSage and Pace (2009, p. 32), though without pointing out what this acronym is standing for. Elhorst (2010) labels this model the Kelejian–Prucha model after their article in 1998 since they were the first to set out an estimation method for this model, also when the spatial weights matrix used to specify the spatial lag and the spatial error structure is the same. Kelejian and Prucha themselves alternately use the terms SARAR or Cliff-Ord type spatial model.

In his presidential address at the 54th North American Meeting of the Regional Science Association International in 2007, James LeSage advocated models that include both endogenous and exogenous interaction effects. This idea is worked out in the textbook which he published together with Kelley Pace in 2009 (LeSage and Pace 2009). In analogy to Durbin (1960) for the time series case, Anselin (1988) labeled the latter model as the spatial Durbin model (SDM).

Gibbons and Overman (2012) criticize the SAR, SEM and SDM models for reasons of identification, and advocate the SLX (spatial lag of *X*) model. To provide a better understanding, Section 2.2 first gives an overview of all linear spatial econometric models with different combinations of interaction effects that can be considered, as well as the relationships between them. Section 2.3 discusses the stationarity conditions that need to be imposed on the spatial interaction parameters and the spatial weights matrix. Section 2.4 explains and, more importantly, also critically discusses the row-normalization procedure of the spatial weights matrix. Too often this procedure leads to a misspecification problem, which can easily be avoided. Section 2.5 examines the parameter space on which the spatial interaction parameters are defined. Too often this parameter space is simply assumed to be $(-1, 1)$, just as in a time-series model. It is shown that this interval in a second order spatial autoregressive process is too restrictive, because it would lead to the exclusion of feasible and perhaps also relevant parameter combinations. Section 2.6 discusses some strengths and weakness of different estimation methods of spatial econometric models. Section 2.7 gives a detailed overview of direct and indirect effects estimates. The latter are also known as spatial spillover effects. Until recently, empirical studies used the coefficient estimates of a spatial econometric model to test the hypothesis as to whether or not spatial spillovers exist. However, LeSage and Pace (2009) point out that a partial derivative interpretation of the impact from changes to the variables represents a more valid basis for testing this hypothesis. By considering these partial derivatives, it is shown that some models are more flexible in modeling spatial spillovers than others. Section 2.8 lists software to estimate the models discussed in this chapter and presents Matlab routines the author of this book has made available at his Web site. Section 2.9 empirically illustrates the results of different spatial econometric models. Finally, Section 2.10 concludes.

2.2 A Taxonomy of Linear Spatial Dependence Models for Cross-Section Data

The standard approach in most spatial analyses is to start with a non-spatial linear regression model and then to test whether or not this so-called benchmark model needs to be extended with spatial interaction effects. This approach is known as the specific-to-general approach.[2] The non-spatial linear regression model takes the form

$$\boldsymbol{Y} = \alpha \boldsymbol{\iota}_N + \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\boldsymbol{Y}$ denotes an $N \times 1$ vector consisting of one observation on the dependent variable for every unit in the sample ($i = 1,\ldots,N$), $\boldsymbol{\iota}_N$ is an $N \times 1$ vector of ones associated with the constant term parameter α to be estimated, $\boldsymbol{X}$ denotes an $N \times K$ matrix of exogenous explanatory variables, $\boldsymbol{\beta}$ is an associated $K \times 1$ vector with unknown parameters to be estimated, and $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots, \varepsilon_N)^T$ is a vector of disturbance terms, where ε_i is assumed to be independently and identically distributed for all i with zero mean and variance σ^2.[3] Since the linear regression model is commonly estimated by Ordinary Least Squares (OLS), it is often labeled the OLS model.

The opposite approach is to start with a more general model containing, nested within it as special cases, a series of simpler models that ideally should represent all the alternative economic hypotheses requiring consideration. Generally, three different types of interaction effects may explain why an observation associated with a specific location may be dependent on observations at other locations. The first are endogenous interaction effects, where the dependent variable of a particular unit A depends on the dependent variable of other units, among which, say, unit B, and vice versa,

$$\text{Dependent variable } y \text{ of unit } A \leftrightarrow \text{Dependent variable } y \text{ of unit } B \tag{2.2}$$

Endogenous interaction effects are typically considered as the formal specification for the equilibrium outcome of a spatial or social interaction process, in which the value of the dependent variable for one agent is jointly determined with that of neighboring agents. In the empirical literature on strategic interaction among local governments, for example, endogenous interaction effects are

[2] For an explanation of this terminology see Hendry (1995).

[3] The superscript T indicates the transpose of a vector or matrix.

theoretically consistent with the situation where taxation and expenditures on public services interact with taxation and expenditures on public services in nearby jurisdictions (Brueckner 2003).

The second are exogenous interaction effects, where the dependent variable of a particular unit depends on independent explanatory variables of other units

$$\text{Independent variable } x \text{ of unit } B \leftrightarrow \text{ Dependent variable } y \text{ of unit } A \tag{2.3}$$

Consider, for example, the savings rate. According to standard economic theory, saving and investment are always equal. People cannot save without investing their money somewhere, and they cannot invest without using somebody's savings. This is true for the world as a whole, but it is not true for individual economies. Capital can flow across borders; hence the amount an individual economy saves does not have to be the same as the amount it invests. In other words, per capita income in one economy also depends on the savings rates of neighboring economies. It should be stressed that, if the number of independent explanatory variables in a linear regression model is K, the number of exogenous interaction effects might also be K, provided that the intercept is considered as a separate variable. In other words, not only the savings rate but also other explanatory variables may affect per capita income in neighboring economies. It is for this reason that in both the theoretical and the empirical literature on economic growth and convergence among countries or regions, the economic growth variable is taken to depend not only on the initial income level and the rates of saving, population growth, technological change and depreciation in the own economy, but also on those variables in neighboring economies (Ertur and Koch 2007; Elhorst et al. 2010).

The third type of interaction effects are those among the error terms

$$\text{Error term } u \text{ of unit } A \leftrightarrow \text{ Error term } u \text{ of unit } B \tag{2.4}$$

Interaction effects among the error terms do not require a theoretical model for a spatial or social interaction process, but instead, are consistent with a situation where determinants of the dependent variable omitted from the model are spatially autocorrelated, or with a situation where unobserved shocks follow a spatial pattern. Interaction effects among the error terms may also be interpreted to reflect a mechanism to correct rent-seeking politicians for unanticipated fiscal policy changes (Allers and Elhorst 2005).

A full model with all types of interaction effects takes the form

$$\boldsymbol{Y} = \delta \boldsymbol{WY} + \alpha \boldsymbol{\iota}_N + \boldsymbol{X\beta} + \boldsymbol{WX\theta} + \boldsymbol{u} \tag{2.5a}$$

$$\boldsymbol{u} = \lambda \boldsymbol{Wu} + \boldsymbol{\varepsilon} \tag{2.5b}$$

where $\boldsymbol{WY}$ denotes the endogenous interaction effects among the dependent variable, $\boldsymbol{WX}$ the exogenous interaction effects among the independent variables, and $\boldsymbol{Wu}$ the interaction effects among the disturbance term of the different units. We

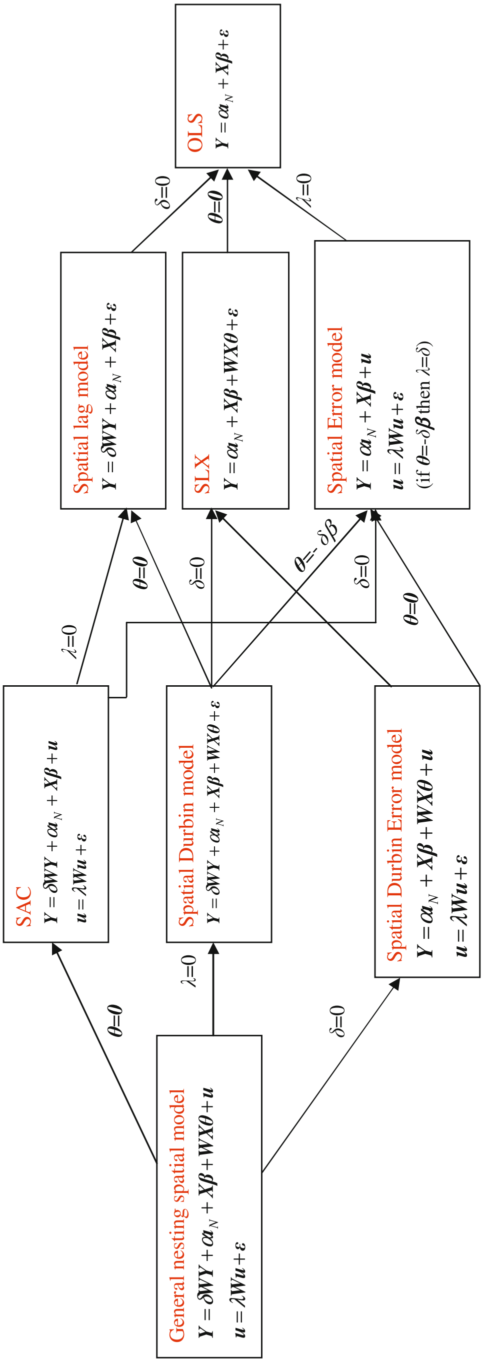

Fig. 2.1 The relationships between different spatial dependence models for cross-section data (*source* Halleck Vega and Elhorst 2012)

will refer to model (2.5a, b) as the general nesting spatial (GNS) model[4] since it includes all types of interaction effects. δ is called the spatial autoregressive coefficient, λ the spatial autocorrelation coefficient, while $\boldsymbol{\theta}$, just as $\boldsymbol{\beta}$, represents a $K \times 1$ vector of fixed but unknown parameters to be estimated. $\boldsymbol{W}$ is a nonnegative $N \times N$ matrix describing the spatial configuration or arrangement of the units in the sample. The next section discusses stationarity conditions that need to be imposed on $\boldsymbol{W}$ to obtain consistent estimators of the parameters in the GNS model.

Figure 2.1 summarizes a family of nine linear spatial econometric models, among which are the OLS model in (2.1) on the right-hand side and the GNS model in (2.5a, b) on the left-hand side. Each model to the right of the GNS model can be obtained from that model by imposing restrictions on one or more of its parameters. The restrictions are shown next to the arrows in Fig. 2.1. This figure shows that there are spatial econometric models that are hardly considered or used in econometric-theoretic and empirical research. The spatial Durbin error model (SDEM), which contains exogenous interaction effects and interaction effects among the error terms, is the best example. In this respect, it should be stressed that there is a large gap in the level of interest in different types of interaction effects between theoreticians and practitioners. Theoreticians are mainly interested in the SAR and SEM models, as well as the SAC model that combines endogenous interaction effects and interaction effects among the error terms, because of the econometric problems accompanying the estimation of these models. Some of these problems will be dealt with in the remainder of this chapter. The reason they generally do not focus on spatial econometric models with exogenous interaction effects is because the estimation of such models does not pose any econometric problems; standard estimation techniques suffice under these circumstances. Consequently, the SLX model is generally not part of the toolbox of researchers interested in the econometric theory of spatial models.

2.3 Stationarity Conditions for δ, λ and W

Spatial weights matrices commonly used in applied research are: (i) p-order binary contiguity matrices (if $p = 1$ only first-order neighbors are included, if $p = 2$ first and second order neighbors are considered, and so on); (ii) inverse distance matrices (with or without a cut-off point); (iii) q-nearest neighbor matrices (where q is a positive integer); (iv) block diagonal matrices where each block represents a group of spatial units that interact with each other but not with observations in other groups. Generally, spatial weights matrices are symmetric, but there are exceptions in which the spatial weights matrix is asymmetric. One example is a commuting flow matrix used to explain regional labor market performance.

[4] LeSage and Pace (2009) neither name nor assign an equation number to model (2.5a, b), which reflects the fact that this model is typically not used in applied research.

A symmetric matrix has the property that all its characteristic roots are real, also when it is row-normalized (see Sect. 2.4) and becomes asymmetric as a result, while an asymmetric matrix will also have complex characteristic roots.

Kelejian and Prucha (1998, 1999) and Lee (2004) make the following assumptions to prove consistency of respectively the GMM estimator of the parameters in the SAR and SAC models and the ML estimator in the SAR model. The spatial weights matrix $\boldsymbol{W}$ is a nonnegative matrix of known constants. The diagonal elements are set to zero by assumption, since no spatial unit can be viewed as its own neighbor. The matrices $\boldsymbol{I}_N$-$\delta\boldsymbol{W}$ and $\boldsymbol{I}_N$-$\lambda\boldsymbol{W}$ are non-singular, where $\boldsymbol{I}_N$ represents the identity matrix of order N. For a symmetric $\boldsymbol{W}$, this condition is satisfied as long as δ and λ are in the interior of $(1/\omega_{min}, 1/\omega_{max})$, where ω_{min} denotes the smallest (i.e. most negative) and ω_{max} the largest real characteristic root of $\boldsymbol{W}$. If $\boldsymbol{W}$ is normalized subsequently, the latter interval takes the form $(1/\omega_{min}, 1)$, since the largest characteristic root of $\boldsymbol{W}$ equals unity in this situation. If $\boldsymbol{W}$ is an asymmetric matrix before it is normalized, it may have complex characteristic roots. LeSage and Pace (2009, pp. 88–89) demonstrate that in that case δ and λ are restricted to the interval $(1/r_{min}, 1)$, where r_{min} equals the most negative purely real characteristic root of $\boldsymbol{W}$ after this matrix is row-normalized. Kelejian and Prucha (1998, 1999) assume that δ and λ are restricted to the interval $(-1, 1)$. We come back to this in Sect. 2.5. Finally, one of the following two conditions should be satisfied: (a) the row and column sums of the matrices $\boldsymbol{W}$, $(\boldsymbol{I}_N\text{-}\delta\boldsymbol{W})^{-1}$ and $(\boldsymbol{I}_N\text{-}\lambda\boldsymbol{W})^{-1}$ before $\boldsymbol{W}$ is row-normalized should be uniformly bounded in absolute value as N goes to infinity, or (b) the row and column sums of $\boldsymbol{W}$ before $\boldsymbol{W}$ is row-normalized should not diverge to infinity at a rate equal to or faster than the rate of the sample size N. Condition (a) is originated by Kelejian and Prucha (1998, 1999), and condition (b) by Lee (2004). Both conditions limit the cross-sectional correlation to a manageable degree, i.e. the correlation between two spatial units should converge to zero as the distance separating them increases to infinity. Below we discuss which of the four matrices introduced above satisfy both conditions (a) and (b), which only satisfy (b), and which satisfy neither (a) and (b).

When the spatial weights matrix is a p-order binary contiguity matrix and p is small, (a) is satisfied. Normally, no spatial unit is assumed to be a neighbor to more than a given number, say q, of other units. Automatically, condition (b) is also satisfied. By contrast, when the spatial weights matrix is an inverse distance matrix, (a) may not be satisfied. To see this, consider an infinite number of spatial units that are arranged linearly. Let the distance of each spatial unit to its first left- and right-hand neighbor be d; to its second left- and right-hand neighbor, the distance $2d$; and so on. When $\boldsymbol{W}$ is an inverse distance matrix and its off-diagonal elements are of the form $1/d_{ij}$, where d_{ij} is the distance between two spatial units i and j, each row sum is $2 \times (1/d + 1/(2d) + 1/(3d) + \ldots)$, representing a series that is not finite. This is perhaps the reason why some empirical applications introduce a cut-off point d^* such that $w_{ij} = 0$ if $d_{ij} > d^*$. However, since the ratio $2 \times (1/d + 1/(2d) + 1/(3d) + \ldots)/N \rightarrow 0$ as N goes to infinity, condition (b) is satisfied, which implies that an inverse distance matrix without a cut-off point does not necessarily have to be excluded in an empirical study for reasons of consistency. Nevertheless, an inverse distance matrix is a border case, which explains why it sometimes leads to numerical problems or unexpected outcomes in empirical applications. This is because the number of units in the sample generally does not go to infinity, but is finite.

Another situation occurs when all cross-sectional units are assumed to be neighbors of each other and are given equal weights. In that case all off-diagonal elements of the spatial weights matrix are $w_{ij} = 1$. Since the row and column sums are $N - 1$, these sums diverge to infinity as N goes to infinity. In contrast to the previous case, however, $(N - 1)/N \to 1$ instead of 0 as N goes to infinity. This implies that a spatial weights matrix that has equal weights and that is row-normalized subsequently, $w_{ij} = 1/(N - 1)$, must be excluded for reasons of consistency since it satisfies neither condition (a) nor (b). The alternative is a group interaction matrix, introduced by Case (1991). Suppose there are G groups and that there are N_g cross-sectional units in each group. Let $w_{ij} = 1/(N_g - 1)$ if units i and j belong to the same group, and zero otherwise. If both N and N_g tend to infinity, with at least two units in each group, or if the number of units in each group does not tend to infinity faster than or equal to the number of groups, condition (b) is restored (Lee 2007).

2.4 Normalizing W

For ease of interpretation, it is common practice to normalize $\boldsymbol{W}$ such that the elements of each row sum to unity. Since $\boldsymbol{W}$ is nonnegative, this ensures that all weights are between 0 and 1, and has the effect that the weighting operation can be interpreted as an averaging of neighboring values.

As an alternative, $\boldsymbol{W}$ might be normalized such that the elements of each column sum to one. This type of normalization is sometimes used in the new social economics literature (Leenders 2002). Note that the column elements of a spatial weights matrix display the impact *of* a particular unit *on* all other units, while the row elements of a spatial weights matrix display the impact *on* a particular unit *by* all other units. Consequently, row normalization has the effect that the impact on each unit by all other units is equalized, while column normalization has the effect that the impact of each unit on all other units is equalized.

Although common practice, row normalization is not free of criticism. Kelejian and Prucha (2010) demonstrate that normalization of the elements of the spatial weights matrix by a different factor for each row as opposed to a single factor is likely to lead to misspecification problem. This problem occurs especially when an inverse distance matrix is row normalized, because its economic interpretation in terms of distance decay will then no longer be valid (Anselin 1988, pp. 23–24; Elhorst 2001). There are (at least) two reasons for this. First of all, because of row-normalization the spatial weights matrix may become asymmetric, as a result of which the impact of unit i on unit j is not the same as that of unit j on unit *i*. Secondly, as a consequence of row normalization remote and central regions will end up having the same impact, i.e. independent on their relative location. The following example may illustrate this. Consider a centrally located spatial unit and a remote unit that both have two neighbors. The distance of the first unit to its neighbors is *d*, while the distance of the second unit to its neighbors is a multiple of *d*. Despite this difference in location, the entries in the inverse distance matrix describing the spatial arrangement of the units in the sample will be 1/2 in both cases, provided that the spatial weights matrix is row-normalized.

If $\boldsymbol{W}_0$ denotes the spatial weights matrix before normalization, Elhorst (2001) and Kelejian and Prucha (2010) propose a normalization procedure where each element of $\boldsymbol{W}_0$ is divided by its largest characteristic root, $r_{0,max}$, to get $\boldsymbol{W} = (1/r_{0,max})\boldsymbol{W}_0$.[5] Alternatively, one may normalize $\boldsymbol{W}_0$ by $\boldsymbol{W} = \boldsymbol{D}^{-1/2}\boldsymbol{W}_0\boldsymbol{D}^{-1/2}$, where $\boldsymbol{D}$ is a diagonal matrix containing the row sums of the matrix $\boldsymbol{W}_0$. The first operation has the effect that the characteristic roots of $\boldsymbol{W}_0$ are also divided by $r_{0,max}$, as a result of which $r_{max} = 1$, just like the largest characteristic root of a row-normalized matrix. However, the smallest (purely) real characteristic root of a matrix that is normalized by a single factor is generally not the same as that of a matrix that is row-normalized. The second operation has been proposed by Ord (1975) and has the effect that the characteristic roots of $\boldsymbol{W}$ are identical to the characteristic roots of a row-normalized $\boldsymbol{W}_0$. Importantly, the mutual proportions between the elements of $\boldsymbol{W}$ remain unchanged as a result of these two normalizations. This is an important property when $\boldsymbol{W}$ represents an inverse distance matrix, since it avoids that this matrix would lose its economic interpretation of distance decay.

2.5 The Parameter Space of δ and λ

To investigate the asymptotic properties of the GMM estimator, Kelejian and Prucha (1998, 1999, and related work) presume that δ is restricted to the interval $(-1, 1)$. This presumption is based on earlier work of Kelejian and Robinson (1995), who demonstrate that the restriction $1/r_{min} < \delta < 1/r_{max}$, before $\boldsymbol{W}$ is row-normalized, may be unnecessarily restrictive since any first-order spatial autoregressive process is defined for every δ as long as the matrix $(\boldsymbol{I}_N\text{-}\delta\boldsymbol{W})$ is non-singular. The following example taken from Elhorst (2001) illustrates this. Let $N = 2$ and $\boldsymbol{W}$ the corresponding spatial weights matrix of two spatial units whose off-diagonal elements are unity, as a result of which $r_{min} = -1$ and $r_{max} = 1$. If $\boldsymbol{\varepsilon} \sim \text{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_N)$, then $\boldsymbol{Y} \sim \text{N}(\boldsymbol{0}, (1 + \delta)/(1 - \delta^2)\sigma^2\boldsymbol{I}_N)$, which shows that the variance of $\boldsymbol{Y}$ is finite when the variance of $\boldsymbol{\varepsilon}$ is finite for every δ unless $\delta = 1/r_{min}$ or $\delta = 1/r_{max}$. In other words, the matrix $(\boldsymbol{I}_N\text{-}\delta\boldsymbol{W})$ is non-singular and its inverse is finite only when δ is not

---

[5] We use the symbol r rather that ω to denote that both symmetric and asymmetric spatial weights matrices are covered here.

equal to the reciprocal of just one of the two characteristic roots of the spatial weights matrix $\boldsymbol{W}$ (see Kelejian and Prucha 2010 for a generalization). According to Bell and Bockstael (2000), this feature is rather curious. The values of δ that make the problem undefined are related directly to the characteristic roots of $\boldsymbol{W}$, which will change if the sample size changes. With no further restrictions, the problem is characterized by a non-continuous parameter space, changing with the addition or the elimination of any observation. To avoid these difficulties and to facilitate the estimation of δ, as well as to ensure the invertibility of the matrix ($\boldsymbol{I}_N$-$\delta\boldsymbol{W}$), Ord (1981) suggests to restrict δ to $1/r_{min} < \delta < 1/r_{max}$ before $\boldsymbol{W}$ is row-normalized and to $1/r_{min} < \delta < 1$ after this. Kelejian and Robinson (1995), on their turn, suggest to restrict δ to $-1 < \delta < 1$, to stress the similarity between time-series and spatial econometrics. A first-order serial autoregressive process

$$y_t = \rho y_{t-1} + \varepsilon_t \tag{2.6}$$

with T observations is stationary if ρ lies in the interval $(-1, 1)$. However, the same interval for a first-order spatial autoregressive process would be too restrictive. For normalized spatial weights, the largest characteristic root is indeed +1, but no general result holds for the smallest characteristic root, and the lower bound will be typically less than -1.

Although there might be some similarities for first-order models, substantive differences occur when considering second-order models. The time-series literature (see Beach and MacKinnon 1978, and the references therein) has pointed out that a second-order serial autoregressive process

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t \tag{2.7}$$

with T observations is stationary if $\rho_1 + \rho_2 < 1$, $1 + \rho_2 - \rho_1 > 0$ and $\rho_2 > -1$. These constraints define a triangular region with vertices at $(-2, -1)$, $(0, 1)$ and $(2, -1)$.

A second-order spatial autoregressive process takes the form

$$\boldsymbol{Y} = \delta_1 \boldsymbol{W}_1 \boldsymbol{Y} + \delta_2 \boldsymbol{W}_2 \boldsymbol{Y} + \boldsymbol{\varepsilon} \tag{2.8}$$

where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are assumed to be normalized. This model as well as some extensions of it have been considered in many studies. Examples are Brandsma and Ketellapper (1979), Sherrell (1990), Hepple (1995), Bell and Bockstael (2000), Bordignon et al. (2003), Lacombe (2004), Allers and Elhorst (2005), McMillen et al. (2007), Ward and Gleditsch (2008); Dall'Erba et al. (2008), Elhorst and Fréret (2009), Lee and Liu (2010), Badinger and Egger (2011), and Elhorst et al. (2012). However, most of these studies do not specify a parameter space for δ_1 and δ_2. Only Lee and Liu (2010) and Badinger and Egger (2011) mention that the sum of the absolute values of the two spatial parameters should be less than one ($|\delta_1| + |\delta_2| < 1$). However, it can be easily seen that this constraint proves to be too restrictive. The fact that δ in a first-order spatial autoregressive process should lie in the interval $(1/r_{min}, 1)$ immediately determines four coordinates of the stationarity region: $(1, 0)$ and $(1/r_{1,min}, 0)$ in case $\delta_2 = 0$, and $(0, 1)$

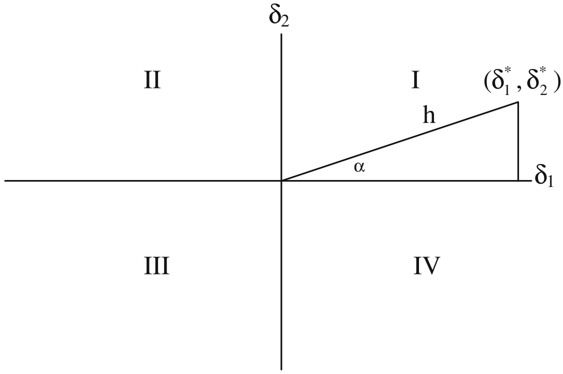

Fig. 2.2 δ_1 and δ_2 and the four quadrants in a two-dimensional space

and (0, $1/r_{2,min}$) in case $\delta_1 = 0$. These four coordinates define a region that is wider than the one assumed in Lee and Liu (2010) and Badinger and Egger (2011). Additionally, these four coordinates demonstrate that the stationarity region does not coincide with the time-series version of the model.

Elhorst et al. (2012) have developed the following procedure to determine the exact boundaries of the curves connecting these four coordinates, depending on the specification of $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$. Consider the four quadrants defined by the two axes in Fig. 2.2 and the angle α between δ_1 and the hypotenuse (h) that connects the origin with the coordinates of a point located at the border of the feasible region, denoted by (δ_1^*, δ_2^*). Since $\tan(\alpha) = \delta_2^*/\delta_1^*$, we get

$$\delta_2^* = \tan(\alpha)\delta_1^*, \quad \text{for } -270^\circ < \alpha < -90^\circ \quad \text{or } -90^\circ < \alpha < 90^\circ \tag{2.9}$$

Consequently, Eq. (2.8) can be rewritten as

$$\boldsymbol{Y} = \delta_1^*[\boldsymbol{W}_1 + \tan(\alpha)\boldsymbol{W}_2]\boldsymbol{Y} + \varepsilon = \delta_1^*\boldsymbol{W}^*\boldsymbol{Y} + \varepsilon \tag{2.10}$$

This implies that the model is stationary for the following parameter combinations (depending on α)

$$0 < \delta_1 < 1/r_{\max}[\boldsymbol{W}^*], \quad 0 \le \delta_2 \le \tan(\alpha)/r_{\max}[\boldsymbol{W}^*], \quad 0^\circ \le \alpha < 90^\circ \tag{2.11a}$$

$$0 < \delta_1 < 1/r_{\max}[\boldsymbol{W}^*], \quad \tan(\alpha)/r_{\max}[\boldsymbol{W}^*] < \delta_2 < 0, \quad -90^\circ < \alpha < 0^\circ \tag{2.11b}$$

$$\delta_1 = 0, \quad 1/r_{\min}[\boldsymbol{W}_2] < \delta_2 < 0 \quad \alpha = -90^\circ \tag{2.11c}$$

$$1/r_{\min}[\boldsymbol{W}^*] < \delta_1 < 0, \quad \tan(\alpha)/r_{\max}[\boldsymbol{W}^*] \le \delta_2 \le 0, \quad -180^\circ \le \alpha < -90^\circ \tag{2.11d}$$

$$1/r_{\min}[\boldsymbol{W}^*] < \delta_1 < 0, \quad 0 < \delta_2 < \tan(\alpha)/r_{\min}[\boldsymbol{W}^*], \quad -270^\circ < \alpha < -180^\circ \tag{2.11e}$$

$$\delta_1 = 0, \quad 0 < \delta_2 < 1/r_{\max}[\boldsymbol{W}_2], \quad \alpha = -270^\circ \tag{2.11f}$$

where r_{max}[.] and r_{min}[.] are the largest (positive) and smallest (negative) purely real characteristic roots of the matrix in square brackets. When $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are

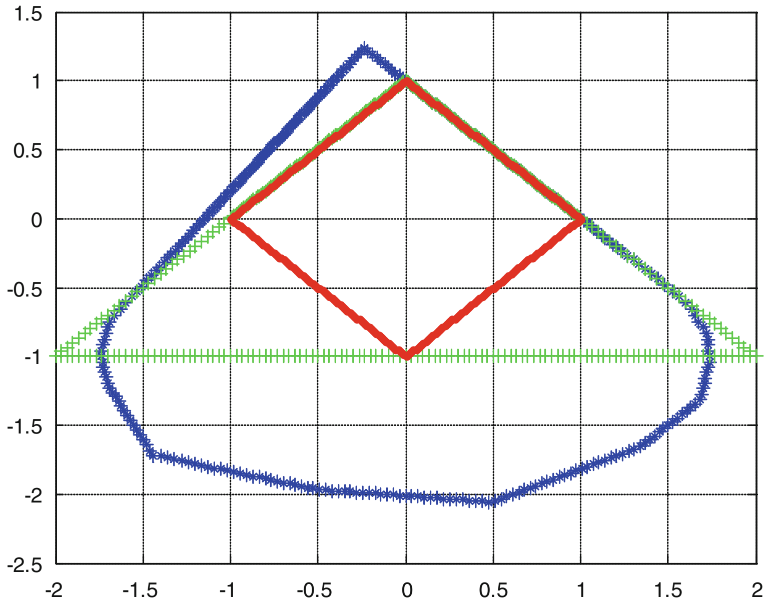

Fig. 2.3 The potential shape of the stationarity region of a second-order spatial autoregressive process (*source* Elhorst et al. 2012)

normalized, the largest characteristic root of the matrix $\delta_1 \boldsymbol{W}_1 + \delta_2 \boldsymbol{W}_2$ in the first quadrant is $\delta_1 + \delta_2$. This implies that the model is stationary for values of δ_1 and δ_2 in the first quadrant if $\delta_1 + \delta_2 < 1$. This expression shows that the parameter space in the first quadrant is independent of $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ and thus identical to both the time-series form and the definition given by Lee and Liu (2010) and Badinger and Egger (2011). The curves connecting the coordinates in the other quadrants, depend on $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ and, therefore, might define different parameter spaces.

Elhorst et al. (2012) illustrate the potential shape of the stationarity region for different pairs of spatial weights matrices. Typically, the stationarity region takes the form graphed (in blue) in Fig. 2.3. In addition, Fig. 2.3 graphs the rhombus (in red) implied by the restriction $|\delta_1| + |\delta_2| < 1$ and the triangle (in green) that corresponds to the stationarity region of the second-order serial autoregressive process.

The rhombus in Fig. 2.3 shows that the naïve adoption of the restriction $|\delta_1| + |\delta_2| < 1$ in a second-order spatial autoregressive process is not recommended, because it would lead to the exclusion of feasible and perhaps also relevant parameter combinations. Up to now, positive spatial autocorrelation has been encountered in empirical data more frequently than negative spatial autocorrelation, and researchers tend to consider negative autocorrelation less relevant. Typically, if a particular variable increases (decreases) in one area, it also tends to increase (decrease) in neighboring areas. However, Griffith and Arbia (2010) present three examples of negatively spatially autocorrelated phenomena that are all based on the economic notion of competitive locational processes. If the manifestation of a certain phenomenon in one area is at the expense of its neighboring areas, then negative spatial autocorrelation is likely to occur. Let $(\delta_1,\delta_2) = (0.8, -0.3)$ or $(\delta_1,\delta_2) = (1.1,-0.3)$ be the outcome of a spatial econometric model. Based on the restriction $|\delta_1| + |\delta_2| < 1$, these combinations of values would be rejected but based on the results presented in this section, they should not be excluded. Additionally, as discussed in Griffith and Arbia (2010), they also make sense from a theoretical point of view.

The triangle that corresponds to the stationarity region of the second-order serial autoregressive process shows that the naïve adoption of the time-series region is not recommended either. It would not only lead to the exclusion of feasible parameter combinations, but also to include some infeasible ones.

In other words, the knowledge of the exact boundary is important, both for estimation and inference. More discussion on these issues can also be found in LeSage and Pace (2011). Elhorst et al. (2012) have made a Matlab routine downloadable for free on their web sites to determine the exact boundaries of any second-order spatial autoregressive process.[6]

2.6 Methods of Estimation

Spatial econometric models can be estimated by maximum likelihood (ML) (Ord 1975), quasi-maximum likelihood (QML) (Lee 2004), instrumental variables (IV) (Anselin 1988, pp. 82–86), generalized method of moments (GMM) (Kelejian and Prucha 1998, 1999), or by Bayesian Markov Chain Monte Carlo methods (Bayesian MCMC) (LeSage 1997). In the next two chapters we extensively discuss the ML estimation procedure of (dynamic) spatial panel data models. Due to the overlap with the ML estimation procedure of cross-sectional spatial econometric models, this section only discusses some strengths and weaknesses of the different estimation methods. Furthermore, updated overviews of these estimation methods can be found in the Handbook of Regional Science that appeared in 2013.

One advantage of QML and IV/GMM estimators is that they do not rely on the assumption of normality of the disturbances ε. Nonetheless, both estimators assume that the disturbance terms ε_i are independently and identically distributed for all i with zero mean and variance σ^2. One disadvantage of the IV/GMM estimator is the possibility of ending up with a coefficient estimate for δ in the SAR model or for λ in the SEM model outside its parameter space. Whereas these coefficients are restricted to the interval $(1/r_{min}, 1)$ by the Jacobian term in the log-likelihood function of ML estimators or in the conditional distribution of the spatial parameter of Bayesian estimators, they are unrestricted using IV/GMM since these estimators ignore the Jacobian term.

---

[6] http://www.regroningen.nl/elhorst and http://community.wvu.edu/~djl041/.

To avoid computational difficulties was one of the reasons to develop IV/GMM estimators (Kelejian and Prucha, 1998, 1999). Estimation of spatial econometric models involves the manipulation of $N \times N$ matrices, such as matrix multiplication, matrix inversion, the computation of characteristic roots and/or Cholesky decomposition. These manipulations may be computationally intensive and/or may require significant amounts of memory if N is large. Since IV/GMM estimators ignore the Jacobian term, many of these problems could be avoided. In Chap. 4 of their book, however, LeSage and Pace (2009) produce conclusive evidence that many of these computational difficulties have become a thing of the past for ML and Bayesian estimators.

In spite of this, Fingleton and Le Gallo (2007, 2008), Drukker et al. (2013) and Liu and Lee (2013) show that IV/GMM estimators are extremely useful in those cases where linear spatial dependence models contain one or more endogenous explanatory variables (other than the spatially lagged dependent variable) that need to be instrumented, because of measurement errors in explanatory variables, omitted variables correlated with included explanatory variables, or because of the existence of an underlying (perhaps unspecified or unknown) set of simultaneous structural equations. ML or Bayesian estimators of single equation models with a spatial lag (i.e. the spatial lag model and the spatial Durbin model) and additional endogenous variables do not feature in the spatial econometrics literature and would be difficult, if not impossible, to derive. The same applies to single equation models with a spatial error process (i.e. the spatial error model and the spatial Durbin error model). By contrast, models including a spatial lag and additional endogenous variables can be straightforwardly estimated by two-stage least squares (2SLS). To instrument the spatially lagged dependent variable, Kelejian et al. (2004) suggest $[\boldsymbol{X}\ \boldsymbol{W}\boldsymbol{X} \ldots \boldsymbol{W}^g\boldsymbol{X}]$, where g is a pre-selected constant.[7] Typically, researchers take $g = 1$ or $g = 2$, dependent on the number of regressors and the type of model. One potential problem in case of the spatial Durbin model is that g should be at least two, since this model already contains the variables $\boldsymbol{X}$ and $\boldsymbol{W}\boldsymbol{X}$ on the right-hand side. This means that the number of potential strong instruments diminishes considerably.

If one or more of the explanatory variables are endogenous, the set of instruments must be limited to $[\boldsymbol{X}^{ex}\ \boldsymbol{W}\boldsymbol{X}^{ex} \ldots \boldsymbol{W}^d\boldsymbol{X}^{ex}]$, where 'ex' denotes the $\boldsymbol{X}$ variables that are exogenous. Furthermore, this set should be used to instrument the additional endogenous explanatory variables. A similar type of extension applies to Kelejian and Prucha's (1999) GMM estimator for models including a spatial error process together with endogenous explanatory variables (Fingleton and Le Gallo 2007). In addition, Fingleton and Le Gallo (2008) consider a mixed 2SLS/GMM estimator of the Kelejian-Prucha model extended to include endogenous explanatory variables.

[7] Lee (2003) introduces the optimal instrument 2SLS estimator, but Kelejian et al. (2004) show that the 2SLS estimator based on this set of instruments has quite similar small sample properties.

Liu and Lee (2013) consider the IV estimation of the spatial lag model with endogenous regressors when the number of instruments grows with the sample size. They suggest a bias-correction procedure based on the leading-order many-instrument bias. To choose among different instruments, they also suggest minimizing an approximation of the mean square error of both the 2SLS and bias-corrected 2SLS estimators.

In an overview paper, Drukker et al. (2013) consider the GMM estimation of the Kelejian-Prucha model with endogenous regressors. Since the model contains more endogenous explanatory variables than $\boldsymbol{WY}$, they suggest (p. 693) "to use a set of instruments as above … augmented by other exogenous variables expected to be part of the reduced form of the system".

One major weakness of spatial econometric models is that the spatial weights matrix $\boldsymbol{W}$ cannot be estimated but needs to be specified in advance and that economic theory underlying spatial econometric applications often has little to say about the specification of $\boldsymbol{W}$ (Leenders 2002). For this reason, it has become common practice to investigate whether the results are robust to the specification of $\boldsymbol{W}$. The same spatial econometric model is estimated, say, S times, every time with a different spatial weights matrix, to investigate whether the estimation results are sensitive to the choice of $\boldsymbol{W}$. One advantage of the Bayesian MCMC estimator is that it offers a criterion, the Bayesian posterior model probability, to select the spatial weights matrix that best describes the data. Whereas tests for significant differences between log-likelihood function values, such as the LR-test, can formally not be used if models are non-nested (i.e. based on different spatial weights matrices), Bayesian posterior model probabilities do not require nested models to carry out these comparisons. The basic idea is to set prior probabilities equal to $1/S$, making each model equally likely a priori, to estimate each model by Bayesian methods, and then to compute posterior probabilities based on the data and the estimation results of this set of S models. Successful applications of this methodology can be found in LeSage and Page (2009, Chap. 6) and Seldadyo et al. (2010).

A Monte-Carlo study of Stakhovych and Bijmolt (2009) demonstrates that a weights matrix selection procedure that is based on 'goodness-of-fit' criteria increases the probability of finding the true specification. If a spatial interaction model is estimated based on S different spatial weights matrices and the log-likelihood function value of every model is estimated, one may select the spatial weights matrix exhibiting the highest log-likelihood function value. However, since LR-tests may formally not be used, one better selects the spatial weights matrix exhibiting the highest Bayesian posterior model probability. Alternatively, one may use J-type statistics to discriminate between different specifications of $\boldsymbol{W}$ (Anselin 1986; Kelejian 2008; Burridge and Fingleton 2010; Burridge 2012).

Harris et al. (2011) criticize these empirical approaches, because they would only find a local maximum among the competing spatial weights matrices and not necessarily a correctly specified $\boldsymbol{W}$ (unless it is unknowingly included in the set of competing matrices considered). However, the Monte Carlo results found by Stakhovych and Bijmolt (2009) partly refute this critique. Although there is a serious probability of selecting the wrong spatial weights matrix if spatial dependence is weak (δ or λ are relatively small in magnitude), the consequences of this poor choice are limited because the coefficient estimates are quite close to the true ones. Conversely, although the wrong choice of a spatial weights matrix can distort the coefficient estimates severely, the probability that this really happens is small if spatial dependence is strong (δ or λ are relatively large in magnitude).

Corrado and Fingleton (2012) strongly argue for the use of more substantive theory in empirical spatial econometric modeling, especially regarding $\boldsymbol{W}$. Despite their criticism, they point out that alternatives to $\boldsymbol{W}$ that have been proposed by e.g., Folmer and Oud (2008) and Harris et al. (2011), such as entering variables in the regression model that proxy spillovers, also require identifying assumptions. In other words, this approach also involves an a priori specification of the spatial relation between units in the sample.

2.7 Direct and Indirect (or Spillover) Effects

Many empirical studies use the point estimates of one or more spatial regression model specifications (δ, $\boldsymbol{\theta}$ and/or λ) to draw conclusions as to whether or not spatial spillovers exist. One of the key contributions of LeSage and Pace's book (2009, p. 74) is the observation that this may lead to erroneous conclusions, and that a partial derivative interpretation of the impact from changes to the variables of different model specifications represents a more valid basis for testing this hypothesis. To illustrate this, they give an example of a spatially lagged independent variable *WX* whose coefficient is negative and insignificant (ibid, Table 3.3), while it's spatial spillover effect is positive and significant (ibid, Table 3.4). The explanation for this can be seen by the derivation below.

By rewriting the general nesting spatial (GNS) model in (2.5a, b) as

$$\boldsymbol{Y} = (\boldsymbol{I} - \delta\boldsymbol{W})^{-1}(\boldsymbol{X\beta} + \boldsymbol{WX\theta}) + \boldsymbol{R} \tag{2.12}$$

where $\boldsymbol{R}$ is a rest term containing the intercept and the error terms, the matrix of partial derivatives of the expected value of $\boldsymbol{Y}$ with respect to the kth explanatory variable of X in unit 1 up to unit N in time can be seen to be

$$\begin{aligned}\left[\frac{\partial E(\boldsymbol{Y})}{\partial x_{1k}} \quad . \quad \frac{\partial E(\boldsymbol{Y})}{\partial x_{Nk}}\right] &= \begin{bmatrix} \frac{\partial E(y_1)}{\partial x_{1k}} & . & \frac{\partial E(y_1)}{\partial x_{Nk}} \\ . & . & . \\ \frac{\partial E(y_N)}{\partial x_{1k}} & . & \frac{\partial E(y_N)}{\partial x_{Nk}} \end{bmatrix} \\ &= (\boldsymbol{I} - \delta\boldsymbol{W})^{-1} \begin{bmatrix} \beta_k & w_{12}\theta_k & . & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & . & w_{2N}\theta_k \\ . & . & . & . \\ w_{N1}\theta_k & w_{N2}\theta_k & . & \beta_k \end{bmatrix}\end{aligned} \tag{2.13}$$

where w_{ij} is the (i, j)th element of $\boldsymbol{W}$. This result illustrates that the partial derivatives of E($\boldsymbol{Y}$) with respect to the kth explanatory variable have three important properties. First, if a particular explanatory variable in a particular unit changes, not only will the dependent variable in that unit itself change but also the dependent variables in other units. The first is called a *direct effect* and the second an *indirect effect*. Note that every diagonal element of the matrix of partial derivatives represents a direct effect, and that every off-diagonal element represents an indirect effect. Consequently, indirect effects do not occur if both $\delta = 0$ and $\theta_k = 0$, since all off-diagonal elements will then be zero [see (2.13)].

Second, direct and indirect effects are different for different units in the sample. Direct effects are different because the diagonal elements of the matrix $(\boldsymbol{I}_N\text{-}\delta\boldsymbol{W})^{-1}$ are different for different units, provided that $\delta \neq 0$ [see the diagonal elements of (2.13)]. Indirect effects are different because both the off-diagonal elements of the matrix $(\boldsymbol{I}_N\text{-}\delta\boldsymbol{W})^{-1}$ and of the matrix $\boldsymbol{W}$ are different for different units, provided that $\delta \neq 0$ and/or $\theta_k \neq 0$ [see the off-diagonal elements of (2.13)].

Third, indirect effects that occur if $\theta_k \neq 0$ are known as *local effects*, as opposed to indirect effects that occur if $\delta \neq 0$ and that are known as *global effects*. Local effects got their name because they arise only from a unit's neighborhood set; if the element w_{ij} of the spatial weights matrix is non-zero (zero), then the effect of x_{jk} on y_i is also non-zero (zero). Global effects got their name because they also arise from units that do not belong to a unit's neighborhood set. This follows from the fact that the matrix $(\boldsymbol{I}_N\text{-}\delta\boldsymbol{W})^{-1}$, in contrast to $\boldsymbol{W}$, does not contain zero elements (provided that $\delta \neq 0$) [see $\boldsymbol{W}$ and $(\boldsymbol{I}_N\text{-}\delta\boldsymbol{W})^{-1}$ in (2.13)]. If both $\delta \neq 0$ and $\theta_k \neq 0$, both global and local effects occur which cannot be separated from each other.

Since both the direct and indirect effects are different for different units in the sample, the presentation of these effects is a problem. If we have N spatial units and K explanatory variables, we obtain K different $N \times N$ matrices of direct and indirect effects. Even for small values of N and K, it may already be rather difficult to report these results compactly. To improve the surveyability of the estimation results of spatial regression model specifications, LeSage and Pace (2009) therefore propose to report one summary indicator for the direct effect, measured by the average of the diagonal elements of the matrix on the right-hand side of (2.13), and one summary indicator for the indirect effect, measured by the average of either the row sums or the column sums of the off-diagonal elements of that matrix. The average row effect represents the impact on a particular element of the dependent variable as a result of a unit change in all elements of an exogenous variable, while the average column effect represents the impact of changing a particular element of an exogenous variable on the dependent variable of all other units. However, since the numerical magnitudes of these two calculations of the indirect effect are the same, it does not matter which one is used. Generally, the indirect effect is interpreted as the impact of changing a particular element of an exogenous variable on the dependent variable of all other units, which corresponds to the average column effect.

2.7.1 Direct and Indirect Effects of Different Spatial Econometric Models

The direct and indirect effects corresponding to the different spatial econometric models introduced in Sect. 2.2 and presented in Fig. 2.1 and for an arbitrary spatial weights matrix are reported in Table 2.1 (Halleck Vega and Elhorst 2012).

If the OLS model is adopted, the direct effect of an explanatory variable is equal to the coefficient estimate of that variable (β_k), while its indirect effect is zero by construction. If the OLS model is augmented with a spatially autocorrelated error term to obtain the SEM model, the direct and the indirect effects remain the same. This is because the disturbances do not come into play when considering the partial derivative of the dependent variable with respect to changes in the explanatory variables (see [2.13]). This property also holds for the extension of the SAR, SLX and the SDM model with spatial autocorrelation, i.e. the SAC, SDEM and the GNS model, respectively.

If the SLX or the SDEM model is adopted, the direct effect of an explanatory variable is equal to the coefficient estimate of that variable (β_k), while its indirect effect is equal to the coefficient estimate of its spatial lagged value (θ_k). The advantage of these models is that the direct and indirect effects do not require further calculations and that both these effects might be different from one explanatory variable to another.

Things get complicated when moving to one of the other models due to the multiplication with the spatial multiplier matrix $(\boldsymbol{I}\text{-}\delta\boldsymbol{W})^{-1}$. Whereas the direct effect of the kth explanatory variable in the OLS, SEM, SLX and SDEM models is β_k, the direct effect in the SAR and SAC models is β_k premultiplied with a number that will eventually be greater than or equal to unity. This can be seen by decomposing the spatial multiplier matrix as follows

$$(\boldsymbol{I}-\delta\boldsymbol{W})^{-1} = \boldsymbol{I} + \delta\boldsymbol{W} + \delta^2\boldsymbol{W}^2 + \delta^3\boldsymbol{W}^3\ldots \tag{2.14}$$

Since the non-diagonal elements of the first matrix term on the right-hand side (the identity matrix $\boldsymbol{I}$) are zero, this term represents a direct effect of a change in X only. Conversely, since the diagonal elements of the second matrix term on the right-hand side ($\delta\boldsymbol{W}$) were assumed to be zero (see Sect. 2.2), this term represents

Table 2.1 Direct and spillover effects of different model specifications

	Direct effect	Indirect effect
OLS/SEM	β_k	0
SAR/SAC	Diagonal elements of $(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}\beta_k$	Off-diagonal elements of $(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}\beta_k$
SLX/SDEM	β_k	θ_k
SDM/GNS	Diagonal elements of $(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_k + \boldsymbol{W}\theta_k)$	Off-diagonal elements of $(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_k + \boldsymbol{W}\theta_k)$

Source Halleck Vega and Elhorst (2012) an indirect effect of a change in X only. Furthermore, since $\boldsymbol{W}$ is taken to the power 1 here, this indirect effect is limited to first-order neighbors only, i.e. the units that belong to the neighborhood set of every spatial unit. All other terms on the right-hand side represent second- and higher-order direct and indirect effects. Higher-order direct effects arise as a result of feedback effects, i.e. impacts passing through neighboring units and back to the unit itself (e.g. 1→2→1 and 1→2→3→2→1). It is these feedback effects that are responsible for the fact that the overall direct effect is eventually greater than unity.[8]

One important limitation of the spatial lag model is that the ratio between the indirect and the direct effect of a particular explanatory variable is independent of β_k. This is because β_k in the numerator and β_k in the denominator of this ratio cancel each other out. This property implies that the ratio between the indirect and direct effects in the spatial lag model is the same for every explanatory variable, and that its magnitude depends on the spatial autoregressive parameter δ and the specification of the spatial weights matrix $\boldsymbol{W}$ only. In many empirical applications, this is not very likely.

If the SDM model is adopted, both the direct effect and the indirect effect of a particular explanatory variable will also depend on the coefficient estimate θ_k of the spatially lagged value of that variable (see Table 2.1). The result is that no prior restrictions are imposed on the magnitude of both the direct and indirect effects and thus that the ratio between the indirect and the direct effect may be different for different explanatory variables, just as in the SLX and SDEM model. Due to this flexibility, the SLX, SDM, SDEM models are a more attractive point of departure in an empirical study than other spatial regression specifications.

Figure 2.1 and Table 2.1 seem to indicate that the best strategy to test for spatial interaction effects and to determine indirect effects is to start with the most general model. The direct and indirect effects of the GNS model, which were derived in Eq. (2.13), are similar to those of the SDM model. However, one major problem is that the parameters of this GNS model are only weakly identified. The empirical illustration in Sect. 2.9 will show that the SDM and the SDEM models are already difficult to distinguish from each other. This problem is strengthened when estimating the GNS model; it often leads to a model that is overparameterized. Parameters have the tendency to become insignificant as a result of which this model does not outperform the SDM and SDEM models.

---

[8] This also holds if the spatial autoregressive parameter is negative. The first term that produces feedback effects is $\delta^2\boldsymbol{W}^2$. This term will always be positive. The second term is $\delta^3\boldsymbol{W}^3$. Since δ is restricted to the interval $(1/r_{min}, 1)$ and the non-negative elements of $\boldsymbol{W}$ after row-normalisation are smaller than or equal to 1, the diagonal elements of $\delta^3\boldsymbol{W}^3$ are smaller in absolute value than those of $\delta^2\boldsymbol{W}^2$. Since the series $\delta^2\boldsymbol{W}^2 + \delta^3\boldsymbol{W}^3 + \delta^4\boldsymbol{W}^4 + \ldots$ alternates in sign if δ is negative, the sum of the diagonal elements of the matrix represented by this series will always be positive.

2.7.2 Testing for Spatial Spillovers

The estimated indirect effects of the independent explanatory variables should eventually be used to test the hypothesis as to whether or not spatial spillovers exist, rather than the coefficient estimate of endogenous interaction effects (***WY***) and/or the coefficients estimates of the exogenous interaction effects (***WX***). However, one difficulty is that it cannot be seen from the coefficient estimates and the corresponding standard errors or t-values (derived from the variance–covariance matrix) whether the indirect effects in models containing endogenous interaction effects (SAR, SAC, SDM, GNS) are significant. This is because the indirect effects are composed of different coefficient estimates according to complex mathematical formulas and the dispersion of these indirect effects depends on the dispersion of all coefficient estimates involved (see Table 2.1). For example, if the coefficients δ, β_k and θ_k in the spatial Durbin model happen to be significant, this does not automatically mean that the indirect effect of the k^{th} explanatory variable is also significant. Conversely, if one or two of these coefficients are insignificant, the indirect effect may still be significant.

One possible way to calculate the dispersion of the direct and indirect effects is to apply formulas for the sum, the difference, the product and the quotient of random variables (see, among others, Mood et al. 1974, pp. 178–181). However, due to the complexity of the matrix of partial derivatives and because every empirical application will have its own unique number of observations (N) and spatial weights matrix (***W***), it is almost impossible to derive one general approach that can be applied under all circumstances. In order to draw inferences regarding the statistical significance of the direct and indirect effects, LeSage and Pace (2009, p. 39) therefore suggest simulating the distribution of the direct and indirect effects using the variance–covariance matrix implied by the maximum likelihood estimates.

The variance–covariance matrix of the parameter estimates of the GNS model takes the form (rewritten from Anselin 1988, pp. 64–65 without heteroskedasticity)

$$
\begin{aligned}
&Var(\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\delta}, \hat{\lambda}, \hat{\sigma}^2) \\
&= \begin{bmatrix}
\frac{1}{\sigma^2}(\boldsymbol{B}\tilde{\boldsymbol{X}})^T\boldsymbol{B}\tilde{\boldsymbol{X}} & \frac{1}{\sigma^2}(\boldsymbol{B}\tilde{\boldsymbol{X}})^T\boldsymbol{B}\tilde{\boldsymbol{W}}_\delta\tilde{\boldsymbol{X}}\hat{\gamma} & \boldsymbol{0} & \boldsymbol{0} \\
\cdot & trace(\tilde{\boldsymbol{W}}_\delta\tilde{\boldsymbol{W}}_\delta + \boldsymbol{B}\tilde{\boldsymbol{W}}_\delta\boldsymbol{B}^{-1}) + \frac{1}{\sigma^2}(\boldsymbol{B}\tilde{\boldsymbol{W}}_\delta\tilde{\boldsymbol{X}}\hat{\gamma})^T(\boldsymbol{B}\tilde{\boldsymbol{W}}_\delta\tilde{\boldsymbol{X}}\hat{\gamma}) & trace(\tilde{\boldsymbol{W}}_\lambda^T\boldsymbol{B}\tilde{\boldsymbol{W}}_\delta\boldsymbol{B}^{-1} + \boldsymbol{W}\tilde{\boldsymbol{W}}_\delta\boldsymbol{B}^{-1}) & \frac{1}{\sigma^2}trace(\boldsymbol{B}\tilde{\boldsymbol{W}}_\delta\boldsymbol{B}^{-1}) \\
\cdot & \cdot & trace(\tilde{\boldsymbol{W}}_\lambda\tilde{\boldsymbol{W}}_\lambda + \tilde{\boldsymbol{W}}_\lambda^T\tilde{\boldsymbol{W}}_\lambda) & 0 \\
\cdot & \cdot & \cdot & \frac{N}{2\sigma^4}
\end{bmatrix}^{-1}
\end{aligned}
\tag{2.15}
$$

where $\mathbf{B} = \boldsymbol{I} - \hat{\lambda}\boldsymbol{W}$, $\tilde{\boldsymbol{W}}_{\delta} = \boldsymbol{W}(\boldsymbol{I} - \hat{\delta}\boldsymbol{W})^{-1}$, $\tilde{\boldsymbol{W}}_{\lambda} = \boldsymbol{W}(\boldsymbol{I} - \hat{\lambda}\boldsymbol{W})^{-1}$, $\tilde{\boldsymbol{X}} = [\boldsymbol{\iota}_N \quad \boldsymbol{X} \quad \boldsymbol{WX}]$ and $\hat{\gamma} = \begin{bmatrix} \hat{\alpha} & \hat{\boldsymbol{\beta}}^T & \hat{\boldsymbol{\theta}}^T \end{bmatrix}^T$ to simplify notation. Since this matrix is symmetric the lower diagonal elements are not shown.

One particular parameter combination drawn from this variance–covariance matrix (indexed by d) can be obtained by

$$[\alpha_d \quad \boldsymbol{\beta}_d^T \quad \boldsymbol{\theta}_d^T \quad \delta_d \quad \lambda_d \quad \sigma_d^2]^T = \boldsymbol{P}^T\vartheta + [\hat{\alpha} \quad \hat{\boldsymbol{\beta}} \quad \hat{\boldsymbol{\theta}} \quad \hat{\delta} \quad \hat{\lambda} \quad \hat{\sigma}^2]^T \tag{2.16}$$

where $\boldsymbol{P}$ denotes the upper-triangular Cholesky decomposition of the variance–covariance matrix and ϑ is a vector of length $4 + 2K$ (the number of parameters that have been estimated) containing random values drawn from a normal distribution with mean zero and standard deviation one. If D parameter combinations are drawn like this[9] and the (in)direct effect of a particular explanatory variable is determined for every parameter combination, the overall (in)direct effect can be approximated by computing the mean value over these D draws and its significance level (t-value) by dividing this mean by the corresponding standard deviation. If μ_{kd} denotes the indirect effect of the kth explanatory variable of draw d, the overall indirect effect over all draws and the corresponding t-value will be

$$\bar{\mu}_k \text{ (ind. eff. kth var.)} = \frac{1}{D}\sum_{d=1}^{D} \mu_{kd} \tag{2.17a}$$

$$\text{t - value (of ind.eff. kth var.)} = \bar{\mu}_k \Bigg/ \left[\frac{1}{D-1}\sum_{d=1}^{D} (\mu_{kd} - \bar{\mu}_k)^2\right] \tag{2.17b}$$

Given the t-value of this indirect effect, one can finally test whether the kth variable causes spatial spillover effects.

There are two possible approaches to program this. One is to determine the matrix on the right-hand side of (2.13) for every draw and then to calculate the direct and indirect effects corresponding to this draw. The disadvantage of using this approach is that ($\boldsymbol{I}_N$-$\delta\boldsymbol{W}$) needs to be inverted for every draw, which will be rather time-consuming and even might break down due to memory problems in case N is large. The other approach, proposed by LeSage and Pace (2009, pp. 114–115), is to exploit the decomposition shown in Eq. (2.14) and to store the traces of the matrices $\boldsymbol{I}$ up to and including $\boldsymbol{W}^{100}$ on the right-hand side of (2.14) in advance. The calculation of the direct and indirect effects then no longer requires the inversion of the matrix ($\boldsymbol{I}_N$-$\delta\boldsymbol{W}$) for every parameter combination drawn from the variance–covariance matrix in (2.15), but only a matrix operation based on the stored traces which, as a result, does not require much computational effort.

---

9 The default value is 1,000, but for models with large N this number might be decreased.

2.8 Software

Software packages and/or routines to estimate spatial econometric models are Stata, Geoda, R and Matlab. The latter three are all freely downloadable. The results to be reported in the next section have been estimated using Matlab routines. At www.spatial-econometrics.com, the routines SAR, SEM and SAC can be downloaded, written by James LeSage, to estimate to SAR, SEM and SAC models, respectively. By changing the argument $\boldsymbol{X}$ of these routines into [$\boldsymbol{X}$ $\boldsymbol{WX}$] it is also possible to estimate the SDM, SDEM and GNS models.

One disadvantage of the routines made available at this Web site is that the Jacobian term, $\ln|\boldsymbol{I}_N\text{-}\delta\boldsymbol{W}|$, $\ln|\boldsymbol{I}_N\text{-}\lambda\boldsymbol{W}|$ or both, in the log-likelihood functions of these models is approached by a numerical approach. To overcome potential numerical difficulties one might face in evaluating the log determinant, Pace and Barry (1997) and Barry and Pace (1999) propose computing this determinant *once* over a grid of values for the parameter δ (λ) ranging from $1/r_{min}$ to $1/r_{max}$ prior to estimation. This only requires the determination of the smallest and largest characteristic root of $\boldsymbol{W}$. They suggest a grid based on 0.001 increments for δ over the feasible range. Given these predetermined values for the determinant of ($\boldsymbol{I}_N$-$\delta\boldsymbol{W}$), one can quickly evaluate the log determinant of ($\boldsymbol{I}_N$-$\delta\boldsymbol{W}$) for a particular value of δ. To compute the log determinant over the feasible range for small values of N (< 500), they compute $\Sigma_i \log|\varsigma_{ii}|$, where ς_{ii} (i = 1,…, N) denotes the diagonal elements of the upper triangular LU decomposition matrix of ($\boldsymbol{I}_N$-$\delta\boldsymbol{W}$). When the sparse structure of the spatial weights matrix is exploited, the required computation time of this decomposition can be reduced from order N^3 to order N^2. For larger values of N (≥ 500), they suggest approaching the log determinant for a particular value of δ over the feasible range by

$$\frac{1}{J}\sum_{j=1}^{J}\left[-N\sum_{k=1}^{H}\left(z_j'\boldsymbol{W}^k z_j \Big/ z_j' z_j\right)\left(\delta^k/k\right)\right] \tag{2.18}$$

where z_j denotes an $N \times 1$ vector of independent standard normal variates. The precision of this estimate can be manipulated by means of the tuning parameters J (the number of simulations generated over which the estimate is averaged) and H (the number of elements in the sum of ratios of quadratic forms). The required computation time of this simulation approach can be reduced to order $N\log N$ and allows for the estimation of models with very large numbers of observations in the cross-sectional domain.

The disadvantage of this approach is that the parameter estimate of δ (λ) and therefore of $\boldsymbol{\beta}$ changes slightly every time the routine is run again. Generally, researchers do not appreciate this. This problem can be avoided by calculating the log determinant by

$$\ln|\boldsymbol{I}-\delta\boldsymbol{W}| = \sum\nolimits_i \ln(1-\delta\omega_i) \tag{2.19}$$

where ω_i (i = 1,…,N) denote the characteristic roots of $\boldsymbol{W}$. Generally, this calculation works well for values of N smaller than 1,000. At www.regroningen.nl the routines SARp, SEMp and SACp have been made available, written by Paul Elhorst, to estimate the SAR, SEM and SAC models based on (2.19). Furthermore, by changing the argument $\boldsymbol{X}$ of these routines into [$\boldsymbol{X}$ $\boldsymbol{WX}$] it is also possible to estimate the SDM, SDEM and GNS models. The advantage of these routines is that the estimation results will be exactly the same every time this routine is run.[10] The routine "demo_crime_rates" posted at this Web site can be used to reproduce the estimation results reported in Tables 2.2 and 2.3 in the next section. By changing the specification of $\boldsymbol{Y}$, $\boldsymbol{X}$, $\boldsymbol{W}$ and N in these routines and by reading a different data set, researchers can use this file to estimate these models for their own research problems.

2.9 Empirical Illustration

To demonstrate the performance of the different spatial econometric models in an empirical setting, Anselin's (1988) cross-sectional dataset of 49 Columbus, Ohio neighborhoods is used to explain the crime rate as a function of household income and housing values. The spatial weights matrix $\boldsymbol{W}$ is specified as a row-normalized binary contiguity matrix, with elements $w_{ij} = 1$ if two spatial neighborhoods share a common border, and zero otherwise. It should be stressed that this specification of the spatial weights matrix is also used in Anselin (1988). The estimation results are reported in Table 2.2 and the direct and spatial spillover effects in Table 2.3. Eight different models are considered. The GNS model includes all types of interaction effects, while the other models ignore one or more interaction effects. If a particular entry in Table 2.2 is empty, the interaction effect reported in the left column is not present in the model. The parameter estimates are obtained by applying ML.

One of the main questions is which model best describes the data. One of the criterions that may be used for this purpose is the likelihood ratio (LR) test based on the log-likelihood function values of the different models. The LR test is based on minus two times the difference between the value of the log-likelihood function in the restricted model and the value of the log-likelihood function of the unrestricted model: $-2*(\text{logL}_{\text{restricted}} - \text{logL}_{\text{unrestricted}})$. This test statistic has a Chi squared distribution with degrees of freedom equal to the number of restrictions imposed.

[10] One of the constants in the log-likelihood function of the routines of James LeSage is $\ln(\pi)$, while this should be $\ln(2\pi)$. This error is probably based on Anselin's (1988) textbook, where the same mistake is made. See, e.g., Eqs. (6.15), (8.4), and p. 181. This innocent error has been removed from the SARp, SEMp and SACp routines.

Table 2.2 Model comparison of the estimation results explaining the crime rate

	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS
Intercept	0.686**	0.451**	0.599**	0.750**	0.478**	0.428**	0.735**	0.509
	(14.49)	(6.28)	(11.32)	(11.32)	(4.83)	(3.38)	(8.37)	(0.75)
Income	−1.597**	−1.031**	−0.942**	−1.109**	−1.026**	−0.914**	−1.052**	−0.951**
	(−4.78)	(−3.38)	(−2.85)	(−2.97)	(−3.14)	(−2.76)	(−3.29)	(−2.16)
House value	−0.274**	−0.266**	−0.302**	−0.290**	−0.282**	−0.294**	−0.276**	−0.286**
	(−2.65)	(−3.01)	(−3.34)	(−2.86)	(−3.13)	(−3.29)	(−3.02)	(−2.87)
W * Crime rate		0.431**			0.368*	0.426**		0.315
		(3.66)			(1.87)	(2.73)		(0.33)
W * Income				−1.371**		−0.520	−1.157**	−0.693
				(−2.44)		(−0.92)	(−2.00)	(−0.41)
W * House value				0.192		0.246	0.112	0.208
				(0.96)		(1.37)	(0.56)	(0.73)
W * Error term			0.562**		0.166		0.425**	0.154
			(4.19)		(0.56)		(2.69)	(0.15)
R^2	0.552	0.652	0.651	0.609	0.651	0.665	0.663	0.651
Log-Likelihood	13.776	43.263	42.273	17.075	43.419	44.260	44.069	44.311

**Significant at 5 %; *Significant at 10 %; T-values in parentheses, W = Binary contiguity matrix

Table 2.3 Model comparison of the marginal effects of the explanatory variables on the crime rate

	OLS	SAR	SEM	SLX	SAC	SDM	SDEM	GNS
Direct effects								
Income	−1.597**	−1.086**	−0.942**	−1.109**	−1.063**	−1.024**	−1.052**	−1.032**
	(−4.78)	(−3.44)	(−2.85)	(−2.97)	(−3.25)	(−3.19)	(−3.29)	(−3.28)
House value	−0.274**	−0.280**	−0.302**	−0.290**	−0.292**	−0.279**	−0.276**	−0.277
	(−2.65)	(−2.96)	(−3.34)	(−2.86)	(−3.10)	(−3.13)	(−3.02)	(0.32)
Indirect or spatial spillover effects								
Income		−0.727*		−1.371**	−0.560	−1.477*	−1.157**	−1.369
		(−1.95)		(−2.44)	(−0.18)	(−1.83)	(−2.00)	(0.02)
House value		−0.188*		0.192	−0.154	0.195	0.112	0.163
		(−1.71)		(0.96)	(−0.39)	(0.66)	(0.56)	(−0.03)

**Significant at 5 %; *Significant at 10 %; T-values in parentheses, W = Binary contiguity matrix

The log-likelihood function value of the OLS model increases from 13.776 to 17.075 when this model is extended to include exogenous interaction effects (***WX***), known as the SLX model. The LR-test of the SLX model versus the OLS model takes the value of 6.598 with 2 degrees of freedom (df), while the 5 % critical value is 6.0. This implies that the OLS model needs to be rejected in favor of the SLX model. However, if the OLS is extended to include endogenous interaction effects (***WY***) or interaction effects among the error terms (***Wu***), the log-likelihood function value increases even more, even though in these two cases only one interaction effect is added to the model. Whether it is this SAR or SEM model that better describes the data is difficult to say, since these two models are not nested. One solution is to test whether the spatial lag model or the spatial error model is more appropriate to describe the data, provided that the OLS model is taken as point of departure. For this purpose, one may use the classic LM-tests proposed by Anselin (1988), or the robust LM-tests proposed by Anselin et al. (1996).[11] Both the classic and the robust tests are based on the residuals of the OLS model and follow a Chi squared distribution with 1 degree of freedom. Using the classic tests, both the hypothesis of no spatially lagged dependent variable and the hypothesis of no spatially autocorrelated error term must be rejected at five per cent significance; the LM test for the spatial lag amounts to 9.36 and for the spatial error to 5.72. When using the robust tests, the hypothesis of no spatially lagged dependent variable must still be rejected, though only at ten per cent significance, whereas the hypothesis of no spatially autocorrelated error term can no longer be rejected; the robust LM test for the spatial lag amounts to 3.72 and for the spatial error to 0.08. This indicates that on the basis of these robust LM tests the spatial lag model is more appropriate.

Another solution is to consider the SAC model, which considers both endogenous interaction effects and interaction effects among the error terms, and therefore nests both the SAR and SEM models. The SAC model produces coefficient estimates of the ***WY*** and the ***Wu*** variables that are not significantly different from their counterparts in the SAR model and the SEM model, respectively.[12] Similarly, the LR-test of the SAC model versus the SAR model takes the value of 0.312 with 1 df, and the LR-test of the SAC model versus the SEM model the value of 2.292 with 1 df, while the 5 % critical value in both cases is 3.84. This implies that it is difficult to choose among these three models. However, since the coefficient of ***WY*** is significant in the SAC model, whereas the coefficient of ***Wu*** is not, and the log-likelihood function value of the SAR model is higher than that of the SEM model, the SAR model seems to be the better choice.

[11] The latter tests are called robust because the existence of one type of spatial dependence does not bias the test for the other type of spatial dependence.

[12] The coefficient of the spatially autocorrelated error term in the SAC model amounts to 0.166. The corresponding t-value is so low that this coefficient plus two times its standard error also covers the coefficient estimate of the spatially autocorrelated error term in the SEM model of 0.562. The fact that the latter is significant does not change this conclusion.

Another way to look at the SAR, SEM and SLX models, on their turn, is to consider the SDM model, since the SDM model nests these three models. The SDM appears to outperform the SLX model (LR-test 54.370, 2 df, critical value 5.99), but not the SAR model (LR-test 1.994, 1 df, critical value 3.84) and the SEM model (LR-test 3.974, 2 df, critical value 5.99). Alternatively, one might consider the SDEM model which also nests the SLX and the SEM models. The SDEM model also appears to outperform the SLX model (LR-test 53.998, 1 df, critical value 3.84) but not the SEM model (LR-test, 3.592, 1 df, critical value 3.84). Whether it is the SDM model or the SDEM model that better describes the data is difficult to say, since these two models are not nested. Unfortunately, estimation of the GNS model which nests these two models does not provide an answer. The increase of the log-likelihood function value when estimating this model is so small that, on the basis of the results reported in Table 2.2, it is impossible to draw any conclusion as to whether it is SDM, SDEM or GNS that best describes the data. In contrast to the SAC model, the extension to the GNS model also provides no answer whether endogenous interaction or error correlation effects are more important.

We now consider the direct and indirect effects estimates of the different explanatory variables (see Sect. 2.7) to see whether they can be used as an additional mean to select the best model. The general pattern that emerges from Table 2.3 is the following. First, the differences between the direct effects and the coefficient estimates reported in Table 2.2 are relatively small. In the OLS, SEM, SLX and SDEM models they are exactly the same by construction; in the SAR, SDM, SAC and the GNS models they may be different due to the endogenous interaction effects ***WY***. These interaction effects cause feedback effects, i.e., impacts affecting crime rates in certain neighborhoods that pass on to surrounding neighborhoods and back to the neighborhood instigating the change. For example, the direct effect of the income variable in the GNS model amounts to −1.032, while the coefficient estimate of this variable is −0.951. This implies that the feedback effect is −1.032−(−0.951) = −0.081. This feedback effect corresponds to 8.5 % of the coefficient estimate.

Second, the differences between the direct effects estimates in the different models appear to be relatively small. The direct effects of the income variable range between −0.942 in the SEM model and −1.109 in the SLX model. Only in the OLS model the magnitude of the direct effect of −1.597 is much greater. Just as the LM and LR test results, it indicates that the OLS model needs to be rejected. Since this model accounts neither for spatial interaction effects nor for spatial spillover effects, the direct effect is overestimated (in absolute value). Similarly, the coefficient of the house value variable ranges between −0.274 in the OLS model and −0.302 in the SEM model. Overall, it seems as if it does not matter which model is used to obtain the direct effects estimates. Also the t-values do not differ to any great extent, except for the t-value of the direct effect generated for the house value variable. There are two explanations for this. One is that the significance level of the spatial autoregressive coefficient of the ***WY*** variable in the GNS models falls considerably, because this variable competes with the spatial autocorrelation coefficient of the ***Wu*** variable. This result also occurred in the SAC model. If endogenous interaction effects and interaction effects among the error terms are separated from each other, both coefficients turn out to be significant, but if they are combined both become insignificant. Another explanation is that the t-values of the coefficient estimates in the different models are relatively stable, except for the GNS model. The t-values of the variables in this model have the tendency to go down.

In contrast to the direct effects estimates, the differences between the spillover effects are extremely large. Still, one can observe some general patterns. The OLS, SAR, SEM and SAC models produce no or wrong spillover effects compared to the SDM, SDEM and GNS models. For example, whereas the spillover effect of the house value variable is positive in the SLX, SDM, SDEM and GNS models, it is zero by construction in the OLS and SEM models, negative in the SAC model, and negative and weakly significant in the SAR model. The negative and also weakly significant effect in the SAR model can be explained by the fact that this model suffers from the problem that the ratio between the spillover effect and the direct effect is the same for every explanatory variable. Consequently, this model is too rigid to model spillover effects adequately. The negative but insignificant effect in the SAC model can be explained by the fact that this model resembles the SAR model: the spatial autocorrelation coefficient of ***Wu*** appears to be insignificant, whereas the spatial autoregressive coefficient of ***WY*** does not, as a result of which the SAC model is hampered by the same problem as the SAR model. This was pointed out earlier in Table 2.1; mathematically, the SAR and SAC models share the same direct and indirect effects estimates.

The spillover effects produced by the SLX, SDM, SDEM and GNS models are more or less comparable to each other. In these models, the spillover effect of the income variable ranges from −1.157 to −1.477 and of the house value variable from 0.112 to 0.192. By contrast, the t-values do not. The t-values in the SLX model are relatively high. This can be explained by the fact that the SLX has been rejected in favor of the SDM and the SDEM models based on the LR tests. Furthermore, the t-values in the GNS model are relatively low. As recently pointed out by Gibbons and Overman (2012), the explanation for this finding is that interaction effects among the dependent variable on the one hand and interaction effects among the error terms on the other hand are only weakly identified. Considering them both, as in the GNS model, strengthens this problem; it leads to a model that is overparameterized, as a result of which the significance levels of all variables tend to go down. This finding is worrying since the interpretation of both types of interaction effects is completely different. A model with endogenous interaction effects posits that the crime rate in one neighborhood depends on that in other neighborhoods, and on a set of neighborhood characteristics. By contrast, a model with interaction effects among the error terms assumes that the crime rate in one neighborhood depends on a set of observed neighborhood characteristics and unobserved characteristic omitted from the model that neighborhoods have in common. Nevertheless, both models appear to produce spillover effects that are comparable to each other, both in terms of magnitude and significance.

2.9.1 Conclusion

The conclusion from the empirical analysis is twofold. First, for various reasons the OLS, SAR, SEM, SLX, SAC and GNS models need to be rejected. The OLS and SLX models are outperformed by other, more general models. The spillover effects of the SEM model are zero by construction, while the results of more general models show that the spillover effect of the income variable is significant. The SAR and SAC models suffer from the problem that the ratio between the spillover effect and the direct effect is the same for every explanatory variable. Consequently, the spillover effect of the housing value variable gets a wrong sign. Finally, the GNS model is overparameterized, as a result of which the t-values of the coefficient estimates and the effects estimates have the tendency to go down. In sum, only the SDM and SDEM model produce acceptable results. Second, it is not clear which of these two models best describes the data. Even though both models produce spillover effects that are comparable to each other, both in terms of magnitude and significance, this is worrying since these two models have a different interpretation.

2.10 Conclusion

Originally, the central focus of spatial econometrics has been on the spatial lag model (SAR) and the spatial error model (SEM) with one type of interaction effect. The results shown in this chapter make clear that this approach is too limited and that the focus should shift to the spatial Durbin model (SDM) and the spatial Durbin error model (SDEM). At the same time, new test procedures should be developed to choose among these two models, which is difficult because both models tend to produce spillover effects that are comparable to each other in terms of magnitude and significance, and because interaction effects among the dependent variable on the one hand and interaction effects among the error terms on the other hand are only weakly identified. Precisely for this reason, the general nesting spatial (GNS) model is not of much help either. It generally leads to a model that is overparameterzed, as a result of which the significance levels of the variables tend to go down.

Recently, Gibbons and Overman (2012) criticized the SAR, SEM and SDM models. They demonstrate that the reduced form of these models is similar to a model with first, second and higher order exogenous interaction effects, and argue that this reduced form can hardly be distinguished from the SLX model that only contains first order exogenous interaction effects.

Another major weakness of spatial econometric models is that the spatial weights matrix $\boldsymbol{W}$ needs to be specified in advance, although there are exceptions, and that economic theory underlying spatial econometric applications often has little to say about the specification of $\boldsymbol{W}$. For this reason, it has become common practice to investigate whether the results are robust to the specification of $\boldsymbol{W}$, or to test different specifications against each other using the log-likelihood function value, Bayesian posterior model probabilities, or the J-test. In this respect, Corrado and Fingleton (2012) strongly argue for the use of more substantive theory in empirical spatial econometric modeling, especially regarding the modeling of $\boldsymbol{W}$.

In view of these critical notes, Halleck Vega and Elhorst (2012) are currently doing research on finding a better and broader[13] modeling strategy to determine the spatial econometric model, including the spatial weights matrix $\boldsymbol{W}$, that best describes the data.

References

Allers MA, Elhorst JP (2005) Tax mimicking and yardstick competition among governments in the Netherlands. Int Tax Public Finance 12(4):493–513

Anselin L (1986) Non-nested tests on the weight structure in spatial autoregressive models. J Reg Sci 26:267–284

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Anselin L, Bera AK, Florax R, Yoon MJ (1996) Simple diagnostic tests for spatial dependence. Reg Sci Urban Econ 26(1):77–104

Badinger H, Egger P (2011) Estimation of higher-order spatial autoregressive cross-section models with heteroskedastic disturbances. Pap Reg Sci 90:213–235

Barry RP, Pace RK (1999) Monte Carlo estimates of the log determinant of large sparse matrices. Linear Algebra Appl 289:41–54

Beach CM, MacKinnon JG (1978) Full maximum likelihood estimation of second-order autoregressive error models. J Econometrics 7:187–198

Bell KP, Bockstael NE (2000) Applying the generalized-moments estimation approach to spatial problems involving microlevel data. Rev Econ Stat 82:72–82

Bordignon M, Cerniglia F, Revelli F (2003) In search of yardstick competition: a spatial analysis of Italian municipality property tax setting. J Urban Econ 54:199–217

Brandsma AS, Ketellapper RH (1979) A biparametric approach to spatial autocorrelation. Environ Plan A 11:51–58

Brueckner JK (2003) Strategic interaction among local governments: An overview of empirical studies. Int Reg Sci Rev 26(2):175–188

Burridge P (2012) Improving the J test in the SARAR model by likelihood estimation. Spatial Econ Anal 7(1):75–107

Burridge P, Fingleton B (2010) Bootstrap inference in spatial econometrics: the J-test. Spatial Econ Anal 5:93–119

Case AC (1991) Spatial patterns in household demand. Econometrica 59:953–965

Corrado L, Fingleton B (2012) Where is the economics in spatial econometrics? J Reg Sci 52(2):210–239

Dall'Erba S, Percoco M, Piras G (2008) Service industry and cumulative growth in the regions of Europe. Entrepreneurship Reg Dev 21:333–349

Drukker DM, Egger P, Prucha IR (2013) On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. Econometric Rev 32(5–6):686–733

[13] Broader in the sense that it is based on theory, statistics, flexibility and possibilities to parameterize $\boldsymbol{W}$.

Durbin J (1960) Estimation of parameters in time-series regression models. J Roy Stat Soc B 22:139–153
Elhorst JP (2001) Dynamic models in space and time. Geogr Anal 33(2):119–140
Elhorst JP (2010) Applied spatial econometrics: raising the bar. Spat. Econ. Anal. 5(1):9–28
Elhorst JP, Fréret S (2009) Evidence of political yardstick competition in France using a two-regime spatial Durbin model with fixed effects. J Reg Sci 49:931–951
Elhorst JP, Piras G, Arbia G (2010) Growth and convergence in a multi-regional model with space-time dynamics. Geogr Anal 42(3):338–355
Elhorst JP, Lacombe DJ, Piras G (2012) On model specification and parameter space definitions in higher order spatial econometrics models. Reg Sci Urban Econ 42(1–2):211–220
Ertur C, Koch W (2007) Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence. J Appl Econometrics 22(6):1033–1062
Fingleton B, Le Gallo J (2007) Finite sample properties of estimators of spatial models with autoregressive, or moving average disturbances and system feedback. Annales d'économie et de statistique 87(88):39–62
Fingleton B, Le Gallo J (2008) Estimating spatial models with endogenous variables, a spatial lag en spatially dependent disturbances: finite sample properties. Pap Reg Sci 87:319–339
Folmer H, Oud J (2008) How to get rid of W: a latent variables approach to modelling spatially lagged variables. Environ Plan A 40:2526–2538
Gibbons S, Overman HG (2012) Mostly pointless spatial econometrics? J Reg Sci 52(2):172–191
Griffith DA, Arbia G (2010) Detecting negative spatial autocorrelation in georeferenced random variables. Int J Geogr Inf Sci 24:417–437
Halleck Vega S, Elhorst JP (2012) On spatial econometric models, spillover effects, and W. University of Groningen, Working paper
Harris R, Moffat J, Kravtsova V (2011) In Search of W. Spatial Econ Anal 6(3):249–270
Hendry DF (1995) Dynamic econometrics. Oxford University Press, Oxford
Hepple LW (1995) Bayesian techniques in spatial and network econometrics: 2. Computational methods and algorithms. Environ Plan A 27:615–644
Kelejian HH (2008) A spatial J-test for model specification against a single or a set of non-nested alternatives. Lett Spatial Res Sci 1(1):3–11
Kelejian HH, Prucha IR (1998) A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J Real Estate Finance Econ 17(1):99–121
Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. Int Econ Rev 40(2):509–533
Kelejian HH, Prucha IR (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. J Econometrics 157(1):53–67
Kelejian HH, Robinson DP (1995) Spatial correlation; a suggested alternative to the autoregressive model. In: Anselin R, Florax RJGM (eds) New directions in spatial econometrics. Springer, Berlin, pp 75–95
Kelejian HH, Prucha IR, Yuzefovich Y (2004) Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: large and small sample results. In: LeSage JP, Pace K (eds) Spatial and spatiotemporal econometrics. Elsevier, Amsterdam, pp 163–198
Lacombe DJ (2004) Does econometric methodology matter? An analysis of public policy using spatial econometric techniques. Geogr Anal 36:105–118
Lee LF (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Rev 22(4):307–335
Lee LF (2004) Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72(6):1899–1925
Lee LF (2007) Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. J Econometrics 140:333–374
Lee LF, Liu X (2010) Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory 26:187–230

Leenders RTAJ (2002) Modeling social influence through network autocorrelation: constructing the weight matrix. Soc Netw 24(1):21–47
LeSage JP (1997) Bayesian estimation of spatial autoregressive models. Int Reg Sci Rev 20:113–129
LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press, Taylor & Francis Group, Boca Raton
LeSage JP, Pace RK (2011) Pitfalls in higher order model extensions of basic spatial regression methodology. Rev Reg Stud 41(1):13–26
Liu X, Lee LF (2013) Two-stage least squares estimation of spatial autoregressive models with endogenous regressors and many instruments. Econometric Rev 32(5–6):734–753
McMillen D, Singell L, Waddell G (2007) Spatial competition and the price of college. Econ Inq 45:817–833
Mood AM, Graybill F, Boes DC (1974) Introduction to the theory of statistics, 3rd edn. McGraw-Hill, Tokyo
Ord K (1975) Estimation methods for models of spatial interaction. J Am Stat Assoc 70:120–126
Ord JK (1981) Towards a theory of spatial statistics: a comment. Geogr Anal 13:91–93
Pace RK, Barry R (1997) Quick computation of spatial autoregressive estimators. Geogr Anal 29(3):232–246
Seldadyo H, Elhorst JP, De Haan J (2010) Geography and governance: Does space matter? Pap Reg Sci 89:625–640
Sherrell N (1990) The estimation and specification of spatial econometric models. Ph.D. thesis, University of Bristol (unpublished)
Stakhovych S, Bijmolt THA (2009) Specification of spatial models: a simulation study on weights matrices. Pap Reg Sci 88:389–408
Ward MD, Gleditsch KS (2008) Spatial regression models. Sage Publications, Los Angeles, Series: Quantitative Applications in the Social Sciences 155

Chapter 3
Spatial Panel Data Models

Abstract This chapter provides a survey of the specification and estimation of spatial panel data models. Five panel data models commonly used in applied research are considered: the fixed effects model, the random effects model, the fixed coefficients model, the random coefficients model, and the multilevel model. Today a (spatial) econometric researcher has the choice of many models. First, he should ask himself whether or not, and, if so, which type of spatial interaction effects should be accounted for. Second, he should ask himself whether or not spatial-specific and/or time-specific effects should be accounted for and, if so, whether they should be treated as fixed or as random effects. A selection framework is demonstrated to determine which of the first two types of spatial panel data models considered in this chapter best describes the data. The well-known Baltagi and Li (2004) panel dataset, explaining cigarette demand for 46 US states over the period 1963 to 1992, is used to illustrate this framework in an empirical setting.

Keywords Spatial panels · Estimation · Bias correction · Fixed vs. Random · Model comparison · Spatial spillover effects · Cigarette demand

3.1 Introduction

The spatial econometrics literature has exhibited a growing interest in the specification and estimation of econometric relationships based on spatial panels since the turn of this century. This interest can be explained by the increased availability of more data sets in which a number of spatial units are followed over time, and by the fact that panel data offer researchers extended modeling possibilities as compared to the single equation cross-sectional setting, which was the primary focus of the spatial econometrics literature for a long time.

The extension of the general nesting spatial model for a cross-section of N observations, presented in Eq. (2.5), to a space-time model for a panel of N observations over T time periods is obtained by adding a subscript t, which runs from 1 to T, to the variables and the error terms of that model

J. P. Elhorst, *Spatial Econometrics*, SpringerBriefs in Regional Science,
DOI: 10.1007/978-3-642-40340-8_3,

$$\boldsymbol{Y}_t = \delta \boldsymbol{W}\boldsymbol{Y}_t + \alpha \boldsymbol{\iota}_N + \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\theta} + \boldsymbol{u}_t \tag{3.1a}$$

$$\boldsymbol{u}_t = \lambda \boldsymbol{W}\boldsymbol{u}_t + \boldsymbol{\varepsilon}_t \tag{3.1b}$$

This model can be estimated along the same lines as the cross-sectional model, provided that all notations are adjusted from one cross-section to T cross-sections of N observations. The same applies to the other spatial econometric models which can be obtained by imposing restrictions on one or more parameters of the parameters in the GNS model: OLS, SAR, SEM, SLX, SAC, SDM, and SDEM. These restrictions are similar to those shown in Fig. 2.1.

The main objection to pooling the data like this is that the resulting model does not account for spatial and temporal heterogeneity. Spatial units are likely to differ in their background variables, which are usually space-specific time-invariant variables that do affect the dependent variable, but which are difficult to measure or hard to obtain. Examples of such variables abound: one spatial unit is located at the seaside, the other just at the border; one spatial unit is a rural area located in the periphery of a country, the other an urban area located in the center; norms and values regarding labor, crime and religion in one spatial unit might differ substantially from those in another unit, etc. Failing to account for these variables increases the risk of obtaining biased estimation results. One remedy is to introduce a variable intercept μ_i representing the effect of the omitted variables that are peculiar to each spatial unit considered. In sum, spatial specific effects control for all time-invariant variables whose omission could bias the estimates in a typical cross-sectional study.

Similarly, the justification for adding time-period specific effects (ξ_t) is that they control for all spatial-invariant variables whose omission could bias the estimates in a typical time-series study (Arrelano 2003; Hsiao 2003; Baltagi 2005). Examples of such variables also exist: one year is marked by economic recession, the other by a boom; changes in legislation or government policy can significantly affect the functioning of an economy as from the date of implementation, as a result of which before and after observations might be significantly different from one another.

The space–time model in (3.1) extended with spatial specific and time-period specific effects reads as

$$\boldsymbol{Y}_t = \rho \boldsymbol{W}\boldsymbol{Y}_t + \alpha \boldsymbol{\iota}_N + \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\theta} + \boldsymbol{\mu} + \xi_t \boldsymbol{\iota}_N + \boldsymbol{u}_t \tag{3.2a}$$

$$\boldsymbol{u}_t = \lambda \boldsymbol{W}\boldsymbol{u}_t + \boldsymbol{\varepsilon}_t \tag{3.2b}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^T$. The spatial and time-period specific effects may be treated as fixed effects or as random effects. In the fixed effects model, a dummy variable is introduced for each spatial unit and for each time period (except one to avoid perfect multicollinearity), while in the random effects model, μ_i and ξ_t are treated as random variables that are independently and identically distributed with zero mean and variance σ_μ^2 and σ_ξ^2, respectively. Furthermore, it is assumed that the random variables μ_i, ξ_t and ε_{it} are independent of each other.

The discussion on the stationarity conditions on the spatial weights matrix $\boldsymbol{W}$, the normalization procedure of $\boldsymbol{W}$, the parameter space on which δ and λ are defined, and the direct and indirect effects discussed in the previous chapter also apply to the models that will be presented in this chapter. One difference is that the assumption that the row and column sums of $\boldsymbol{W}$ before row-normalization should not diverge to infinity at a rate equal to or faster than the rate of the sample size N, which is made in the cross-sectional setting, is not explicitly made in a panel data setting. In this respect, a couple of studies have paid attention to a spatial weights matrix with equal weights, that is, a matrix where all off-diagonal elements are defined as $1/(N-1)$. Lee (2004) and Kelejian and Prucha (2002) prove that this matrix leads to inconsistent parameter estimates in a cross-sectional setting, since the ratio between the row sums of this matrix before normalization and the sample size, $(N-1)/N$, converges to one instead of zero as N goes to infinity. By contrast, in a panel data setting, this spatial weights matrix causes no problems, provided that time-period effects are not included (see Kelejian and Prucha 2002; Kelejian et al. 2006). However, if time-period fixed effects are also considered, the estimators to be discussed in this chapter become inconsistent again if this situation occurs.

The organization of this chapter is as follows. First, the estimation procedure of standard panel data models, the fixed effects and random effects model, without any spatial interaction effects is discussed in Section 3.2. Next, Section 3.3 outlines the modifications that are needed to estimate the fixed effects model and the random effects model extended to include endogenous interaction effects or interaction effects among the error terms. It is to be noted that these two extensions also cover the SDM and SDEM models. The only thing that needs to be changed is extending the set of explanatory variables to include exogenous interaction effects, $\boldsymbol{X} = [\boldsymbol{X}\ \ \boldsymbol{WX}]$. The estimation method of the SAR model can then be used to estimate the SDM model and of the SEM model to estimate the SDEM model. In contrast to the previous chapter, the SAC and GNS models are not considered since their empirical relevance appeared to be relatively small. Section 3.4 discusses the pros and cons of treating the spatial and time-period specific effects as fixed or random. Section 3.5 deals with issues relevant for model comparison and selection that have not been discussed in the previous chapter, and that require attention when having data in panel. In Section 3.6, a demand model for cigarettes is estimated based on panel data from 46 U.S. states over the period 1963–1992 to empirically illustrate the different spatial econometric models and their effects estimates. This data set is taken from Baltagi (2005) and has been used for illustration purposes in many other studies too. Two routines will be presented to help the researcher choose among different spatial econometric models. The first routine provides (robust) LM tests, generalizing the classic LM-tests proposed by

Burridge (1980) and Anselin (1988) and the robust LM-tests proposed by Anselin et al. (1996) from a cross-sectional setting to a spatial panel setting.[1] The second routine contains a framework to test the spatial lag, the spatial error model, and the spatial Durbin model against each other, as well as a framework to choose among fixed effects, random effects or a model without fixed/random effects.

Although the fixed or random effects model accommodates spatial (and temporal) heterogeneity to a certain extent, the problem remains as to whether the data in such a model are pooled correctly. When spatial heterogeneity is not completely captured by the intercept, a natural generalization is to let the slope parameters of the regressors vary as well. The slope parameters can also be considered fixed or randomly distributed between spatial units. Section 3.7 deals with these fixed and random coefficients models. Section 3.8 continues with multi-level models which consist of a mix of fixed and random coefficients. These models are useful when analyzing data at two or more different levels, such as regions within different countries of the European Union. Section 3.9 considers the spatial SUR model, which are different from fixed and random coefficient models in that the coefficients do not vary over space but over time or over different dependent variables. Finally, Section 3.10 concludes.

3.2 Standard Models for Spatial Panels

To explain the estimation procedure of the fixed and random effects model in this section and the extensions with spatial interaction effects in the next section, the notation in vector form is left for the moment. Instead, a notation in terms of individual observations is used. In addition to this, time-specific effects are left aside. It simplifies notation, while the extension with time-specific effects is straightforward, unless otherwise stated.

A pooled linear regression model with spatial specific effects but without spatial interaction effects reads as

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{\mu}_i + \varepsilon_{it} \tag{3.3}$$

where i is an index for the cross-sectional dimension (spatial units), with $i = 1, \ldots, N$, and t is an index for the time dimension (time periods), with $t = 1, \ldots, T$. In the remainder of this book it is assumed that the data are sorted first by time and then by spatial unit, whereas the classic panel data literature tends to sort the data first by spatial unit and then by time. When y_{it} and $\boldsymbol{x}_{it}$ of these

[1] Baltagi et al. (2003) are the first to consider the testing of spatial interaction effects in a spatial panel data model. They derive a joint LM test which tests for spatial error autocorrelation and spatial random effects, as well as two conditional tests which test for one of these extensions assuming the presence of the other.

T successive cross-sections of N observations are stacked, an $NT \times 1$ vector is obtained for $\boldsymbol{Y}$ and an $NT \times K$ matrix for $\boldsymbol{X}$.

3.2.1 Fixed Effects Model

If the spatial specific effects are treated as fixed effects, the parameters of the model in (3.3) can be estimated in three steps. First, the spatial fixed effects μ_i are eliminated from the regression equation by demeaning the y and x variables. This transformation takes the form

$$y_{it}^* = y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} \quad \text{and} \quad \boldsymbol{x}_{it}^* = \boldsymbol{x}_{it} - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it} \tag{3.4}$$

Second, the transformed regression equation $y_{it}^* = \boldsymbol{x}_{it}^* \boldsymbol{\beta} + \varepsilon_{it}^*$ is estimated by OLS: $\boldsymbol{\beta} = \left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{Y}^*$ and $\sigma^2 = (\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})^T(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})/(NT-N-K)$. This estimator is known as the least squares dummy variables (LSDV) estimator. The main advantage of the demeaning procedure is that the computation of $\boldsymbol{\beta}$ involves the inversion of a $K \times K$ matrix rather than $(K + N) \times (K + N)$ as in (3.3). This would slow down the computation and worsen the accuracy of the estimates considerably for large N.

Instead of estimating the demeaned equation by OLS, it can also be estimated by ML. Since the log-likelihood function of the demeaned equation is

$$\text{LogL} = -\frac{NT}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}^* - \boldsymbol{x}_{it}^*\boldsymbol{\beta}\right)^2 \tag{3.5}$$

the ML estimators of $\boldsymbol{\beta}$ and σ^2 are $\beta = \left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{Y}^*$ and $\sigma^2 = (\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})^T(\boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta})/NT$, respectively. In other words, the ML estimator of σ^2 is slightly different from the LSDV estimator in that it does not correct for degrees of freedom. The asymptotic variance matrix of the parameters is (see Greene 2008, p. 519)

$$\text{Asy.Var}(\boldsymbol{\beta}, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{*T}\boldsymbol{X}^* & 0 \\ 0 & \frac{NT}{2\sigma^4} \end{bmatrix}^{-1} \tag{3.6}$$

Finally, the spatial fixed effects may be recovered by

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \boldsymbol{x}_{it}\boldsymbol{\beta}), \qquad i = 1, \ldots, N \tag{3.7}$$

It should be stressed that the spatial fixed effects can only be estimated consistently when T is sufficiently large, because the number of observations available for the estimation of each μ_i is T. Importantly, sampling more observations in the cross-sectional domain is no solution for insufficient observations in the time domain, since the number of unknown parameters increases as N increases, a situation known as the incidental parameters problem. Fortunately, the inconsistency of μ_i is not transmitted to the estimator of the slope coefficients $\boldsymbol{\beta}$ in the demeaned equation, since this estimator is not a function of the estimated μ_i. Consequently, the incidental parameters problem does not matter when $\boldsymbol{\beta}$ are the coefficients of interest and the spatial fixed effects μ_i are not, which is the case in many empirical studies. Finally, it should be stressed that the incidental parameters problem is independent of the extension of the model with spatial interaction effects.

In case the spatial fixed effects μ_i do happen to be of interest, their standard errors may be computed as the square roots of their asymptotic variances (see Greene 2008, p. 196).

$$\text{Asy.Var}(\hat{\mu}_i) = \frac{\hat{\sigma}^2}{T} + \hat{\sigma}^2 \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_{it} \right) \left(\boldsymbol{X}^{*T} \boldsymbol{X}^{*} \right)^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_{it} \right)^{T} \tag{3.8}$$

An alternative and equivalent formulation of (3.3) is to introduce a mean intercept α, provided that $\Sigma_{\mathrm{i}} \mu_i = 0$. Then the spatial fixed effect μ_i represents the deviation of the i-th spatial unit from the individual mean (see Hsaio 2003, p. 33).

3.2.2 Random Effects Model

To obtain the ML parameter estimates of the random effects model, an iterative two-stage estimation procedure may be used (Breusch 1987). The log-likelihood of the random effects model in (3.3) is

$$\text{LogL} = -\frac{NT}{2} \log\left(2\pi\sigma^2\right) + \frac{N}{2} \log \phi^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it}^{\bullet} - \boldsymbol{x}_{it}^{\bullet} \boldsymbol{\beta}\right)^2, \tag{3.9}$$

where ϕ denotes the weight attached to the cross-sectional component of the data, with $0 \leq \phi^2 = \sigma^2 / \left(T\sigma_\mu^2 + \sigma^2\right) \leq 1$, and the symbol $\bullet$ denotes a transformation of the variables dependent on ϕ

$$y_{it}^{\bullet} = y_{it} - (1-\phi) \frac{1}{T} \sum_{t=1}^{T} y_{it} \quad \text{and} \quad \boldsymbol{x}_{it}^{\bullet} = \boldsymbol{x}_{it} - (1-\phi) \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_{it} \tag{3.10}$$

If $\phi = 0$, this transformation simplifies to the demeaning procedure of Eq. (3.4) and hence the random effects model to the fixed effects model.

Given ϕ, $\boldsymbol{\beta}$ and σ^2 can be solved from their first-order maximizing conditions: $\boldsymbol{\beta} = \left(\boldsymbol{X}^{\bullet T} \boldsymbol{X}^{\bullet}\right)^{-1} \boldsymbol{X}^{\bullet T} \boldsymbol{Y}^{\bullet}$ and $\sigma^2 = (\boldsymbol{Y}^{\bullet} - \boldsymbol{X}^{\bullet} \boldsymbol{\beta})^T (\boldsymbol{Y}^{\bullet} - \boldsymbol{X}^{\bullet} \boldsymbol{\beta}) / NT$. Conversely, ϕ may be estimated by maximizing the concentrated log-likelihood function with respect to ϕ, given $\boldsymbol{\beta}$ and σ^2,

$$\mathrm{LogL} = -\frac{NT}{2}\log\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - (1-\phi)\frac{1}{T}\sum_{t\prime=1}^{T} y_{it\prime} - \left[\boldsymbol{x}_{it} - (1-\phi)\frac{1}{T}\sum_{t\prime=1}^{T}\boldsymbol{x}_{it\prime}\right]\boldsymbol{\beta}\right)^2\right\} + \frac{N}{2}\log\phi^2 \tag{3.11}$$

The use of ϕ^2 instead of ϕ ensures that both the argument of $\log(\phi^2)$ and of $\sqrt{(\phi^2)}$ are positive (see Magnus 1982 for details). The asymptotic variance matrix of the parameters is

$$\mathrm{Asy.Var}(\boldsymbol{\beta}, \phi, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{\bullet T}\boldsymbol{X}^{\bullet} & 0 & 0 \\ 0 & N(1+\frac{1}{\phi^2}) & -\frac{N}{\sigma^2} \\ 0 & -\frac{N}{\sigma^2} & \frac{NT}{2\sigma^4} \end{bmatrix}^{-1} \tag{3.12}$$

One can test whether the spatial random effects are significant by performing a LR test of the hypothesis $\mathrm{H}_0 : \phi = 1$.[2] This test statistic has a Chi squared distribution with one degree of freedom. If the hypothesis is rejected, the spatial random effects are significant.

Finally, it is to be noted that the random effects model may always include a constant term, in which case the number of independent variables is $K + 1$ rather than K.

3.3 Estimation of Spatial Panel Data Models

This section outlines the modifications that are needed to estimate the fixed effects model and the random effects model extended to include a spatially lagged dependent variable or a spatially autocorrelated error. The description spatially lagged dependent variable or shorter spatial lag is synonymous with endogenous interaction effects and the description spatially autocorrelated error or shorter spatial error is synonymous with interaction effects among the error terms.

It is assumed that $\boldsymbol{W}$ is constant over time and that the panel is balanced. It should be noted that the estimators discussed in this chapter can be modified for a spatial weights matrix that changes over time due to changes of economic environments, that is, if the elements of $\boldsymbol{W}$ are based on economic/socioeconomic distances or demographic characteristics. Lee and Yu (2012a) show that the (quasi) ML estimator of spatial dynamic panel data models if the spatial weights

[2] $\phi = 1$ implies $\sigma_\mu^2 = 0$, since σ_μ^2 may be calculated from ϕ by $\sigma_\mu^2 = \frac{1-\phi^2}{\phi^2}\frac{\sigma^2}{T}$.

matrix is time varying due to changes of economic environments is consistent and asymptotically normal. By contrast, even though these estimators can also be modified for unbalanced panels due to missing observations, their asymptotic properties, in the event of missing observations, may become problematic if the reason why data are missing is not known. There are a couple of papers now dealing with missing observations within spatial panels (Pfaffermayr 2009; Wang and Lee 2013), but a general approach is still not available.

The spatial lag model can be specified as

$$y_{it} = \delta \sum_{j=1}^{N} w_{ij} y_{jt} + \boldsymbol{x}_{it}\boldsymbol{\beta} + \mu_i + \varepsilon_{it} \tag{3.13}$$

where w_{ij} is an element of the spatial weights matrix $\boldsymbol{W}$, while the spatial error model reads as

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \mu_i + u_{it} \tag{3.14a}$$

$$u_{it} = \lambda \sum_{j=1}^{N} w_{ij} u_{it} + \varepsilon_{it} \tag{3.14b}$$

3.3.1 Fixed Effects Spatial Lag Model

According to Anselin et al. (2006), the extension of the fixed effects model with a spatially lagged dependent variable raises two complications. First, the endogeneity of $\Sigma_j w_{ij} y_{jt}$ violates the assumption of the standard regression model that $\mathrm{E}\left[\left(\Sigma_j w_{ij} y_{jt}\right)\varepsilon_{it}\right] = 0$. In model estimation, this simultaneity must be accounted for. Second, the spatial dependence among the observations at each point in time may affect the estimation of the fixed effects.

In this section, the ML estimator is derived to account for the endogeneity of $\Sigma_j w_{ij} y_{jt}$. The log-likelihood function of model (3.13) if the spatial specific effects are assumed to be fixed is

$$\begin{aligned} \mathrm{LogL} = & -\frac{NT}{2}\log(2\pi\sigma^2) + T\log|\boldsymbol{I}_N - \delta\boldsymbol{W}| \\ & -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \delta\sum_{j=1}^{N} w_{ij} y_{jt} - \boldsymbol{x}_{it}\boldsymbol{\beta} - \mu_i\right)^2 \end{aligned} \tag{3.15}$$

where the second term on the right-hand side represents the Jacobian term of the transformation from ε to y taking into account the endogeneity of $\Sigma_j w_{ij} y_{jt}$ (Anselin 1988, p. 63).

The partial derivatives of the log-likelihood with respect to μ_i are

$$\frac{\partial \mathrm{LogL}}{\partial \mu_i} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\left(y_{it} - \delta\sum_{j=1}^{N} w_{ij}y_{jt} - \boldsymbol{x}_{it}\boldsymbol{\beta} - \mu_i\right) = 0,\ i = 1,\ldots,N \tag{3.16}$$

When solving μ_i from (3.16), one obtains

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \delta\sum_{j=1}^{N} w_{ij}y_{jt} - \boldsymbol{x}_{it}\boldsymbol{\beta}\right),\ i = 1,\ldots,N \tag{3.17}$$

This equation shows that the standard formula for calculating the spatial fixed effects, Eq. (3.7), applies to the fixed effects spatial lag model in a straightforward manner. Nevertheless, Lee and Yu (2010a, b) show that there are cases, dependent on N and T, in which corrections for some parameters need to be made for the cross-sectional dependence among the observations at each point in time. This is discussed in subsection 3.3.3 below.

Substituting the solution for μ_i into the log-likelihood function, and after rearranging terms, the concentrated log-likelihood function with respect to $\boldsymbol{\beta}$, δ and σ^2 is obtained

$$\begin{aligned}\mathrm{LogL} = &-\frac{NT}{2}\log\left(2\pi\sigma^2\right) + T\log|\boldsymbol{I}_N - \delta\boldsymbol{W}| \\ &-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}^* - \delta\left[\sum_{j=1}^{N} w_{ij}y_{jt}\right]^* - \boldsymbol{x}_{it}^*\boldsymbol{\beta}\right)^2\end{aligned} \tag{3.18}$$

where the asterisk denotes the demeaning procedure introduced in Eq. (3.4).

Anselin and Hudak (1992) have spelled out how the parameters $\boldsymbol{\beta}$, δ and σ^2 of a spatial lag model can be estimated by ML starting with cross-sectional data. This estimation procedure can also be used to maximize the concentrated log-likelihood function in (3.18) with respect to $\boldsymbol{\beta}$, δ and σ^2. The only difference is that the data are extended from a cross-section of N observations to a panel of $N \times T$ observations. This estimation procedure consists of the following steps.

First, stack the observations as successive cross-sections for $t = 1,\ldots,T$ to obtain $NT \times 1$ vectors for $\boldsymbol{Y}^*$ and $(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}^*$, and an $NT \times K$ matrix for $\boldsymbol{X}^*$ of the demeaned variables. Note that these calculations have to be performed only once and that the $NT \times NT$ diagonal matrix $(\boldsymbol{I}_T \otimes \boldsymbol{W})$ does not have to be stored. This would slow down the computation of the ML estimator considerably for large data sets. Second, let $\boldsymbol{b}_0$ and $\boldsymbol{b}_1$ denote the OLS estimators of successively regressing $\boldsymbol{Y}^*$ and $(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}^*$ on $\boldsymbol{X}^*$, and $\boldsymbol{e}_0^*$ and $\boldsymbol{e}_1^*$ the corresponding residuals. Then the ML estimator of δ is obtained by maximizing the concentrated log-likelihood function

$$\mathrm{LogL} = C - \frac{NT}{2}\log\left[\left(\boldsymbol{e}_0^* - \delta\boldsymbol{e}_1^*\right)^T\left(\boldsymbol{e}_0^* - \delta\boldsymbol{e}_1^*\right)\right] + T\log|\boldsymbol{I}_N - \delta\boldsymbol{W}| \tag{3.19}$$

where C is a constant not depending on δ. Unfortunately, this maximization problem can only be solved numerically, since a closed-form solution for δ does not exist. However, since the concentrated log-likelihood function is concave in δ, the numerical solution is unique (Anselin and Hudak 1992). Just as in Chap. 2 (Section 2.8), the Jacobian term might be approached by a numerical approach to speed up computation time. The disadvantage of this approach is that the parameter estimate of δ and therefore of $\boldsymbol{\beta}$ change slightly every time the routine is run again. For this reason, the Matlab routines to estimate spatial panel data models made available at www.regroningen.nl also offer the opportunity to choose between the exact approach (default N < 1000) and the numerical approach (default N > 1000). See Section 2.8 for a more detailed discussion.

Third, the estimators of $\boldsymbol{\beta}$ and σ^2 are computed, given the numerical estimate of δ,

$$\boldsymbol{\beta} = \boldsymbol{b}_0 - \delta \boldsymbol{b}_1 = \left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}[\boldsymbol{Y}^* - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}^*] \tag{3.20a}$$

$$\sigma^2 = \frac{1}{NT}\left(\boldsymbol{e}_0^* - \delta \boldsymbol{e}_1^*\right)^T\left(\boldsymbol{e}_0^* - \delta \boldsymbol{e}_1^*\right) \tag{3.20b}$$

Finally, the asymptotic variance matrix of the parameters is computed for inference (standard errors, t-values). This matrix takes the form (since this matrix is symmetric the upper diagonal elements are left aside)

$$\text{Asy.Var}(\boldsymbol{\beta}, \delta, \sigma^2) = \begin{bmatrix} \frac{\boldsymbol{X}^{*T}\boldsymbol{X}^*}{\sigma^2} & & \\ \frac{\boldsymbol{X}^{*T}\left(\boldsymbol{I}_T \otimes \tilde{\boldsymbol{W}}\right)\boldsymbol{X}^*\boldsymbol{\beta}}{\sigma^2} & T * tr\left(\tilde{\boldsymbol{W}}\tilde{\boldsymbol{W}} + \tilde{\boldsymbol{W}}^T\tilde{\boldsymbol{W}}\right) + \frac{\boldsymbol{\beta}^T\boldsymbol{X}^{*T}\left(\boldsymbol{I}_T \otimes \tilde{\boldsymbol{W}}^T\tilde{\boldsymbol{W}}\right)\boldsymbol{X}^*\boldsymbol{\beta}}{\sigma^2} & \\ 0 & \frac{T}{\sigma^2} tr\left(\tilde{\boldsymbol{W}}\right) & \frac{NT}{2\sigma^4} \end{bmatrix}^{-1} \tag{3.21}$$

where $\tilde{\boldsymbol{W}} = \boldsymbol{W}(\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1}$ and "*tr*" denotes the trace of a matrix. The differences with the asymptotic variance matrix of a spatial lag model in a *cross-sectional* setting (see Anselin and Bera 1998; Lee 2004) are the change in dimension of the matrix $\boldsymbol{X}^*$ from N to $N \times T$ observations and the summation over T cross-sections involving manipulations of the $N \times N$ spatial weights matrix $\boldsymbol{W}$. For large values of N the determination of the elements of the variance matrix may become computationally impossible. In that case the information may be approached by the numerical Hessian matrix using the maximum likelihood estimates of $\boldsymbol{\beta}$, δ and σ^2.

3.3.2 Fixed Effects Spatial Error Model

Anselin and Hudak (1992) have also spelled out how the parameters $\boldsymbol{\beta}$, λ and σ^2 of a linear regression model extended to include a spatially autocorrelated error term can be estimated by ML starting with cross-sectional data. Just as for the spatial lag model, this estimation procedure can be extended to include spatial fixed effects and from a cross-section of N observations to a panel of $N \times T$ observations. The log-likelihood function of model (3.14a, b) if the spatial specific effects are assumed to be fixed is

$$\text{LogL} = -\frac{NT}{2}\log\left(2\pi\sigma^2\right) + T\log|\boldsymbol{I}_N - \lambda\boldsymbol{W}| \\ -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{y_{it}^* - \lambda\left[\sum_{j=1}^{N} w_{ij}y_{jt}\right]^* - \left(\boldsymbol{x}_{it}^* - \lambda\left[\sum_{j=1}^{N} w_{ij}\boldsymbol{x}_{jt}\right]^*\right)\boldsymbol{\beta}\right\}^2 \tag{3.22}$$

Given λ, the ML estimators of $\boldsymbol{\beta}$ and σ^2 can be solved from their first-order maximizing conditions, to get

$$\boldsymbol{\beta} = \left([\boldsymbol{X}^* - \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}^*]^T[\boldsymbol{X}^* - \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}^*]\right)^{-1} \\ \times [\boldsymbol{X}^* - \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}^*]^T[\boldsymbol{Y}^* - \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}^*] \tag{3.23a}$$

$$\sigma^2 = \frac{\boldsymbol{e}(\lambda)^T\boldsymbol{e}(\lambda)}{NT} \tag{3.23b}$$

where $\boldsymbol{e}(\lambda) = \boldsymbol{Y}^* - \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}^* - [\boldsymbol{X}^* - \lambda(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}^*]\boldsymbol{\beta}$. The concentrated log-likelihood function of λ takes the form

$$\text{LogL} = -\frac{NT}{2}\log\left[\boldsymbol{e}(\lambda)^T\boldsymbol{e}(\lambda)\right] + T\log|\boldsymbol{I}_N - \lambda\boldsymbol{W}| \tag{3.24}$$

Maximizing this function with respect to λ yields the ML estimator of λ, given $\boldsymbol{\beta}$ and σ^2. An iterative procedure may be used in which the set of parameters $\boldsymbol{\beta}$ and σ^2 and the parameter λ are alternately estimated until convergence occurs. The asymptotic variance matrix of the parameters takes the form

$$Asy.Var\left(\boldsymbol{\beta}, \lambda, \sigma^2\right) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{*T}\boldsymbol{X}^* & & \\ 0 & T^* tr(\tilde{\tilde{\boldsymbol{W}}}\tilde{\tilde{\boldsymbol{W}}} + \tilde{\tilde{\boldsymbol{W}}}^T\tilde{\tilde{\boldsymbol{W}}}) & \\ 0 & \frac{T}{\sigma^2}tr(\tilde{\tilde{\boldsymbol{W}}}) & \frac{NT}{2\sigma^4} \end{bmatrix}^{-1} \tag{3.25}$$

where $\tilde{\tilde{\boldsymbol{W}}} = \boldsymbol{W}(\boldsymbol{I}_N - \lambda\boldsymbol{W})^{-1}$. The spatial fixed effects can finally be estimated by

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T}(y_{it} - \boldsymbol{x}_{it}\boldsymbol{\beta}), \qquad i = 1, \ldots, N \tag{3.26a}$$

3.3.3 Bias Correction in Fixed Effects Models

The estimation of the fixed effects models is based on the demeaning procedure spelled out in Baltagi (2005). Lee and Yu (2010a) label this procedure the direct approach but show that it will yield biased estimates of (some of) the parameters. Starting with the SAC model, and using rigorous asymptotic theory, they analytically derive the size of these biases. If the model contains spatial fixed effects but no time-period fixed effects, the parameter estimate of σ^2 will be biased if N is large and T is fixed. If the model contains both spatial and time-period fixed effects, the parameter estimates of all parameters will be biased if both N and T are large. By contrast, if T is fixed the time effects can be regarded as a finite number of additional regression coefficients similar to the role of $\boldsymbol{\beta}$. On the basis of these findings, Lee and Yu (2010a) propose two methods to obtain consistent results. Instead of demeaning, they propose an alternative procedure to wipe out the spatial (and time-period) fixed effects, which reduces the number of observations available for estimation by one observation for every spatial unit in the sample, i.e., from NT to $N(T-1)$ observations in case of spatial fixed effects and $(N-1)(T-1)$ observations in case of both spatial and time period fixed effects. This procedure is labeled the transformation approach.

The second approach Lee and Yu propose to obtain consistent results is a bias correction procedure of the parameters estimates obtained by the direct approach based on maximizing the likelihood function that is obtained under the transformation approach. This section adopts the bias correction procedure and translates the biases Lee and Yu (2010a) derived for the SAC model to successively the SAR, SEM, SDM and SDEM models.

First, if the SAR, SEM, SDM and SDEM models contain spatial fixed effects but no time-period fixed effects, the parameter estimate $\hat{\sigma}^2$ of σ^2 obtained by the direct approach will be biased. This bias can easily be corrected (BC) by (Lee and Yu 2010a, Eq. 18)

$$\hat{\sigma}^2_{BC} = \frac{T}{T-1}\hat{\sigma}^2 \tag{3.26b}$$

This bias correction will have hardly any effect if T is large. However, most spatial panels do not meet this requirement. Mathematically, the asymptotic variance matrices of the parameters of the SAR, SEM, SDM and SDEM models do not change as a result of this bias correction. This is the thrust of the bias correction procedure Lee and Yu (2010a) present as a result of Theorem 2 in their paper. Therefore, we may apply the algebraic expressions of the variance matrix when using the direct approach. It concerns the variance matrix in Eq. (3.21) for the SAR model and in Eq. (3.25) for the SEM model. In case of the SDM and the SDEM models, $\boldsymbol{X}$ is replaced by $\boldsymbol{X} = [\boldsymbol{X}\ \boldsymbol{WX}]$, respectively. Since $\hat{\sigma}^2_{BC}$ replaces $\hat{\sigma}^2$ numerically, the standard errors and thus the t-values of the parameter estimates in the SAR, SEM, SDM and SDEM models will change.

Conversely, if the SAR, SEM, SDM and SDEM models contain time-period fixed effects but no spatial fixed effects, the parameter estimate $\hat{\sigma}^2$ of σ^2 obtained by the direct approach can be corrected by

$$\hat{\sigma}^2_{BC} = \frac{N}{N-1}\hat{\sigma}^2 \tag{3.27}$$

This bias correction is taken from Lee et al. (2010), who consider a block diagonal spatial weights matrix where each block represents a group of (spatial) units that interact with each other but not with observations in other groups. Since this setup is equivalent to a spatial panel data model with time dummies where spatial units interact with each other within the same time period but not with observations in other time periods, it might also be used here. From Eq. (3.27), it can be seen that this bias correction will hardly have any effect if N is large, as in most spatial panels.

If the SAR, SEM, SDM and SDEM models contain both spatial and time-period fixed effects, other parameters need to be bias corrected too. Furthermore, the bias correction will be different for each model. The bias correction in the GNS model would take the form

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\delta} \\ \hat{\lambda} \\ \hat{\sigma}^2 \end{bmatrix}_{BC} = \begin{bmatrix} \boldsymbol{1}_K \\ \boldsymbol{1}_K \\ 1 \\ 1 \\ \frac{T}{T-1} \end{bmatrix} \circ \left[\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\theta}} \\ \hat{\delta} \\ \hat{\lambda} \\ \hat{\sigma}^2 \end{bmatrix} - \frac{1}{N}\left[-\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\delta}, \hat{\lambda}, \hat{\sigma}^2\right)\right]^{-1} \begin{bmatrix} \boldsymbol{0}_K \\ \boldsymbol{0}_K \\ \frac{1}{1-\hat{\delta}} \\ \frac{1}{1-\hat{\lambda}} \\ \frac{1}{2\hat{\sigma}^2} \end{bmatrix} \right] \tag{3.28}$$

where $\boldsymbol{\Sigma}\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\delta}, \hat{\lambda}, \hat{\sigma}^2\right)$ represents the expected value of the second-order derivatives of the log-likelihood function multiplied by -1/(NT) (Lee and Yu, 2010a, Eq. 53) and the symbol $\circ$ denotes the element-by-element product of two vectors or matrices (also known as the Hadamard product). The bias correction for the parameters of the other models are obtained by striking out irrelevant rows from the matrix expressions in Eq. (3.28); 2 and 4 for the SAR model, 2 and 3 for the SEM model, 4 for the SDM model and 3 for the SDEM model. The expressions are based on Lee and Yu (2010a, Eq. 34). Mathematically, the asymptotic variance matrices of the parameters of the SAR, SEM, SDM and SDEM models do not change as a result of the bias correction. This is the thrust of the bias correction procedure Lee and Yu (2010a) present as a result of theorems 4 and 5 in their paper. However, since the bias corrected parameter estimates replace the parameter estimates of the direct approach numerically, the standard errors and t-values of the parameter estimates do change.

3.3.4 Random Effects Spatial Lag Model

The log-likelihood of model (3.13) if the spatial effects are assumed to be random is

$$\begin{aligned} \text{LogL} = & -\frac{NT}{2}\log\left(2\pi\sigma^2\right) + T\log|\boldsymbol{I}_N - \delta\boldsymbol{W}| + \frac{N}{2}\log\phi^2 \\ & -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}^{\bullet} - \delta\left[\sum_{j=1}^{N} w_{ij}y_{jt}\right]^{\bullet} - \boldsymbol{x}_{it}^{\bullet}\boldsymbol{\beta}\right)^2 \end{aligned} \tag{3.29}$$

where the symbol · denotes the transformation introduced in Eq. (3.10) dependent on ϕ. Given ϕ, this log-likelihood function is identical to the log-likelihood function of the fixed effects spatial lag model in (3.15). This implies that the same procedure can be used to estimate $\boldsymbol{\beta}$, δ and σ^2 as described above Eqs. (3.19, 3.20a, b), but that the superscript * must be replaced by •. Given $\boldsymbol{\beta}$, δ and σ^2, ϕ can be estimated by maximizing the concentrated log-likelihood function with respect to ϕ

$$\text{LogL} = -\frac{NT}{2}\log\left[\boldsymbol{e}(\phi)^T\boldsymbol{e}(\phi)\right] + \frac{N}{2}\log\phi^2 \tag{3.30}$$

where the typical element of $\boldsymbol{e}(\phi)$ is

$$\begin{aligned}\boldsymbol{e}(\phi)_{it} =& y_{it} - (1-\phi)\frac{1}{T}\sum_{t=1}^{T} y_{it} - \delta\left[\sum_{j=1}^{N} w_{ij}y_{jt} - (1-\phi)\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{N} w_{ij}y_{jt}\right] \\ &- \left[\boldsymbol{x}_{it} - (1-\phi)\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\right]\boldsymbol{\beta}\end{aligned} \tag{3.31}$$

Again an iterative procedure may be used where the set of parameters $\boldsymbol{\beta}$, δ and σ^2 and the parameter ϕ are alternately estimated until convergence occurs. This procedure is a mix of the estimation procedures used to estimate the parameters of the fixed effects spatial lag model and those of the non-spatial random effects model.

The asymptotic variance matrix of the parameters takes the form

$$\text{Asy}\cdot\text{Var}\left(\boldsymbol{\beta},\delta,\theta,\sigma^2\right) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{\bullet T}\boldsymbol{X}^{\bullet} & & & \\ \frac{1}{\sigma^2}\boldsymbol{X}^{\bullet T}(\boldsymbol{I}_T\otimes\tilde{\mathbf{W}})\boldsymbol{X}^{\bullet}\boldsymbol{\beta} & T * tr\left(\tilde{\mathbf{W}}\tilde{\mathbf{W}} + \tilde{\mathbf{W}}^T\tilde{\mathbf{W}}\right) + \frac{1}{\sigma^2}\boldsymbol{\beta}^T\boldsymbol{X}^{\bullet T}\left(\boldsymbol{I}_T\otimes\tilde{\mathbf{W}}^T\tilde{\mathbf{W}}\right)\boldsymbol{X}^{\bullet}\boldsymbol{\beta} & & \\ 0 & -\frac{1}{\sigma^2}tr(\tilde{\mathbf{W}}) & N\left(T+\frac{1}{\varphi^2}\right) & \\ 0 & \frac{T}{\sigma^2}tr(\tilde{\mathbf{W}}) & -\frac{N}{\sigma^2} & \frac{NT}{2\sigma^4} \end{bmatrix}^{-1} \tag{3.32}$$

3.3.5 *Random Effects Spatial Error Model*

The log-likelihood of model (3.14a, b) if the spatial effects are assumed to be random is (Anselin 1988; Elhorst 2003; Baltagi 2005)

$$\begin{aligned}\text{LogL} =& -\frac{NT}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\log|\boldsymbol{V}| + (T-1)\sum_{i=1}^{N}\log|\boldsymbol{B}| \\ &- \frac{1}{2\sigma^2}\boldsymbol{e}^T\left(\frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^T\otimes\boldsymbol{V}^{-1}\right)\boldsymbol{e} - \frac{1}{2\sigma^2}\boldsymbol{e}^T\left(\boldsymbol{I}_T - \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^T\right)\otimes\left(\boldsymbol{B}^T\boldsymbol{B}\right)\boldsymbol{e}\end{aligned} \tag{3.33}$$

where $V = T\varphi I_N + (B^T B)^{-1}$,[3] $B = I_N - \lambda W$ and $e = Y - X\beta$. It is the matrix V that complicates the estimation of this model considerably. First, the Pace and Barry (1997) procedure to overcome numerical difficulties one might face in evaluating $\log|B| = \log|I_N - \lambda W|$ cannot be used to calculate $\log|V| = \log\left|T\varphi I_N + (B^T B)^{-1}\right|$. Second, there is no simple mathematical expression for the inverse of V. Baltagi (2006) solves these problems by considering a random effects spatial error model with equal weights, i.e., a spatial weights matrix W whose off-diagonal elements are all equal to 1/(N–1). Due to this setup, the inverse of V and a feasible GLS estimator of β can be determined mathematically. Furthermore, by considering a GLS estimator the term log|V| in the log-likelihood function does not have to be calculated.

Elhorst (2003) suggests to express $\log|V|$ as a function of the characteristic roots of W based on Griffith (1988, Table 3.1).

$$\log|V| = \log\left|T\varphi I_N + (B^T B)^{-1}\right| = \sum_{i=1}^{N} \log\left[T\varphi + \frac{1}{(1-\lambda\omega_i)^2}\right] \tag{3.34}$$

Furthermore, he suggests adopting the transformation

$$y_{it}^{\circ} = y_{it} - \lambda\sum_{j=1}^{N} w_{ij}y_{jt} + \sum_{j=1}^{N}\left\{\left[p_{ij} - (1-\lambda w_{ij})\right]\frac{1}{T}\sum_{t=1}^{T} y_{jt}\right\} \tag{3.35}$$

and the same for the variables x_{it}, where p_{ij} is an element of an $N \times N$ matrix P such that $P^T P = V^{-1}$. P can be the spectral decomposition of V^{-1}, $P = \Lambda^{-1/2}R$, where R is an $N \times N$ matrix of which the i-th column is the characteristic vector r_i of V, which is the same as the characteristic vector of the spatial weights matrix W (see Griffith 1988, Table 3.1), $R = (r_1, \ldots, r_N)$, and Λ a $N \times N$ diagonal matrix with the ith diagonal element the corresponding characteristic root, $c_i = T\varphi + 1/(1-\lambda\omega_i)^2$. A similar procedure has been adopted by Yang et al. (2006). It is clear that for large N the numerical determination of P can be problematic. However, Hunneman et al. (2007) find that if W is kept symmetric by using one of the alternative normalizations discussed in Section 2.4, this procedure works well within a reasonable amount of time for values of N up to 4000.

[3] Note that $\varphi = \sigma_\mu^2 / \sigma^2$ is different from ϕ in the non-spatial random effects model and in the random effects spatial lag model.

As a result of (3.34) and (3.35), the log-likelihood function simplifies to

$$\begin{aligned}\text{LogL} = &-\frac{NT}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{N}\log\left(1 + T\phi(1-\lambda\omega_i)^2\right)\\ &+ T\sum_{i=1}^{N}\log(1-\lambda\omega_i) - \frac{1}{2\sigma^2}\boldsymbol{e}^{\circ T}\boldsymbol{e}^{\circ}\end{aligned} \tag{3.36}$$

where $\boldsymbol{e}^0 = \boldsymbol{Y}^0 - \boldsymbol{X}^0\boldsymbol{\beta}$. $\boldsymbol{\beta}$ and σ^2 can be solved from their first-order maximizing conditions: $\boldsymbol{\beta} = \left(\boldsymbol{X}^{0\mathrm{T}}\boldsymbol{X}^0\right)^{-1}\boldsymbol{X}^{0\mathrm{T}}\boldsymbol{Y}^0$ and $\sigma^2 = \left(\boldsymbol{Y}^0 - \boldsymbol{X}^0\boldsymbol{\beta}\right)^{\mathrm{T}}\left(\boldsymbol{Y}^0 - \boldsymbol{X}^0\boldsymbol{\beta}\right)\Big/NT$. Upon substituting $\boldsymbol{\beta}$ and σ^2 in the log-likelihood function, the concentrated log-likelihood function of λ and φ is obtained

$$\begin{aligned}\text{LogL} =& C - \frac{NT}{2}\log\left[\boldsymbol{e}(\lambda,\varphi)^T\boldsymbol{e}(\lambda,\varphi)\right] - \frac{1}{2}\sum_{i=1}^{N}\log\left(1 + T\varphi(1-\lambda\omega_i)^2\right)\\ &+ T\sum_{i=1}^{N}\log(1-\lambda\omega_i)\end{aligned} \tag{3.37}$$

where C is a constant not depending on λ and φ and the typical element of $\boldsymbol{e}(\lambda,\varphi)$ is

$$\begin{aligned}\boldsymbol{e}(\lambda,\theta)_{it} =& y_{it} - \lambda\sum_{j=1}^{N}w_{ij}y_{jt} + \sum_{j=1}^{N}\{[p(\lambda,\phi)_{ij} - (1-\lambda w_{ij})]\frac{1}{T}\sum_{t=1}^{T}y_{jt}\}\\ &- [\mathbf{x}_{it} - \lambda\sum_{j=1}^{N}w_{ij}\mathbf{x}_{jt} + \sum_{j=1}^{N}\{[p(\lambda,\phi)_{ij} - (1-\lambda w_{ij})]\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{jt}\}]\boldsymbol{\beta}\end{aligned} \tag{3.38}$$

The notation $p_{ij} = p(\lambda,\varphi)_{ij}$ is used to indicate that the elements of the matrix $\boldsymbol{P}$ depend on λ and φ. One can iterate between $\boldsymbol{\beta}$ and σ^2 on the one hand, and λ and φ on the other, until convergence. The estimators of $\boldsymbol{\beta}$ and σ^2, given λ and φ, can be obtained by OLS regression of the transformed variable $\boldsymbol{Y}^0$ on the transformed variables $\boldsymbol{X}^0$. However, the estimators of λ and φ, given $\boldsymbol{\beta}$ and σ^2, must be attained by numerical methods because the equations cannot be solved analytically.

The asymptotic variance matrix of this model has been derived by Baltagi et al. (2007). In this paper, they develop diagnostics to test for serial error correlation, spatial error correlation and/or spatial random effects. They also derive asymptotic variance matrices provided that one or more of the corresponding coefficients are zero. One objection to this study is that serial and spatial error correlation are modeled sequentially instead of jointly. Elhorst (2008a) demonstrates that jointly modeling serial and spatial error correlation results in a trade-off between the serial and spatial autocorrelation coefficients and that ignoring this trade-off causes inefficiency and may lead to non-stationarity. However, if the serial autocorrelation coefficient is set to zero, this problem disappears. Consequently, the asymptotic variance matrix that is obtained if the serial autocorrelation coefficient is set to zero exactly happens to be the variance matrix of the random effects spatial error model.

One difference is that Baltagi et al. (2007) do not derive the asymptotic variance matrix of $\boldsymbol{\beta}$, λ, φ and σ^2, but of $\boldsymbol{\beta}$, λ, σ_μ^2 and σ^2. This matrix takes the following form[4]

$$Asy.Var(\boldsymbol{\beta}, \lambda, \sigma_\mu^2, \sigma^2) =$$

$$\begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{\circ T}\boldsymbol{X}^{\circ} & & & \\ 0 & \frac{T-1}{2}tr(\boldsymbol{\Gamma})^2 + \frac{1}{2}tr(\boldsymbol{\Sigma}\boldsymbol{\Gamma})^2 & & \\ 0 & \frac{T}{2\sigma^2}tr\left(\boldsymbol{\Sigma}\boldsymbol{\Gamma}\boldsymbol{V^{-1}}\right) & \frac{T^2}{2\sigma^4}tr\left(\boldsymbol{V}^{-1}\right)^2 & \\ 0 & \frac{T-1}{2\sigma^2}tr(\boldsymbol{\Gamma}) + \frac{1}{2\sigma^2}tr[\boldsymbol{\Sigma}\boldsymbol{\Gamma}\boldsymbol{\Sigma}] & \frac{T}{2\sigma^4}tr\left(\boldsymbol{\Sigma}\boldsymbol{V^{-1}}\right) & \frac{1}{2\sigma^4}\left[(T-1)N + tr(\boldsymbol{\Sigma})^2\right] \end{bmatrix}^{-1} \tag{3.39}$$

where $\boldsymbol{\Gamma} = \left(\boldsymbol{W^T}\boldsymbol{B} + \boldsymbol{B^T}\boldsymbol{W}\right)\left(\boldsymbol{B^T}\boldsymbol{B}\right)^{-1}$ and $\boldsymbol{\Sigma} = \boldsymbol{V}^{-1}\left(\boldsymbol{B^T}\boldsymbol{B}\right)^{-1}$. Since $\varphi = \sigma_\mu^2/\sigma^2$, the asymptotic variance of φ can be obtained using the formula (Mood et al. 1974, p. 181)

$$var(\varphi) = \phi^2\left[\frac{var(\sigma_\mu^2)}{(\varphi\sigma^2)^2} + \frac{var(\sigma^2)}{(\sigma^2)^2} - 2\frac{var(\sigma_\mu^2, \sigma^2)}{(\varphi\sigma^2)\sigma^2}\right] \tag{3.40}$$

In conclusion, we can say that the estimation of the random effects spatial error model is far more complicated than that of the other spatial panel data models.

3.4 Fixed or Random Effects

The spatial econometrics literature is characterized by an overwhelming supply of papers taking the random effects specification as point of departure rather than the fixed effects specification. Baltagi et al. (2003) consider the testing of spatial error correlation in a model with spatial random effects. Baltagi et al. (2007) extend this study to include serial autocorrelation. Kapoor et al. (2007) consider GMM estimation of a spatial error model with time-period random effects. Pfaffermayr (2009) considers maximum likelihood estimation of a random effects SAC model not only for a balanced but also for an unbalanced spatial panel data set. Montes-Rojas (2010) considers the testing of serial error correlation in a random effects spatial lag model. Parent and LeSage (2010, 2011) set out the Bayesian MCMC estimator of a dynamic spatial panel data model. Baltagi and Liu (2011) extend the Kelejian-Prucha (1998) and Lee (2003) instrumental variables estimators of the spatial lag

[4] Note that the matrix $\boldsymbol{Z_0}$ in Baltagi et al. (2007, pp. 39–40) has been replaced by $\boldsymbol{Z}_0 = \left(T\sigma_\mu^2\boldsymbol{I}_N + \sigma^2\left(\boldsymbol{B}'\boldsymbol{B}\right)^{-1}\right)^{-1} = \frac{1}{\sigma^2}\left(T\varphi\boldsymbol{I}_N + (\boldsymbol{B}'\boldsymbol{B})^{-1}\right)^{-1} = \frac{1}{\sigma^2}\boldsymbol{V}^{-1}$.

model to the random effects spatial lag model. Baltagi and Pirotte (2010) focuses on inference based on standard non-spatial panel data estimators if the true model is a random effects model with either a spatial autoregressive or spatial moving average error process. Millo (2013) describes the software implementation of panel data models with random effects, a spatially lagged dependent variable and serially correlated errors if they are estimated by ML. Baltagi et al. (2012) consider the testing of spatial autocorrelation in both the remainder error term and the spatial random effects.

The popularity of the random effects specification can be explained by three reasons. First, it may be considered as a compromise solution to the all or nothing way of utilizing the cross-sectional component of the data. Panel data models with controls for spatial fixed effects only utilize the time-series component of the data, whereas these models without such controls employ both time-series and cross-sectional components. The parameter ϕ in random effects models, which can take values on the interval [0, 1], may be used to estimate the weight that may be attached to the cross-sectional component of the data. If this weight equals 0, the random effects model reduces to the fixed effects model; if it goes to 1, it converges to its counterpart without controls for spatial fixed effects.

Second, the random effects model avoids the loss of degrees of freedom incurred in the fixed effects model associated with a relatively large N. Besides, the spatial fixed effects can only be estimated consistently when T is sufficiently large, because the number of observations available for the estimation of each μ_i is T. Recall, however, that the inconsistency of μ_i is not transmitted to the estimator of the slope coefficients $\boldsymbol{\beta}$, since it is not a function of the estimated μ_i. In other words, the incidental parameters problem does not matter when $\boldsymbol{\beta}$ are the coefficients of interest and the spatial fixed effects μ_i are not, which is the case in most empirical studies.

Third, it avoids the problem that the coefficients of time-invariant variables or variables that only vary a little cannot be estimated. This is the main reason for many studies not to control for spatial fixed effects, for example, because such variables are the main focus of the empirical analysis. It should be realized, however, that if one or more relevant explanatory variables are omitted from the regression equation, when they should be included, the estimator of the coefficients of the remaining variables will be biased and inconsistent (Greene 2008, pp. 133–134). This also holds true for spatial fixed effects and is known as the omitted regressor bias. One can test whether the spatial fixed effects are jointly significant by performing a Likelihood Ratio (LR) test of the hypothesis H_0: $\mu_1 = \ldots = \mu_N = \alpha$, where α is the mean intercept. The corresponding test statistic is $-2\ s$, where s measures the difference between the log-likelihood of the restricted model and that of the unrestricted model. The LR test has a Chi squared distribution with degrees of freedom equal to the number of restrictions that must be imposed on the unrestricted model to obtain the restricted model, which in this particular case is N–1. Thanks to the availability of the log-likelihood of the restricted as well as of the unrestricted model when applying ML estimation methods, the LR test can be carried out instead of, or in addition to, the classical F-test spelled out in Baltagi (2005, p. 13).

Despite its popularity, the question whether the random effects model is also an appropriate specification is often left unanswered. Three conditions should be satisfied before the random effects model may be implemented. First, the number of units should potentially be able to go to infinity. Second, the units of observation should be representative of a larger population. Whether these two conditions are satisfied in spatial research is at least controversial, as discussed below. Finally, the traditional assumption of zero correlation between the random effects μ_i and the explanatory variables needs to be made, which in general is particularly restrictive.

There are two types of asymptotics that are commonly used in the context of spatial observations: (a) The 'infill' asymptotic structure, where the sampling region remains bounded as $N \to \infty$. In this case more units of information come from observations taken from between those already observed; and (b) The 'increasing domain' asymptotic structure, where the sampling region grows as $N \to \infty$. In this case there is a minimum distance separating any two spatial units for all N. According to Lahiri (2003), there are also two types of sampling designs: (a) The stochastic design where the spatial units are randomly drawn; and (b) The fixed design where the spatial units lie on a nonrandom field, possibly irregularly spaced. The spatial econometric literature mainly focuses on increasing domain asymptotics under the fixed sample design (Cressie 1993, p. 100; Griffith and Lagona 1998; Lahiri 2003). Although the number of spatial units under the fixed sample design can potentially go to infinity, this design is incompatible with the increasing domain asymptotic structure. If there is a minimum distance separating spatial units and the researcher wants to collect data for a certain type of spatial units within a particular study area, there will be an upper bound on the number of spatial units. Furthermore, when data on all spatial units within a study area are collected it is questionable whether they are still representative of a larger population. For a given set of regions, such as all counties of a state or all regions in a country, the population may be said 'to be sampled exhaustively' (Nerlove and Balestra 1996, p. 4),[5] and 'the individual spatial units have characteristics that actually set them apart from a larger population' (Anselin 1988, p. 51). In other words, if the data happen to be a random sample of the population, unconditional inference about the population necessitates estimation with random effects. If, however, the objective is limited to making conditional inferences about the sample, then fixed effects should be specified. In this respect Beenstock and Felsenstein (2007) point out that the random effects model should be the default option in principle, since researchers are usually interested in making unconditional inferences about the population and the fixed effects model would lead to an enormous loss of degrees of freedom. However, 'if the sample happens to be the

[5] This remark through Balestra and Nerlove is striking especially since they are the devisers of the random effects model (Balestra and Nerlove 1966).

population' (Beenstock and Felsenstein 2007, p. 178), specific effects should be fixed because each spatial unit represents itself and has not been sampled randomly. Similar observations have been made by Beck (2001, p. 272), 'the critical issue is that the spatial units be fixed and not sampled, and that inference be conditional on the observed units' [see also Hsaio (2003, p. 43) for a more general explanation].

In spatial research there is a prominent reason why investigators generally do not draw a limited sample of units from a particular study area, but rather work with cross-sectional or space-time data of adjacent spatial units located in unbroken study areas. This is because otherwise the spatial weights matrix cannot be defined and the impact of spatial interaction effects cannot be consistently estimated. Only when neighboring units are also part of the sample, it is possible to measure the impact of these neighboring units. In other words, this type of research just requires that the data covers the whole population, since it would break down when having a random sample of the population.

Many studies that derive test statistics for spatial effects or that develop estimation methods for the parameters in spatial panel data models with random effects overlook this issue and therefore can be criticized for not paying (sufficient) attention to the reasoning behind the random effects specification. The motivation to consider random effects rather than fixed effects in one of the first studies that derived a Lagrange Multiplier (LM) test for spatial interaction effects when combining time-series cross-section data consists of all-in-all one single sentence (Baltagi et al. 2003, p. 124): 'Heterogeneity across the cross-sectional units is usually modeled with an error component model'. This observation is representative for other studies dealing with the random effects model. Generally, one needs to read these studies several times to find any motivation in favor of the random effects specification, often without result.

In conclusion, we can say that the fixed effects model is generally more appropriate than the random effects model since spatial econometricians tend to work with space-time data of adjacent spatial units located in unbroken study areas, such as all counties of a state or all regions in a country. To explain cigarette demand using a panel of 46 U.S. states over the period 1963–1992, Yang et al. (2006) adopt a dynamic spatial panel data model with random effects. However, since these states cover almost the whole U.S., a fixed effects model would have been a better choice (Elhorst 2005). We come back to this empirical application in Section 3.6.

To test the assumption of zero correlation between the random effects μ_i and the explanatory variables, the Hausman specification test might be used (Baltagi 2005, pp. 66–68). The hypothesis being tested is H_0: $h = 0$, where

$$h = \boldsymbol{d}^T[\mathrm{var}(\boldsymbol{d})]^{-1}\boldsymbol{d} \tag{3.41}$$

$$\boldsymbol{d} = \hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}$$

$$\mathrm{var}(\boldsymbol{d}) = \hat{\sigma}^2_{RE}\left(\boldsymbol{X}^{\bullet T}\boldsymbol{X}^{\bullet}\right)^{-1} - \hat{\sigma}^2_{FE}\left(\boldsymbol{X}^{*T}\boldsymbol{X}^{*}\right)^{-1}$$

Note the reverse sequence with which $\boldsymbol{d}$ and var($\boldsymbol{d}$) are calculated. This test statistic has a Chi squared distribution with K degrees of freedom (the number of explanatory variables in the model, excluding the constant term). Hausman's specification test can also be used when the model is extended to include spatial error autocorrelation or a spatially lagged dependent variable. Since the spatial lag model has one additional explanatory variable, the test statistic for this model, which should be calculated by $\boldsymbol{d} = \begin{bmatrix} \hat{\boldsymbol{\beta}}^T & \hat{\boldsymbol{\delta}} \end{bmatrix}^T_{FE} - \begin{bmatrix} \hat{\boldsymbol{\beta}}^T & \hat{\delta} \end{bmatrix}^T_{RE}$, has a Chi squared distribution with $K + 1$ degrees of freedom. To calculate var($\boldsymbol{d}$) in this particular case, one should extract the first $K + 1$ rows and columns of the variance matrices in (3.21) and (3.32). Lee and Yu (2012b) formally derive the Hausman test for a general spatial panel data model, which nests various spatial panel data models existing in the literature.[6] If the hypothesis is rejected, the random effects models must be rejected in favor of the fixed effects model.

In addition, one might test the hypothesis H_0: $\phi = 0$ to see whether the random effects should be rejected in favor of the fixed effects model. Recently, Debarsy (2012) has extended the Mundlak approach to the SDM model to help the applied researcher to determine the adequacy of the random effects specification of this spatial econometric model.

3.5 Model Comparison and Selection

To test for spatial interaction effects in a cross-sectional setting, Burridge (1980) and Anselin (1988) developed Lagrange Multiplier (LM) tests for a spatially lagged dependent variable and for spatial error correlation. Anselin et al. (1996) also developed robust LM tests which test for a spatially lagged dependent variable in the local presence of spatial error autocorrelation and for spatial error autocorrelation in the local presence of a spatially lagged dependent variable. These tests have become very popular in empirical research.[7] Recently, Anselin et al. (2006) also specified the classical LM tests for a spatial panel

$$\mathrm{LM}_\delta = \frac{[\mathbf{e}^T(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}/\hat{\sigma}^2]^2}{J} \quad \text{and} \quad \mathrm{LM}_\lambda = \frac{[\boldsymbol{e}^T(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{e}/\hat{\sigma}^2]^2}{T \times T_W} \tag{3.42}$$

---

[6] Mutl and Pfaffermayr (2011) derive the Hausman test when the fixed and random effects models are estimated by 2SLS instead of ML.

[7] Software programs, such as Spacestat and Geoda, have built-in routines that automatically report the results of these tests. Matlab routines have been made available at http://oak.cats.ohiou.edu/~lacombe/research.html by Donald Lacombe and at www.regroningen.nl by Paul Elhorst.

where $\boldsymbol{e}$ denotes the residual vector of a pooled regression model without any spatial or time-specific effects or of a panel data model with spatial and/or time period fixed effects. Finally, J and T_W are defined by

$$J = \frac{1}{\hat{\sigma}^2}\left[\left((\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{I}_{NT} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\right)(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}\hat{\boldsymbol{\beta}} + TT_W\hat{\sigma}^2\right] \tag{3.43}$$

$$T_W = tr\left(\boldsymbol{W}\boldsymbol{W} + \boldsymbol{W}^T\boldsymbol{W}\right) \tag{3.44}$$

Elhorst (2010b) shows that the robust counterparts of these LM tests for a spatial panel take the form

$$\text{robust LM}_\delta = \frac{\left[\boldsymbol{e}^T(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}/\hat{\sigma}^2 - \boldsymbol{e}^T(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{e}/\hat{\sigma}^2\right]^2}{J - TT_W}, \tag{3.45}$$

$$\text{robust LM}_\lambda = \frac{\left[\boldsymbol{e}^T(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{e}/\hat{\sigma}^2 - TT_W/J \times \boldsymbol{e}^T(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}/\hat{\sigma}^2\right]^2}{TT_W[1 - TT_W/J]} \tag{3.46}$$

The classical and robust LM tests are based on the residuals of the non-spatial model with or without spatial and/or time-period fixed effects and follow a Chi square distribution with one degree of freedom. Alternatively, one may use conditional LM tests which test for the existence of one type of spatial dependence conditional on the other. A mathematical derivation of these tests for a spatial panel data model with spatial fixed effects can be found in Debarsy and Ertur (2010). The difference between these robust and conditional LM tests is that the first are based on the residuals of non-spatial models and the second on the ML residuals of the spatial lag or spatial error model.

Applied researchers often find weak evidence in favor of spatial interaction effects when time-period fixed effects are also accounted for. The explanation is that most variables tend to increase and decrease together in different spatial units along the national evolution of these variables over time. Examples are the evolutions of the labor force participation rate and the unemployment rate over the business cycle (Elhorst 2008b; Zeilstra and Elhorst 2012). In the long term, after the effects of shocks have been settled, variables return to their equilibrium values. In equilibrium, neighboring values tend to be more similar than those further apart, but this interaction effect is often weaker than its counterpart over time. The mathematical explanation is that time-period fixed effects are identical to a spatially autocorrelated error term with a spatial weights matrix whose elements are all equal to $1/N$, including the diagonal elements. When this spatial weights matrix would be adopted, one obtains

$$y_{it} - \sum_{j=1}^{N} w_{ij}y_{jt} = y_{it} - \frac{1}{N}\sum_{j=1}^{N} y_{jt} \text{ and } \boldsymbol{x}_{it} - \sum_{j=1}^{N} w_{ij}\boldsymbol{x}_{jt} = \boldsymbol{x}_{it} - \frac{1}{N}\sum_{j=1}^{N}\boldsymbol{x}_{jt} \tag{3.47}$$

which is equivalent to the demeaning procedure of Eq. (3.4) but then for fixed effects in time. Even though spatial weights matrices with non-zero diagonal elements are unusual in spatial econometrics, these expressions show that accounting for time-period fixed effects is one way to correct for spatial interaction effects among the error terms. If, in addition to time-period fixed effects, a spatial error term is considered with a spatial weights matrix with zero diagonal elements, the magnitude of this spatial interaction effect will automatically fall as a result. Using Monte Carlo simulation experiments, Lee and Yu (2010b) show that ignoring time-period fixed effects may lead to large upward biases (up to 0.45) in the coefficient of the spatial lag.

3.5.1 Goodness-of-fit

The computation of a goodness-of-fit measure in spatial panel data models is difficult because there is no precise counterpart of the R^2 of an OLS regression model with disturbance covariance $\sigma^2 \boldsymbol{I}$ to a generalized regression model with disturbance covariance matrix $\sigma^2 \boldsymbol{\Omega} (\boldsymbol{\Omega} \neq \boldsymbol{I})$. Most empirical researchers use

$$R^2(\boldsymbol{e}, \boldsymbol{\Omega}) = 1 - \frac{\boldsymbol{e}^T \boldsymbol{\Omega} \boldsymbol{e}}{(\boldsymbol{Y} - \bar{\boldsymbol{Y}})^T (\boldsymbol{Y} - \bar{\boldsymbol{Y}})} \text{ or } R^2(\tilde{\boldsymbol{e}}) = 1 - \frac{\tilde{\boldsymbol{e}}^T \tilde{\boldsymbol{e}}}{(\boldsymbol{Y} - \bar{\boldsymbol{Y}})^T (\boldsymbol{Y} - \bar{\boldsymbol{Y}})} \tag{3.48}$$

where $\bar{\boldsymbol{Y}}$denotes the overall mean of the dependent variable in the sample and $\boldsymbol{e}$ is the residual vector of the model. Alternatively, $\boldsymbol{e}^T \boldsymbol{\Omega} \boldsymbol{e}$ can be replaced by the residual sum of squares of transformed residuals $\tilde{\boldsymbol{e}}^T \tilde{\boldsymbol{e}}$.

One objection to the measures in (3.48) is that there is no assurance that adding (eliminating) a variable to (from) the model will result in an increase (decrease) of R^2. This problem is at issue in the fixed effects spatial error model, the random effects spatial lag model and the random effects spatial error model, because the coefficients λ, θ or φ may change when changing the set of independent variables. The problem is not at issue in the fixed effects spatial lag model, but note another problem in Eq. (3.50) below. This is because the demeaning procedure was only meant to speed up computation time and to improve the accuracy of the estimates of $\boldsymbol{\beta}$. If the R^2 is calculated after the spatial fixed effects have been added back to the model, it will have the same properties as the R^2 of the OLS model.

An alternative goodness-of-fit measure that meets the above objection is the squared correlation coefficient between actual and fitted values (Verbeek 2000, p. 21).

$$corr^2(\boldsymbol{Y}, \hat{\boldsymbol{Y}}) = \frac{\left[(\boldsymbol{Y} - \bar{\boldsymbol{Y}})^T (\hat{\boldsymbol{Y}} - \bar{\boldsymbol{Y}})\right]^2}{\left[(\boldsymbol{Y} - \bar{\boldsymbol{Y}})^T (\boldsymbol{Y} - \bar{\boldsymbol{Y}})\right]\left[(\hat{\boldsymbol{Y}} - \bar{\boldsymbol{Y}})^T (\hat{\boldsymbol{Y}} - \bar{\boldsymbol{Y}})\right]} \tag{3.49}$$

where $\hat{\boldsymbol{Y}}$ is an $NT \times 1$ vector of fitted values. Unlike the R^2, this goodness-of-fit measure ignores the variation explained by the spatial fixed effects. The argumentation is that the estimator of $\boldsymbol{\beta}$ in the fixed effects model is chosen to explain the time-series rather than the cross-sectional component of the data, as well as that the spatial fixed effects capture rather than explain the variation between the spatial units

(Verbeek 2000, p. 320). This is also the reason why the spatial fixed effects are often not computed, let alone reported. The difference between R^2 and $corr^2$ indicates how much of the variation is explained by the fixed effects, which in many cases is quite substantial. A similar type of argument applies to spatial random effects.

Another difficulty is how to cope with a spatially lagged dependent variable. If the spatial lag is seen as a variable that helps to explain the variation in the dependent variable, the first measure (R^2) should be used. By contrast, if the spatial lag is not seen as variable that helps to explain the variation in the dependent variable, simply because it is a left-hand side variable in principle, the second measure ($corr^2$) should be used. The latter measure is adopted by LeSage (1999) to calculate the goodness-of-fit of the spatial lag model in a cross-sectional setting.[8] In vector notation, the reduced form of the spatial lag model in Eq. (3.3) is

$$\boldsymbol{Y} = [\boldsymbol{I}_{NT} - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W})]^{-1}[\boldsymbol{X\beta} + (\boldsymbol{\tau}_T \otimes \boldsymbol{I}_N)\boldsymbol{\mu} + \boldsymbol{\varepsilon}] \tag{3.50}$$

From this equation it can be seen that the squared correlation coefficient between actual and fitted values in spatial lag models, no matter whether $\boldsymbol{\mu}$ is fixed or random, should also account for the spatial multiplier matrix $[\boldsymbol{I}_{NT} - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W})]^{-1}$.

The two measures for the different spatial panel data models are listed in Table 3.1. It shows that in the fixed and random effects spatial lag model not only the spatially lagged dependent variable, but also the spatial fixed or random effects are ignored when calculating the squared correlation coefficient between actual and fitted values.

Table 3.1 Two goodness-of-fit measures of the four spatial panel data models

Fixed effects spatial lag model	
$R^2(\boldsymbol{e}, \boldsymbol{I}_N)$	$\boldsymbol{e} = \boldsymbol{Y} - \hat{\delta}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} - (\boldsymbol{\tau}_T \otimes \boldsymbol{I}_N)\hat{\boldsymbol{\mu}}$
$Corr^2$	$corr^2\left(\boldsymbol{Y}^*, \left[\boldsymbol{I}_{NT} - \hat{\delta}(\boldsymbol{I}_T \otimes \boldsymbol{W})\right]^{-1}\boldsymbol{X}^*\hat{\boldsymbol{\beta}}\right)$
Fixed effects spatial error model	
$R^2(\tilde{\boldsymbol{e}})$	$\tilde{\boldsymbol{e}} = \boldsymbol{Y} - \hat{\lambda}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y} - \left[\boldsymbol{X} - \hat{\lambda}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}\right]\hat{\boldsymbol{\beta}} - (\boldsymbol{\tau}_T \otimes \boldsymbol{I}_N)\hat{\boldsymbol{\mu}}$
$Corr^2$	$corr^2(\boldsymbol{Y}^*, \boldsymbol{X}^*\hat{\boldsymbol{\beta}})$
Random effects spatial lag model	
$R^2(\tilde{\boldsymbol{e}})$	$\tilde{\boldsymbol{e}} = \boldsymbol{Y}^\bullet - \hat{\delta}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{Y}^\bullet - \boldsymbol{X}^\bullet\boldsymbol{\beta}$
$Corr^2$	$corr^2\left(\boldsymbol{Y}, \left[\boldsymbol{I}_{NT} - \hat{\delta}(\boldsymbol{I}_T \otimes \boldsymbol{W})\right]^{-1}\boldsymbol{X}\hat{\boldsymbol{\beta}}\right)$
Random effects spatial error model	
$R^2(\tilde{\boldsymbol{e}})$	$\tilde{\boldsymbol{e}} = \boldsymbol{Y}^\circ - \boldsymbol{X}^\circ\hat{\boldsymbol{\beta}}$
$Corr^2$	$corr^2(\boldsymbol{Y}, \boldsymbol{X}\hat{\boldsymbol{\beta}})$

$R^2(e, \boldsymbol{I}_N)$ and $R^2(\tilde{\boldsymbol{e}})$ are defined by Eq. (3.48), $corr^2$ is defined by Eq. (3.49)

[8] See the routine “sar” posted at LeSage’s website <www.spatial-econometrics.com>.

3.6 Empirical Illustration

To demonstrate the performance of the different spatial econometric models in an empirical setting, Baltagi and Li's (2004) panel data dataset is used to explain cigarette demand in 46 US states spanning a period of 30 years, from 1963 to 1992. The dependent variable is real per capita sales of cigarettes, which is measured in packs per person aged 14 years and older. The explanatory variables are average retail price of a pack of cigarettes and real per capita disposable income. All variables are taken in logs, as is done in Baltagi and Li (2004). The data is available at www.wiley.co.uk/baltagi/. Details on data sources are given in Baltagi and Levin (1986, 1992) and Baltagi, Griffin, and Xiong (2000). For an adapted version refer to www.regroningen.nl/elhorst. This data set is also used in Elhorst (2005, 2012, 2013), Debarsy et al. (2012), and Kelejian and Piras (2012). The spatial weights matrix $\boldsymbol{W}$ is specified as a row-normalized binary contiguity matrix, with elements $w_{ij} = 1$ if two states share a common border, and zero otherwise. It should be stressed that this specification of the spatial weights matrix is also used in Elhorst (2005, 2012, 2013). Debarsy et al. (2012) specify a row-normalized $\boldsymbol{W}$ based on state border miles in common between the states. Kelejian and Piras (2012) assume interaction effects between states if the price of cigarettes in adjacent states is lower than in the home state.

3.6.1 Software

At www.regroningen.nl the routines sar_panel_FE and sem_panel_FE have been made available, written by Paul Elhorst, to estimate the SAR and SEM models without fixed effects, with spatial fixed effects, with time-period fixed effects, or with both spatial and time-period fixed effects. Furthermore, by replacing the argument $\boldsymbol{X}$ of these routines by [$\boldsymbol{X}$ $\boldsymbol{WX}$] it is also possible to estimate the SDM and SDEM models. In addition, the routines sar_panel_RE and sem_panel_RE can be used to estimate the SAR/SDM and SEM/SDEM models including spatial random effects. The demonstration file "demoLMsarsem_panel" posted at the Web site can be used to reproduce the results of non-spatial models without or with different sets of fixed effects and of the (robust) LM tests to test for spatial dependence reported in Table 3.2 below. The demonstration file "demopanels-compare" can be used to reproduce the results that will be reported in Tables 3.3 and 3.4. This program covers the code of all conceivable models; by changing the specification of $\boldsymbol{Y}$, $\boldsymbol{X}$, $\boldsymbol{W}$, N and T in this routine, by reading a different data set, and by selecting the relevant code depending on the results produced by the different test statistics, the researcher can use these demonstration files to estimate these models for his or her own research problem.

Table 3.2 Estimation results of cigarette demand using panel data models without spatial interaction effects

Determinants	(1)	(2)	(3)	(4)
	Pooled OLS	Spatial fixed effects	Time-period fixed effects	Spatial and time-period fixed effects
Log(P)	−0.859 (−25.16)	−0.702 (−38.88)	−1.205 (−22.66)	−1.035 (−25.63)
Log(Y)	0.268 (10.85)	−0.011 (−0.66)	0.565 (18.66)	0.529 (11.67)
Intercept	3.485 (30.75)			
σ^2	0.034	0.007	0.028	0.005
R^2	0.321	0.853	0.440	0.896
LogL	370.3	1425.2	503.9	1661.7
LM spatial lag	66.47	136.43	44.04	46.90
LM spatial error	153.04	255.72	62.86	54.65
Robust LM spatial lag	58.26	29.51	0.33	1.16
Robust LM spatial error	144.84	148.80	19.15	8.91

Notes: t-values in parentheses

Table 3.3 Estimation results of cigarette demand: spatial Durbin model specification with spatial and time-period specific effects

Determinants	(1)		(2)		(3)	
	Spatial and time-period fixed effects		Spatial and time-period fixed effects bias-corrected		Random spatial effects, Fixed time-period effects	
W*Log(C)	0.219	(6.67)	0.264	(8.25)	0.224	(6.82)
Log(P)	−1.003	(−25.02)	−1.001	(−24.36)	−1.007	(−24.91)
Log(Y)	0.601	(10.51)	0.603	(10.27)	0.593	(10.71)
W*Log(P)	0.045	(0.55)	0.093	(1.13)	0.066	(0.81)
W*Log(Y)	−0.292	(−3.73)	−0.314	(−3.93)	−0.271	(−3.55)
Phi					0.087	(6.81)
σ^2		0.005		0.005		0.005
R^2		0.901		0.902		0.880
Corrected R^2		0.400		0.400		0.317
LogL		1691.4		1691.4		1555.5
Wald test spatial lag	14.83 ($p = 0.006$)		17.96 ($p = 0.001$)		13.90 ($p = 0.001$)	
LR test spatial lag	15.75 ($p = 0.004$)		15.80 ($p = 0.004$)		14.48 ($p = 0.000$)	
Wald test spatial error	8.98 ($p = 0.011$)		8.18 ($p = 0.017$)		7.38 ($p = 0.025$)	
LR test spatial error	8.23 ($p = 0.016$)		8.28 ($p = 0.016$)		7.27 ($p = 0.026$)	

Notes t-values of coefficient estimates and p-values of test results in parentheses, corrected R^2 is R^2 without the contribution of fixed effects

3.6.2 Cigarette Demand

Table 3.2 reports the estimation results when adopting a non-spatial panel data model and test results to determine whether the spatial lag model or the spatial error model is more appropriate. When using the classic LM tests, both the hypothesis of no spatially lagged dependent variable and the hypothesis of no spatially autocorrelated error term must be rejected at 5 % as well as 1 % significance, irrespective of the inclusion of spatial and/or time-period fixed effects. When using the robust tests, the hypothesis of no spatially autocorrelated error term must still be rejected at 5 % as well as 1 % significance. However, the hypothesis of no spatially lagged dependent variable can no longer be rejected at 5 % as well as 1 % significance, provided that time-period or spatial and time-period fixed effects are included.[9] Apparently, the decision to control for spatial and/or time-period fixed effects represents an important issue.

To investigate the (null) hypothesis that the spatial fixed effects are jointly insignificant, one may perform a likelihood ratio (LR) test.[10] The results (2315.7, with 46 degrees of freedom [df], $p < 0.01$) indicate that this hypothesis must be rejected. Similarly, the hypothesis that the time-period fixed effects are jointly insignificant must be rejected (473.1, 30 df, $p < 0.01$). These test results justify the extension of the model with spatial and time-period fixed effects, which is also known as the two-way fixed effects model (Baltagi 2005).

Up to this point, the test results point to the spatial error specification of the two-way fixed effects model. However, if a non-spatial model on the basis of (robust) LM tests is rejected in favor of the spatial lag model or the spatial error model, one should be careful to endorse one of these two models. LeSage and Pace (2009, Chap. 6) recommend to also consider the spatial Durbin model. The results obtained by estimating the parameters of this model can then be used to test the hypotheses H_0: $\boldsymbol{\theta} = \boldsymbol{0}$ and $H_0 : \boldsymbol{\theta} + \delta\boldsymbol{\beta} = \boldsymbol{0}$. The first hypothesis examines whether the spatial Durbin can be simplified to the spatial lag model, and the second hypothesis whether it can be simplified to the spatial error model (Burridge 1981). Both tests follow a Chi squared distribution with K degrees of freedom. If the spatial lag and the spatial error model are estimated too, these tests can take the form of a Likelihood Ratio (LR) test. If these models are not estimated, these tests can only take the form of a Wald test. LR tests have the disadvantage that they require more models to be estimated, while Wald tests are more sensitive to the parameterization of nonlinear constraints (Hayashi 2000, p.122).

---

[9] Note that the test results satisfy the condition that LM spatial lag + robust LM spatial error = LM spatial error + robust LM spatial lag (Anselin et al. 1996).

[10] These tests are based on the log-likelihood function values of the different models. Table 3.2 shows that these values are positive, even though the log-likelihood functions only contain terms with a minus sign. However, since $\sigma^2 < 1$, we have $-\log(\sigma^2) > 0$. Furthermore, since this positive term dominates the negative terms in the log-likelihood function, we eventually have $\text{LogL} > 0$.

If both hypotheses H_0: $\boldsymbol{\theta} = \boldsymbol{0}$ and $H_0 : \boldsymbol{\theta} + \delta\boldsymbol{\beta} = \boldsymbol{0}$ are rejected, then the spatial Durbin best describes the data. Conversely, if the first hypothesis cannot be rejected, then the spatial lag model best describes the data, provided that the (robust) LM tests also pointed to the spatial lag model. Similarly, if the second hypothesis cannot be rejected, then the spatial error model best describes the data, provided that the (robust) LM tests also pointed to the spatial error model. If one of these conditions is not satisfied, i.e. if the (robust) LM tests point to another model than the Wald/LR tests, then the spatial Durbin model should be adopted. This is because this model generalizes both the spatial lag and the spatial error model.

The spatial econometrics literature is divided about whether to apply the specific-to-general approach or the general-to-specific approach (Florax et al. 2003; Mur and Angula 2009). The testing procedure outlined above mixes both approaches. First, the non-spatial model is estimated to test it against the spatial lag and the spatial error model (specific-to-general approach). In case the non-spatial model is rejected, the spatial Durbin model is estimated to test whether it can be simplified to the spatial lag or the spatial error model (general-to-specific approach). If both tests point to either the spatial lag or the spatial error model, it is safe to conclude that that model best describes the data. By contrast, if the non-spatial model is rejected in favor of the spatial lag or the spatial error model while the spatial Durbin model is not, one better adopts this more general model. One weakness of this testing procedure is that the SDEM model is not considered. This is a relatively new issue, currently under investigation by Halleck Vega and Elhorst (2012), discussed before in Section 2.10.

The results that are obtained by estimating the SDM model are reported in Table 3.3. The first column gives the results when this model is estimated using the direct approach and the second column when the coefficients are bias corrected according to Eq. (3.28), after eliminating the fourth row. These results show that the differences between the coefficient estimates of the direct approach and of the bias corrected approach are small for the independent variables ($\boldsymbol{X}$) and σ^2. By contrast, the coefficients of the spatially lagged dependent variable ($\boldsymbol{WY}$) and of the independent variables ($\boldsymbol{WX}$) appear to be quite sensitive to the bias correction procedure. This is the main reason why the bias correction procedure is part of the Matlab routines dealing with the fixed effects spatial lag and the fixed effects spatial error model (the routines "sar_panel_FE" and "sem_panel_FE"). Furthermore, bias correction is the default option in these SAR and SEM panel data estimation routines, but the user can set an input option (info.bc = 0) to turn off bias correction, resulting in uncorrected parameter estimates.

The Wald test (8.98, with 2 degrees of freedom [df], p = 0.011) and the LR test (8.23, 2 df, p = 0.016) indicate that the hypothesis whether the spatial Durbin model can be simplified to the spatial error model, $H_0 : \boldsymbol{\theta} + \delta\boldsymbol{\beta} = \boldsymbol{0}$, must be rejected. Similarly, the hypothesis that the spatial Durbin model can be simplified to the spatial lag model, H_0: $\boldsymbol{\theta} = \boldsymbol{0}$, must be rejected (Wald test: 14.83, 2 df,

$p = 0.006$; LR test: 15.75, 2 df, $p = 0.004$). This implies that both the spatial error model and the spatial lag model must be rejected in favor of the spatial Durbin model.

The third column in Table 3.3 reports the parameter estimates if μ_i is treated as a random variable rather than a set of fixed effects. Hausman's specification test can be used to test the random effects model against the fixed effects model. The results (30.61, 5 df, $p < 0.01$) indicate that the random effects model must be rejected. Another way to test the random effects model against the fixed effects model is to estimate the parameter "phi" (ϕ^2 in Eq. [3.29]), which measures the weight attached to the cross-sectional component of the data and which can take values on the interval [0, 1]. If this parameter equals 0, the random effects model converges to its fixed effectscounterpart; if it goes to 1, it converges to a model without any controls for spatial specific effects. We find phi = 0.087, with t-value of 6.81, which just as Hausman's specification test indicates that the fixed and random effects models are significantly different from each other.

The coefficients of the two explanatory variables in the non-spatial model are significantly different from zero and have the expected signs. In the two-way fixed effects version of this model (the last column of Table 3.2), higher prices restrain people from smoking, while higher income levels have a positive effect on cigarette demand. The price elasticity amounts to −1.035 and the income elasticity to 0.529. However, as the spatial Durbin model specification of this model was found to be more appropriate, we identify these elasticities as biased. To investigate this, it is tempting to compare the coefficient estimates in the non-spatial model with their counterparts in the two-way spatial Durbin model, but this comparison is invalid. Whereas the parameter estimates in the non-spatial model represent the marginal effect of a change in the price or income level on cigarette demand, the coefficients in the spatial Durbin model do not. For this purpose, one should use the direct and indirect effects estimates derived from Eq. (2.13). These effects are reported in Table 3.4. The reason that the direct effects of the explanatory variables are different from their coefficient estimates is due to the feedback effects that arise as a result of impacts passing through neighboring states and back to the states themselves. These feedback effects are partly due to the coefficient of the spatially lagged dependent variable $[\boldsymbol{W}^{*}\mathrm{Log}(\boldsymbol{C})]$, which turns out to be positive and significant, and partly due to the coefficient of the spatially lagged value of the explanatory variable itself. The latter coefficient turns out to be negative and significant for the income variable $[\boldsymbol{W}^{*}\mathrm{Log}(\boldsymbol{Y})]$, and to be positive but insignificant for the price variable $[\boldsymbol{W}^{*}\mathrm{Log}(\boldsymbol{P})]$. The direct and indirect effects estimates and their t-values are computed using the two methods explained in Section 2.7: the first estimate is obtained by computing $(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}$ for every draw, while the second estimate is obtained by storing the traces of the matrices $\boldsymbol{I}$ up to and including $\boldsymbol{W}^{100}$ in advance.

Table 3.4 Direct and indirect effects estimates based on the coefficient estimates of the spatial Durbin model reported in Table 3.3

Determinants	(1)		(2)		(3)	
	Spatial and time-period fixed effects		Spatial and time-period fixed effects bias-corrected		Random spatial effects, Fixed time-period effects	
Direct effect Log(P)	−1.015	−1.014	−1.013	−1.012	−1.018	−1.018
	(−24.34)	(−25.44)	(−24.73)	(−23.93)	(−24.64)	(−25.03)
Indirect effect Log(P)	−0.210	−0.211	−0.220	−0.215	−0.199	−0.195
	(−2.40)	(−2.37)	(−2.26)	(−2.12)	(−2.28)	(−2.19)
Total effect Log(P)	−1.225	−1.225	−1.232	−1.228	−1.217	−1.213
	(−12.56)	(−12.37)	(−11.31)	(−11.26)	(−12.43)	(−12.21)
Direct effect Log(Y)	0.591	0.594	0.594	0.594	0.586	0.583
	(10.62)	(10.44)	(10.45)	(10.67)	(10.68)	(10.53)
Indirect effect Log(Y)	−0.194	−0.194	−0.197	−0.196	−0.169	−0.171
	(−2.29)	(−2.27)	(−2.15)	(−2.18)	(−2.03)	(−2.06)
Total effect Log(Y)	0.397	0.400	0.397	0.398	0.417	0.412
	(5.05)	(5.19)	(4.61)	(4.62)	(5.45)	(5.37)

Left column $(\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1}$ computed every draw, right column $(\boldsymbol{I}_N - \delta \boldsymbol{W})^{-1}$ calculated by Eq. (2.14)

Since the differences are negligible, we focus on the first numbers below. For large values of N, however, it is generally better to turn off the first method and to apply the second method in order to reduce computation time.

In the two-way fixed effects spatial Durbin model (column (2) of Table 3.4) the direct effect of the income variable appears to be 0.594 and of the price variable to be −1.013. This means that the income elasticity of 0.529 in the non-spatial model is underestimated by 10.9 % and the price elasticity of −1.035 by 2.1 %. Since the direct effect of the income variable is 0.594 and its coefficient estimate 0.601, its feedback effect amounts to −0.007 or −1.2 % of the direct effect. Similarly, the feedback effect of the price variable amounts to 0.012 or 1.2 % of the direct effect. In other words, these feedback effects turn out to be relatively small. By contrast, whereas the indirect effects in the non-spatial model are set to zero by construction, the indirect effect of a change in the explanatory variables in the spatial Durbin model appears to be 21.7 % of the direct effect in case of the price variable and −33.2 % in case of the income variable. Furthermore, based on the t-statistics calculated from a set of 1,000 simulated parameter values, these two indirect effects appear to be significantly different from zero. In other words, if the price or the income level in a particular state increases, not only cigarette consumption in that state itself but also in that of its neighboring states will change; the change in neighboring states to the change in the state itself is in the proportion of approximately 1 to −4.6 in case of a price change and 1 to −3.0 in case of an income change.

Up to now, many empirical studies used point estimates of one or more spatial regression model specifications to test the hypothesis as to whether or not spatial spillovers exist. The results above illustrate that this may lead to erroneous conclusions. More specifically, whereas the coefficient of the spatial lagged value of the price variable is positive and insignificant, it's indirect or spillover effect is negative and significant.

The results reported in Table 3.4 illustrate that the t-values of the indirect effects compared to those of the direct effects are relatively small, −24.73 versus −2.26 for the price variable and 10.45 versus −2.15 for the income variable. Experience shows that one needs quite of lot of observations over time to find significant coefficient estimates of the spatially lagged independent variables and, related to that, significant estimates of the indirect effects. It is one of the obstacles to the spatial Durbin model in empirical research. Since most practitioners use cross-sectional data or panel data over a relatively short period of time, they often cannot reject the hypothesis that the coefficients of the spatially lagged independent variables are jointly insignificant (H_0: $\boldsymbol{\theta} = \boldsymbol{0}$), as a result of which they are inclined to accept the spatial lag model. However, one important limitation of the spatial lag model is that the ratio between the direct and indirect effects is the same for every explanatory variable by construction (Elhorst 2010a). In other words, whereas we find that the ratio between the indirect and the direct effects is positive and significant for the price variable (21.7 %) and negative and significant (−33.2 %) for the income variable, these percentages cannot be different from each other when adopting the spatial lag model. In this case, both would amount to approximately 27.1 %. Therefore, practitioners should think twice before abandoning the spatial Durbin model, since not only significance levels count but also flexibility.

The finding that own-state price increases will restrain people not only from buying cigarettes in their own state (elasticity −1.01) but to a limited extent also from buying cigarettes in neighboring states (elasticity −0.22) is not consistent with Baltagi and Levin (1992). They found that price increases in a particular state—due to tax increases meant to reduce cigarette smoking and to limit the exposure of non-smokers to cigarette smoke— encourage consumers in that state to search for cheaper cigarettes in neighboring states. Since Baltagi and Levin (1992) estimate a dynamic but non-spatial panel data model, an interesting topic for further research is whether our spatial spillover effect will change sign when considering a dynamic spatial panel data model. This is investigated in the next chapter.

3.7 Fixed and Random Coefficients Models

In the previous sections spatial heterogeneity was captured by the intercept, but a natural generalization would be to let the slope parameters of the regressors vary as well. Just as the intercept, the slope parameters can also be considered fixed or randomly distributed between spatial units.

If the parameters are fixed but different across spatial units, each spatial unit is treated separately. If $\boldsymbol{Y}_{\mathrm{i}} = \boldsymbol{X}_{\mathrm{i}}\boldsymbol{\beta}_{\mathrm{i}} + \boldsymbol{\varepsilon}_{\mathrm{i}}$ represents the i-th equation in a set of N equations, with the observations stacked by spatial unit over time, the N separate regressions can be related by assuming correlation between the error terms in different equations, a phenomenon that is known as contemporaneous error correlation. Such a specification is reasonable when the error terms for different spatial units, at a given point in time, are likely to reflect some common immeasurable or omitted factor. In full-sample notation, the set of N equations can be written as

$$\begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ . \\ \boldsymbol{Y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & 0 & . & 0 \\ 0 & \boldsymbol{X}_2 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \boldsymbol{X}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ . \\ \boldsymbol{\beta}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ . \\ \boldsymbol{\varepsilon}_N \end{bmatrix} \tag{3.51}$$

where $\mathrm{E}(\boldsymbol{\varepsilon}_i) = 0, \mathrm{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j^{\mathrm{T}}\right) = \sigma_{ij}^2 \boldsymbol{I}_T$ and $i,j = 1,\ldots,N$. This model is also known as the seemingly unrelated regressions (SUR) model.

If the parameters are treated as outcomes of random experiments between spatial units, the data can be pooled into one model in order to estimate the unknown parameters. This is known as the Swamy random coefficients model (Swamy 1970):

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad \mathrm{E}(\boldsymbol{\varepsilon}_i) = \boldsymbol{0}, \quad \mathrm{E}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T) = \sigma_i^2 \boldsymbol{I}_T \tag{3.52a}$$

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\nu}_i, \quad \mathrm{E}(\boldsymbol{\nu}_i) = 0, \quad \mathrm{E}(\boldsymbol{\nu}_i \boldsymbol{\nu}_i^T) = V \tag{3.52b}$$

where the vector $\boldsymbol{\beta}_i$ applying to a particular spatial unit is the outcome of a random process with a common-mean coefficient vector $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{V}$, which is a symmetric $K \times K$ matrix. In addition, it is assumed that $\mathrm{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j^T\right) = 0$ and $\mathrm{E}\left(\boldsymbol{\nu}_i \boldsymbol{\nu}_j^T\right) = 0$ for $i \neq j$ and that the random vectors $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\nu}_i$ are independent of each other.

3.7.1 Fixed Coefficients Spatial Error Model

The fixed coefficients or SUR model given in (3.51), with one equation for every spatial unit over time and with contemporaneous error correlation, does not have to be changed to cope with the spatial error case since the set of $\sigma_{ij}(i,j = 1, \ldots, N)$ already reflects the interactions between the spatial units. In the literature, this is regarded as an advantage because no a priori assumptions are required about the nature of interactions over space (White and Hewings 1982). The explanation is that the specification of a particular spatial weight matrix does not alter the estimates of the response parameters $\boldsymbol{\beta}$, because the estimate of each σ_{ij} would immediately adapt itself to the value of w_{ij} by which it is multiplied. As the SUR model is discussed in almost every econometric textbook, and it is available in almost every commercial econometrics software package, it hardly requires any further explanation.

The standard method to attain the maximum likelihood estimates of the parameters in a SUR model is by iterating the feasible GLS procedure. In every iteration, the residuals of the separate regressions are used to update the elements of the covariance matrix $\sigma_{ij} = \boldsymbol{e}_i^T \boldsymbol{e}_j / T$, until convergence. It should be observed that the estimates of $\boldsymbol{\beta}_i$ and σ_{ij} obtained by iterating the feasible GLS procedure are equivalent to those that would be obtained by maximizing the log-likelihood function of the model, assuming that there are no restrictions on the response parameters $\boldsymbol{\beta}$ across or within the equations.

The efficiency gain in the fixed coefficients spatial error model is greater, the greater the correlation of the error terms, the less correlation exists among variables across equations, and the more correlation exists among variables within an equation (Fiebig 2001). When $\sigma_{ij} = 0$ for $i \neq j$, joint estimation of the set of N equations is not required. Shiba and Tsurumi (1988) provide a complete set of LM and LR tests for this null hypothesis. A hypothesis of particular interest is the homogeneity restriction of equal coefficient vectors $\boldsymbol{\beta}_i$. This hypothesis can be investigated using F or LR tests (Greene 2008).

3.7.2 *Fixed Coefficients Spatial Lag Model*

The set of N equations, with one equation for every spatial unit over time, in a model with fixed coefficients and spatially lagged dependent variables can be expressed as

$$\begin{gathered}
\begin{bmatrix} \boldsymbol{Y}_1 & \boldsymbol{Y}_2 & \ldots & \boldsymbol{Y}_N \end{bmatrix} \begin{bmatrix} 1 & -\delta_{21} & . & -\delta_{N1} \\ -\delta_{12} & 1 & . & -\delta_{N2} \\ . & . & . & . \\ -\delta_{1N} & -\delta_{2N} & . & 1 \end{bmatrix} \\
= \begin{bmatrix} \boldsymbol{X}_1 & 0 & . & 0 \\ 0 & \boldsymbol{X}_2 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \boldsymbol{X}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ . \\ \boldsymbol{\beta}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ . \\ \boldsymbol{\varepsilon}_N \end{bmatrix}
\end{gathered} \tag{3.53}$$

or equivalently

$$\boldsymbol{Y\Gamma} = \boldsymbol{XB} + \boldsymbol{\varepsilon}, \mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}, \mathrm{E}\left(\boldsymbol{\varepsilon\varepsilon}^T\right) = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_N \quad \text{with} \quad \boldsymbol{\Sigma} = \sigma_{ii} \boldsymbol{I}_T \, (i = 1, \ldots, N) \tag{3.54}$$

Each equation can also be written as

$$\boldsymbol{Y}_i = [\boldsymbol{Y}_1 \ldots \boldsymbol{Y}_{i-1} \boldsymbol{Y}_{i+1} \ldots \boldsymbol{Y}_N \ \boldsymbol{X}_i] \times \begin{bmatrix} \delta_{i1} \\ . \\ \delta_{ii-1} \\ \delta_{ii+1} \\ . \\ \delta_{iN} \\ \boldsymbol{\beta}_i \end{bmatrix} + \boldsymbol{\varepsilon}_i \equiv \boldsymbol{Z}_i \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i. \tag{3.55}$$

Note that the δs as well as the $\boldsymbol{\beta}$s are assumed to differ across spatial units. Furthermore, the assumption of contemporaneous error correlation is dropped, and the assumption $\mathrm{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j^T\right) = \sigma_{ij} \boldsymbol{I}_T$ is changed to $\mathrm{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T\right) = \sigma_{ii} \boldsymbol{I}_T$ and $\mathrm{E}\left(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j^T\right) = 0$ for $i \neq j$. Although the latter is not strictly necessary, this change is made to discriminate between the spatial error specification and the spatial lag specification.

The log-likelihood function and the first-order maximizing conditions of a linear simultaneous equations model are given in Hausman (1975, 1983). Due to dropping the assumption of contemporaneous error correlation, the full information maximum likelihood (FIML) estimator of each single $\boldsymbol{\eta}_{\mathrm{i}}$ is

$$\boldsymbol{\eta}_i = \left(\hat{\boldsymbol{Z}}_i^T \boldsymbol{Z}_i\right)^{-1} \hat{\boldsymbol{Z}}_i^T \boldsymbol{Y}_i \tag{3.56a}$$

$$\text{where} \quad \hat{\boldsymbol{Z}}_i = \left[\left(\boldsymbol{XB}\boldsymbol{\Gamma}^{-1}\right)_i \boldsymbol{X}_i\right], \text{ while } \sigma_{ii} = \frac{\left(\boldsymbol{Y}_i - \boldsymbol{Z}_i \boldsymbol{\eta}_i\right)^T \left(\boldsymbol{Y}_i - \boldsymbol{Z}_i \boldsymbol{\eta}_i\right)}{T} \tag{3.56b}$$

The matrix $\boldsymbol{XB}\boldsymbol{\Gamma}^{-1}$ consists of N columns. In the case where $Y_j (j = 1, \ldots, N)$ is an explanatory variable of $Y_i (i = 1, \ldots, N)$, the j-th column of $\boldsymbol{XB}\boldsymbol{\Gamma}^{-1}$ is part of the matrix of estimated values of $\boldsymbol{Z}_i$. The matrix of estimated values of $\boldsymbol{Z}_i$, $\hat{\boldsymbol{Z}}_i$, consists of (N–1 + K) columns: (N–1) columns with respect to the spatially lagged dependent variables explaining $\boldsymbol{Y}_i$, and K columns with respect to the independent variables explaining $\boldsymbol{Y}_i$. Note that the estimated values of $\boldsymbol{Z}_i$ can also be seen as instrumental variables (Hausman 1975, 1983). Since the estimated values of $\boldsymbol{Z}_i$ at the right-hand side of (3.56b) depend on $\boldsymbol{\eta}$, Eqs. (3.56a, b) define no closed form solution for $\boldsymbol{\eta}$. One can attempt to solve for $\boldsymbol{\eta}$ by the Jacobi iteration method. Since a solution $\boldsymbol{\eta} = \mathrm{f}(\boldsymbol{\eta})$ is required, the Jacobi iteration method iterates according to $\boldsymbol{\eta}^{h+1} = \mathrm{f}(\boldsymbol{\eta}^{h})$. This method is available in some commercial econometric software packages.

Because a fixed coefficients spatial lag model has different spatial autoregressive coefficients δ for different spatial units, it follows that the Jacobian term, $T\ln|\boldsymbol{\Gamma}|$, cannot be expressed in function of the characteristic roots of the spatial weight matrix. This difference with the fixed coefficients spatial error model complicates the numerical determination of the FIML estimator. As an alternative, one can use two stage least squares (2SLS), since this estimator has the same asymptotic distribution as the FIML estimator. The benefit of the 2SLS estimator is that it is considerably easier to compute. The incurred cost is a loss in asymptotic efficiency, because 2SLS does not take account of possible restrictions on the coefficients within the matrices $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$.

3.7.3 Additional Remarks

A disadvantage of a model with different parameters for different spatial units is the large number of parameters to be estimated: ($N \times K$) different regression coefficients (β) and (½$N(N+1)$) different (σ) parameters of the (symmetric) covariance matrix in the spatial error model, and ($N \times K$+N(N–1)) different regression coefficients (β, δ) and N different (σ) parameters of the (diagonal) covariance matrix in the spatial lag model. These models are therefore only of use when T is large, and N is small. Another practical problem is that the value of N in most commercial econometrics software is restricted.

Driscoll and Kraay (1998) have pointed out that if N is too large relative to T, it will not be possible to estimate all parameters in a manner that yields a nonsingular estimate. In this case, it is necessary to place prior restrictions on the parameters in order to reduce the dimensionality of the problem. However, even if these restricted estimators are feasible, the quality of the asymptotic approximation used to justify their use is suspect, unless the ratio N/T is close to zero.

One way to reduce the number of parameters of the covariance matrix to 2N in the spatial error model, which at the same time re-establishes the use of the spatial weight matrix, is obtained by imposing the restrictions $\sigma_{ij} = \psi_i w_{ij}$ for $i \neq j$. These restrictions are reasonable if one has prior information about the nature of interactions over space. Under these restrictions, the elements of the covariance matrix must be updated by

$$\sigma_{ii} = \frac{\boldsymbol{e}_i^T \boldsymbol{e}_i}{T}, \quad \psi_i = \sum_{j=1, j\neq i}^{N} w_{ij} \boldsymbol{e}_i^T \boldsymbol{e}_j / T \sum_{j=1, j\neq i}^{N} w_{ij} \tag{3.57}$$

in each iteration. Similarly, the number of regression coefficients in the spatial lag model can be reduced to ($N \times K$+N). Under these circumstances we obtain

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & -\delta_2 w_{21} & . & -\delta_N w_{N1} \\ -\delta_1 w_{12} & 1 & . & -\delta_N w_{N2} \\ . & . & . & . \\ -\delta_1 w_{1N} & -\delta_2 w_{2N} & . & 1 \end{bmatrix} \quad \text{and } \hat{\boldsymbol{Z}}_i = \sum_{j=1}^{N} w_{ij} [\boldsymbol{XB\Gamma}^{-1}]^j \tag{3.58}$$

where $[\boldsymbol{XB\Gamma}^{-1}]^j$ denotes the j-th column of the matrix $\boldsymbol{XB\Gamma}^{-1}$. Although these restrictions simplify the estimation procedure, use of the Jacobi iteration method cannot be avoided.

In both cases, the number of parameters still depends on N, causing the appropriateness of the asymptotic approximation to be suspect. An alternative, more rigorous, way to reduce the number of parameters is to make a compromise between estimating a uniform equation that is valid for all spatial units, and a separate equation for each single spatial unit. First, homogeneous spatial units are joined within groups, and then a separate equation is considered for each group. Schubert (1982) uses this approach in building an interregional labor market model for Austria, and Murphy and Hofler (1984) in estimating a regional unemployment rate equation for the US. Froot (1989) suggests this approach, in more formal terms, in the accounting and finance literature in order to deal with cross-sectional time series data of firms. In addition, one can choose between spatial dependence among the observations within groups (as in Froot 1989), or spatial dependence between groups. The former may be applicable when neighboring spatial units are grouped, and the latter when spatial units with comparable characteristics are put together. Let P denote the number of groups $p(=1, \ldots, P)$, and N_p the number of spatial units in each group, so that $\Sigma_p N_p = N$. Then, the number of parameters for spatial dependence within groups reduces to $P \times K + \Sigma_p {}^1\!/_2 N_p \left(N_\mathrm{p} + 1\right)$ in the spatial error model, and to $P \times K + \Sigma_p N_p \left(N_\mathrm{p} - 1\right)$ in the spatial lag model. In the case of spatial dependence between groups, the number of parameters reduces to $P \times K + {}^1\!/_2 P(P + 1)$ in the spatial error model, and to $P \times K + P(P - 1) + P$ in the spatial lag model.

Another possibility of dealing with spatial error autocorrelation is to employ groups and a nonparametric covariance estimation technique (such as GMM). The GMM technique avoids the estimation of the parameters of the covariance matrix (Driscoll and Kraay 1998). However, these parameter reduction techniques as well as the nonparametric covariance estimation technique (Driscoll and Kraay 1998, fn. 5) rule out applications where the parameters are allowed to vary across all spatial units, which constitutes the initial purpose of the fixed coefficients model.

3.7.4 Random Coefficients Spatial Error Model

The number of parameters to be estimated can also be reduced by treating the coefficients in the regression equation as outcomes of random experiments between spatial units. Consequently, the number of response coefficients no longer grows with the number of spatial units. This approach also improves the efficiency of the estimators due to the availability of substantially more degrees of freedom. Unfortunately, the random effects approach does not reduce the number of parameters of the covariance matrix in the spatial error model, or the number of parameters associated with the spatially lagged dependent variables in the spatial lag model. Therefore, a large value of N relative to T remains a problem.

The random coefficient model with spatial error autocorrelation can be specified as in Eqs. (3.52a, b), incorporating the extension

$$\mathrm{E}(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_j^T) = \sigma_{ij} \boldsymbol{I}_T \tag{3.59}$$

Note that we change the notation slightly by using σ_{ii} for $i = j$ instead of σ_i^2 as in Eq. (3.52a). Similar to the fixed coefficients model, no prior assumptions are required about the nature of the interactions over space. In this model, the random vector $\boldsymbol{Y} \equiv \left(\boldsymbol{Y}_1^T, \ldots, \boldsymbol{Y}_N^T\right)^T$ can be assumed to be distributed with mean $\boldsymbol{X\beta}$, where $\boldsymbol{X} \equiv \left(\boldsymbol{X}_1^T, \ldots, \boldsymbol{X}_N^T\right)^T$, and covariance matrix

$$\begin{aligned}\boldsymbol{\Sigma} &= \begin{bmatrix} \boldsymbol{X}_1\boldsymbol{V}\boldsymbol{X}_1^T + \sigma_{11}\boldsymbol{I}_T & \boldsymbol{\sigma}_{12}\boldsymbol{I}_T & . & \sigma_{1N}\boldsymbol{I}_T \\ \sigma_{21}\boldsymbol{I}_T & \boldsymbol{X}_2\boldsymbol{V}\boldsymbol{X}_2^T + \sigma_{22}\boldsymbol{I}_T & . & \sigma_{2N}\boldsymbol{I}_T \\ . & . & . & . \\ \sigma_{N1}\boldsymbol{I}_T & \sigma_{N2}\boldsymbol{I}_T & . & \boldsymbol{X}_N\boldsymbol{V}\boldsymbol{X}_N^T + \sigma_{NN}\boldsymbol{I}_T \end{bmatrix} \\ &= \boldsymbol{D}(\boldsymbol{I}_N \otimes \boldsymbol{V})\boldsymbol{D}^T + (\boldsymbol{\Sigma}_\sigma \otimes \boldsymbol{I}_T)\end{aligned} \tag{3.60}$$

where $\boldsymbol{D}$ is a $NT \times NK$ block-diagonal matrix, $D = \text{diag}[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N]$, and $\boldsymbol{\Sigma}_\sigma$ is a $N \times N$ matrix with $\boldsymbol{\Sigma}_\sigma = (\sigma_{ij})$. The ML and the GLS estimator of $\boldsymbol{\beta}$ are known to be equivalent (Lindstrom and Bates 1988), and equal to

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y} \tag{3.61}$$

although the major problem is that $\boldsymbol{\Sigma}$ contains unknown parameters $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\Sigma}_{\boldsymbol{\sigma}}, \boldsymbol{V})$ that must also be estimated. There are two ways to proceed. A feasible GLS estimator of $\boldsymbol{\beta}$ can be constructed on the basis of a consistent estimate of $\boldsymbol{\Sigma}_\sigma$ and $\boldsymbol{V}$. To obtain this estimator, the following steps must be carried out. First, estimate the model assuming that all response parameters are fixed and different for differing spatial units. We use the mnemonic *FC* to refer to these estimates. This model is actually the fixed coefficients model without restrictions on the covariance matrix as given in Eq. (3.51). This step results in estimates for $\boldsymbol{\beta}_i^{FC}$ and σ_{ij}^{FC}. Second, estimate $\boldsymbol{V}$ by (see Swamy 1974)

$$\begin{aligned}\boldsymbol{V} = &\frac{1}{N-1}\boldsymbol{S} - \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_{ii}^{FC}\left(\boldsymbol{X}_i^T\boldsymbol{X}_i\right)^{-1} \\ &+ \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}\hat{\sigma}_{ij}^{FC}\left(\boldsymbol{X}_i^T\boldsymbol{X}_i\right)^{-1}\boldsymbol{X}_i^T\boldsymbol{X}_j\left(\boldsymbol{X}_j^T\boldsymbol{X}_j\right)^{-1}\end{aligned} \tag{3.62a}$$

$$\text{where } S = \sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i^{FC} - \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i^{FC}\right)\left(\hat{\boldsymbol{\beta}}_i^{FC} - \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i^{FC}\right)^T \tag{3.62b}$$

The estimator of $\boldsymbol{V}$, although unbiased, may not be positive definite. To ensure the positive definiteness of the estimated matrix, one can also use the consistent estimator $\boldsymbol{V} = 1/(N–1) \times \boldsymbol{S}$ (for details, see Swamy 1970). Finally, estimate the common-mean coefficient vector $\boldsymbol{\beta}$ by GLS according to Eq. (3.61). A distinct problem of the final step is that it requires a matrix inversion of order ($N \times T$). As an alternative, the inverse of $\boldsymbol{\Sigma}$ can be computed with the expression

$$\begin{aligned}\boldsymbol{\Sigma}^{-1} =& (\boldsymbol{\Sigma}_{\sigma}^{-1} \otimes \boldsymbol{I}_T) \\ &- (\boldsymbol{\Sigma}_{\sigma}^{-1} \otimes \boldsymbol{I}_T)\boldsymbol{D}\left[\boldsymbol{D}^T(\boldsymbol{\Sigma}_{\sigma}^{-1} \otimes \boldsymbol{I}_T)\boldsymbol{D} + \boldsymbol{I}_N \otimes \boldsymbol{V}^{-1}\right]^{-1}\boldsymbol{D}^T\left(\boldsymbol{\Sigma}_{\sigma}^{-1} \otimes \boldsymbol{I}_T\right)\end{aligned} \tag{3.63}$$

which requires the inversion of three matrices, one of order K ($\boldsymbol{V}$), one of order N ($\boldsymbol{\Sigma}_{\sigma}$), and one of order $N \times K$ for the matrix between square brackets. In the case where T is large and/or $K \ll T$, this alternative computation is to be preferred, although the inversion of a matrix of order $N \times K$ may still create computational difficulties in some of the commercial econometric software packages.

Despite the mathematical equivalence, the feasible GLS estimator of $\boldsymbol{\beta}$ does not coincide with the ML estimator of $\boldsymbol{\beta}$. This is the case because the feasible GLS estimator of $\boldsymbol{\beta}$ is based on a consistent but not on the ML estimate of $\boldsymbol{\Sigma}_{\sigma}$ and $\boldsymbol{V}$. The statistical literature shows that ML estimation of $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_{\sigma}$ and $\boldsymbol{V}$, although possible, is laborious. There are three reasons for this. First, $\boldsymbol{\Sigma}_{\sigma}$ and $\boldsymbol{V}$ cannot be solved algebraically from the first-order maximizing conditions of the log-likelihood function. Consequently, $\boldsymbol{\Sigma}_{\sigma}$ and $\boldsymbol{V}$ must be solved by numerical methods. Second, a common estimation problem is associated with the restrictions on the parameters of the covariance matrix. A variance estimate should be nonnegative, and a covariance matrix estimate should be nonnegative definite. Moreover, it must be feasible that an estimate takes on values at the boundary of the parameter space. Thus, a variance estimate may be zero, and a covariance matrix estimate may be a nonnegative definite matrix of any rank. In fact, such boundary cases provide useful exploratory information during the model building process. It is desirable that numerical algorithms for ML estimators can successfully produce the defined estimates for all possible samples including those where the maximum is attained at the boundary of the parameter space. However, this parameter space problem often causes difficulties with existing ML algorithms (Shin and Amemiya 1997, p. 190). Third, although some studies assert to have developed efficient and effective algorithms for the likelihood-based estimation of the parameters, they generally assume that $\mathrm{E}(\boldsymbol{\varepsilon}_{\mathrm{i}}\boldsymbol{\varepsilon}_{\mathrm{i}}^{\mathrm{T}}) = \sigma^2\boldsymbol{I}_T$ and $\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j^T) = 0$ for $i,j = 1, \ldots, N$ and $i \neq j$ instead of $\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j^T) = \sigma_{ij}^2\boldsymbol{I}_T$ (Jenrich and Schluchter 1986; Lindstrom and Bates 1988, p. 1014, left column; Longford 1993; Goldstein 1995; Shin and Amemiya 1997, p. 189). This naturally simplifies the parameter space problem, and it is therefore not clear whether these algorithms work for the more general case.

3.7.5 Random Coefficients Spatial Lag Model

A full random parameter model with spatial lags of the dependent variables does in fact not exist. The main reason for this is that the assumption of a random element in the coefficients of lagged dependent variables raises intractable difficulties at the level of identification and estimation (Kelejian 1974; Balestra and Negassi 1992; Hsaio 1996). Instead, a mixed model can be used that contains fixed coefficients for the spatial dependent variables, and random coefficients for the exogenous variables. This model reads as

$$\boldsymbol{Y}_{it} = \delta_{1i}\boldsymbol{Y}_{1t} + \ldots + \delta_{ii-1}\boldsymbol{Y}_{i-1t} + \delta_{i+1t}\boldsymbol{Y}_{i+1t} + \ldots + \delta_{Nt}\boldsymbol{Y}_{Nt} + \boldsymbol{X}_{it}\boldsymbol{\beta}_i + \varepsilon_{it} \quad (3.64)$$

A problem that causes this model not to be used very often is the number of observations needed for its estimation. The minimum number of observations on each spatial unit amounts to ($N + K$), as the number of regressors is (N–1 + K). Most panels do not meet this requirement, even if N is small. Provided that information is available about the nature of interactions over space, we therefore impose the restrictions $\delta_{ij} = \delta_i w_{ij}$ for $j \neq i$, in order to attain

$$\boldsymbol{Y}_{it} = \delta_i \sum_{j=1}^{N} w_{ij}\boldsymbol{Y}_{jt} + \boldsymbol{X}_{it}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_{it} \equiv \delta_i \boldsymbol{Y}_i(w) + X_{it}\boldsymbol{\beta}_i \equiv \boldsymbol{Z}_{it}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_t \quad (3.65)$$

In this case, the minimum number of observations needed on each spatial unit reduces to ($K + 1$), which is independent of N.

Stacking the observations by time for each spatial unit and taking account of Eq. (3.52a, b), the full model can be expressed as

$$\begin{bmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ . \\ \boldsymbol{Y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{Y}_1(w) & 0 & . & 0 \\ 0 & \boldsymbol{Y}_2(w) & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \boldsymbol{Y}_N(w) \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ . \\ \delta_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ . \\ \boldsymbol{X}_N \end{bmatrix} \boldsymbol{\beta}$$

$$+ \begin{bmatrix} \boldsymbol{X}_1 & 0 & . & 0 \\ 0 & \boldsymbol{X}_2 & . & 0 \\ . & . & . & . \\ 0 & 0 & . & \boldsymbol{X}_N \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ . \\ v_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ . \\ \boldsymbol{\varepsilon}_N \end{bmatrix}$$

$$\equiv \text{diag}[\boldsymbol{Y}_1(w), \ldots., \boldsymbol{Y}_N(w)] \times \boldsymbol{\delta} + \boldsymbol{X\beta} + \text{diag}[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N] \times \boldsymbol{v} + \boldsymbol{\varepsilon} \quad (3.66)$$

The covariance matrix of the composite disturbance term $\text{diag}[\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N] \times \boldsymbol{v} + \boldsymbol{\varepsilon}$ is block-diagonal with the i-th diagonal block given by

$$\boldsymbol{\Phi}_i = \boldsymbol{X}_i \boldsymbol{V} \boldsymbol{X}_i^T + \sigma_i^2 \boldsymbol{I}_T \quad (3.67)$$

Similar to the spatial error case, there are two ways to proceed. A feasible GLS estimator of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ may be constructed, extended to instrumental variables and based on a consistent estimate of $\sigma_1^2, \ldots, \sigma_N^2$ and $\boldsymbol{V}$. Alternatively, $\boldsymbol{\delta}, \boldsymbol{\beta}, \sigma_1^2, \ldots, \sigma_N^2$

and $\boldsymbol{V}$ may be estimated by ML.[11] The following feasible GLS analog instrumental variables estimator is suggested by Bowden and Turkington (1984, chap. 3).[12]

Let $\boldsymbol{X}_i$ denote the $(T \times K)$ matrix of the exogenous variables in the i-th equation, $\boldsymbol{Z}_i$ the $(T \times (1 + K))$ matrix of the spatially lagged dependent variable and the exogenous variables in the i-th equation, and $\boldsymbol{X}$ the $(T \times K_{ALL})$ matrix of all the explanatory variables in the full model, where K_{ALL} equals $(N(1 + K))$. Consequently, the inversion of the matrix $\boldsymbol{X}^T\boldsymbol{X}$ of order $(K_{ALL} \times K_{ALL})$ may constitute a problem when N and/or K are large.

First, estimate the model assuming that all coefficients are fixed. We again use the mnemonic FC to refer to these estimates. The model is in effect the fixed coefficients model extended with spatially lagged dependent variables as described above in Eqs. (3.53–3.55), but now we stick to the use of instrumental variable estimators. This results in the following estimates for $\boldsymbol{\eta}_i^{FC}$ and $\sigma_i^{2,FC}$

$$\hat{\boldsymbol{\eta}}_i^{FC} = \left[\boldsymbol{Z}_i^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}_i\right]^{-1}\boldsymbol{Z}_i^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}_i \tag{3.68a}$$

$$\hat{\sigma}_i^{2,FC} = \frac{(\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\eta}}_i^{FC})^T(\boldsymbol{Y}_i - \boldsymbol{Z}_i\hat{\boldsymbol{\eta}}_i^{FC})}{T - K} \tag{3.68b}$$

Second, estimate $\boldsymbol{V}$ by (see Balestra and Negassi 1992; Hsiao and Tahmiscioglu 1997)

$$\hat{\boldsymbol{V}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i^{FC} - \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i^{FC}\right)\left(\hat{\boldsymbol{\beta}}_i^{FC} - \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i^{FC}\right)^T \tag{3.69}$$

Let $\boldsymbol{Z}_i^p$ denote the predictive values from the multi-equation regression of $\boldsymbol{Z}_i = [\boldsymbol{Y}_i(w)\ \boldsymbol{X}_i]$ on $\boldsymbol{X}$, with the observations for each spatial unit weighted by $\boldsymbol{\Phi}_i^{-1}$

$$\boldsymbol{Z}_i^p = \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{\Phi}_i^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Phi}_i^{-1}\boldsymbol{Z}_i = [\boldsymbol{Y}_i^p(w)\,\boldsymbol{X}_i]. \tag{3.70}$$

The inverse of $\boldsymbol{\Phi}_i$ can be computed by the expression

$$\boldsymbol{\Phi}_i^{-1} = \frac{1}{\sigma_i^2}\boldsymbol{I}_T - \frac{1}{\sigma_i^2}\boldsymbol{X}_i\left[\boldsymbol{V}^{-1} + \frac{1}{\sigma_i^2}\boldsymbol{X}_i^T\boldsymbol{X}_i\right]^{-1}\frac{1}{\sigma_i^2}\boldsymbol{X}_i^T \tag{3.71}$$

as a result of which the formula for $\boldsymbol{Z}_i^p$ changes to

[11] One application of this model in the literature is of Sampson et al. (1999), but this paper does not describe the estimation procedure in detail.

[12] Bowden and Turkington start from the regression equation $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\mu}$, where $\mathrm{E}(\boldsymbol{\mu\mu}^{\mathrm{T}}) = \boldsymbol{\Omega}$, and some of the $\boldsymbol{X}$ variables are endogenous. Let $\boldsymbol{Z}$ denote the set of instrumental variables. Then, the GLS analog instrumental variables estimator is $\boldsymbol{b} = \left(\boldsymbol{X}^{pT}\boldsymbol{\Omega}^{-1}\boldsymbol{X}^{P}\right)^{-1}\boldsymbol{X}^{pT}\boldsymbol{\Omega}^{-1}\boldsymbol{Y}$, where $\boldsymbol{X}^p = \boldsymbol{Z}(\boldsymbol{Z}^T\boldsymbol{\Omega}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}^T\boldsymbol{\Omega}^{-1}\boldsymbol{X}$.

$$\begin{aligned}\boldsymbol{Z}_i^p = X \times &\left[\frac{1}{\sigma_{\mathrm{i}}^2}\boldsymbol{X}_i^T\boldsymbol{X} - \frac{1}{\sigma_{\mathrm{i}}^2}\boldsymbol{X}_i^T\boldsymbol{X}_i\left[\boldsymbol{V}^{-1} + \frac{1}{\sigma_{\mathrm{i}}^2}\boldsymbol{X}_i^T\boldsymbol{X}_i\right]^{-1}\frac{1}{\sigma_{\mathrm{i}}^2}\boldsymbol{X}_i^T\boldsymbol{X}_i\right]^{-1} \\ &\times \left[\frac{1}{\sigma_{\mathrm{i}}^2}\boldsymbol{X}_i^T\boldsymbol{Z}_i - \frac{1}{\sigma_{\mathrm{i}}^2}\boldsymbol{X}_i^T\boldsymbol{X}_i\left[\boldsymbol{V}^{-1} + \frac{1}{\sigma_{\mathrm{i}}^2}\boldsymbol{X}_i^T\boldsymbol{X}_i\right]^{-1}\frac{1}{\sigma_{\mathrm{i}}^2}\boldsymbol{X}_i^T\boldsymbol{Z}_i\right]\end{aligned} \quad (3.72)$$

Finally, estimate the parameters δ and $\boldsymbol{\beta}$ by

$$\begin{bmatrix}\delta_1 \\ \cdot \\ \delta_N \\ \boldsymbol{\beta}\end{bmatrix} = \begin{bmatrix}\boldsymbol{Y}_1^p(w)^T\boldsymbol{\phi}_1^{-1}\boldsymbol{Y}_1(w) & \cdot & 0 & \boldsymbol{Y}_1^p(w)^T\boldsymbol{\phi}_1^{-1}\boldsymbol{X}_1 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \boldsymbol{Y}_N^p(w)^T\boldsymbol{\phi}_N^{-1}\boldsymbol{Y}_N(w) & \boldsymbol{Y}_N^p(w)^T\boldsymbol{\phi}_N^{-1}\boldsymbol{X}_N \\ \boldsymbol{X}_1^T\boldsymbol{\phi}_1^{-1}\boldsymbol{Y}_1(w) & \cdot & \boldsymbol{X}_N^T\boldsymbol{\phi}_N^{-1}\boldsymbol{Y}_N(w) & \sum_{i=1}^{N}\boldsymbol{X}_i^T\boldsymbol{\phi}_i^{-1}\boldsymbol{X}_i\end{bmatrix}^{-1} \times \begin{bmatrix}Y_1^p(w)^T\boldsymbol{\phi}_1^{-1}Y_1(w) \\ \cdot \\ Y_N^p(w)^T\boldsymbol{\phi}_N^{-1}Y_N(w) \\ \sum_{i=1}^{N}X_i^T\boldsymbol{\phi}_i^{-1}Y_i(w)\end{bmatrix}, \quad (3.73)$$

where $\boldsymbol{\Phi}_i^{-1}$ can be substituted for the expression given in Eq. (3.71).

3.7.6 Additional Remarks

Although the random coefficients spatial error and spatial lag models have only K response coefficients β, and thus $((N–1) \times K)$ less parameters than their fixed counterparts, the problem that N may be too large relative to T remains. This implies that techniques to reduce the number of σ or δ parameters, as already presented for the fixed coefficients model, also apply to the random coefficients model. To test the homogeneity restriction of equal coefficient vectors $\boldsymbol{\beta}$ a χ^2 test may be used (see Greene 2008). The estimation of the parameters of a random coefficients model is obviously not a simple calculation, but it is feasible. A practical problem is that the fixed coefficients, which must be estimated first, cannot be determined when T is smaller than K. In this case, one has to resort to studies asserting to have developed efficient and effective algorithms for the likelihood-based estimation of the parameters (see above).

Similar to the random effects model not necessarily being an appropriate specification when observations on space-time data of adjacent units in unbroken study areas are used (see Section 3.4), the random coefficients model may not be either. In that case, the fixed coefficients model is compelling, even when N is large.

3.8 Multilevel Models

Many empirical studies use regional data across multiple countries, especially studies that try to explain regional phenomena in different member states of the European Union.The data in these studies may be said to be grouped at two different levels. Regions are so-called level 1 units grouped within countries that are the level 2 units. According to Goldstein (1995, pp. 1–2), the existence of such groupings should not be ignored in the empirical analysis. While most macroeconomic studies focus on how national characteristics affect the dependent variable, one may also view the possibility that the dependent variable observed at the regional level deviates from the national average due to local circumstances. Similarly, while many regional studies based on European data focus on how regional characteristics affect the dependent variable, one may also view this process as embedded in country-level institutional peculiarities, since even among the fairly homogenous group of EU member countries, institutions do differ. Consequently, working at a single level, estimating a macroeconomic equation based on macro data or a regional economic equation based on regional data, is likely to lead to a distorted representation of reality. A single-level model assumes that the data do not follow a hierarchical structure, and thus that all the relevant variation is at one scale. A modeling strategy which does not allow for these national institutions effectively assumes that regions are independent of each other. This is evidently not a safe assumption to make as these national institutions influence the dependent variable at the regional level. For example, whereas it is reasonable to assume that the dependent variable observed in one region is independent of that of a region in another country, the dependent variable of two regions within one country cannot be assumed to be independent within the same country, as both regions share the same institutional framework. Proceeding with a standard regression analysis under the false assumption of independent observations leads to standard errors for the estimates that are too small, giving false impressions of the importance of explanatory variables.

A two-level model takes the hierarchical structure between regions and countries into account by modeling the variation at both levels. In this model a distinction is made between explanatory variables that vary between countries only and explanatory variables that also vary between regions within countries. The former may be denoted as national-level variables and the latter as regional-level variables. The coefficients of the regional-level variables may vary from one country to another and be treated as random, while the coefficients of the national-level variables are the same for all countries and should be treated as fixed. This mixed random and fixed coefficients model reads as

$$Y_{crt} = \boldsymbol{X}_{crt}\boldsymbol{\beta}_c + \boldsymbol{Z}_{ct}\boldsymbol{\alpha} + \varepsilon_{crt} \tag{3.74a}$$

$$\boldsymbol{\beta}_c = \boldsymbol{\beta} + \boldsymbol{v}_c \tag{3.74b}$$

$$E(\varepsilon_{crt}) = 0, Var(\varepsilon_{crt}) = \sigma_c^2 \tag{3.74c}$$

$$E(\boldsymbol{v}_c) = 0, \quad Var(\boldsymbol{v}_c) = \boldsymbol{V} \tag{3.74d}$$

where c (= 1,…, N) refers to a country, r refers to a region (= 1,…, R_c with R_c the number of regions in country c), and t (= 1,…, T) refers to a given time period. Y_{crt} is the dependent variable in region r of country c at time t, X_{crt} is a vector of explanatory variables measured in region r of country c at time t, and $\boldsymbol{Z}_{ct}$ is a vector of explanatory variables in region r but only observed at the national level of country c at time t, since these variables do not differ between regions within countries. ε_{crt} is a heteroskedastic disturbance term with variance σ_c^2, which is assumed to be different for different countries. $\boldsymbol{\beta}$ represents a vector of random response parameters and $\boldsymbol{\alpha}$ a vector of fixed response parameters in the regression equation. The $\boldsymbol{\beta}_c$ applying to a particular country is the outcome of a random process with common-mean-coefficient vector $\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{V}$. When the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_c$ (c = 1,…, N) are of size K, $\boldsymbol{V}$ is of size $K \times K$.

This model belongs to the class of mixed linear models. Frees (2004) gives a general and detailed overview of the mathematical and statistical fundamentals of this class of models, as well as substantive applications across the social sciences. These mixed models are also known as two-level or multilevel models, the difference being that in this type of model as well as in many regional economic applications the error term is commonly assumed to be homoskedastic, $\text{var}(\varepsilon_{\text{crt}}) = \sigma^2$ for every country c.[13] However, it is better to consider a generalization of the two-level model since the assumption of a homoskedastic error term often needs to be rejected in favor of a heteroskedastic error term, $\sigma_c^2 \neq \sigma^2$ for different countries c (see Frees 2004, pp. 45–52).

Two other problems that frequently occur when using space–time data are serial dependence between the observations of each spatial unit over time and spatial dependence between the observations of the spatial units at each point in time. To deal with serial dependence, one might add the dependent variable lagged in time to the model. Similarly, to deal with spatial dependence, one might add the dependent variable lagged in space. Although high or low values of the dependent variable in a particular region often go hand in hand with similar values in surrounding regions, the latter extension may be criticized. This is because the dependent variable often also tends to go up and down in different regions along the national evolution of this variable over time. In general, there are three explanations for these observations. One is business cycle effects, which affect all regions and countries. To control for these effects, one might include time-specific effects. Another is country-specific common spatial interaction effects to all regions. If the government of a particular country changes its national institutional framework, that is, if it changes one of the national factors $\boldsymbol{Z}$ affecting the

[13] Applications based on the multilevel approach in regional economic research are Jones (1991), Ward and Dale (1992), Gould and Fieldhouse (1997), McCall (1998), Elhorst and Zeilstra (2007), Chasco and Lopez (2009), and Zeilstra and Elhorst (2012).

dependent variable in Eq. (3.74a), the dependent variable of all regions located in that country may change. This clustering of regional observations within countries implies spatial dependence. Not accounting for this clustering effect, that is, not controlling for institutional variables observed at the national level may lead to biased results. According to Corrado and Fingleton (2012), these group effects (read: institutional variables) also differ advantageously from the use of regional or country dummy variables that capture the effects of several omitted variables. A final explanation is unobserved national or subnational variables. To account for these unobserved variables, one might incorporate a spatial autoregressive process among the error terms within countries. The best-known spatial dependence model starts with a first-order spatial autoregressive process in the error terms $\boldsymbol{\varepsilon}_{ct} = \lambda_c \boldsymbol{W}_c \boldsymbol{\varepsilon}_{ct} + \boldsymbol{\mu}_{ct}$, where $\boldsymbol{\varepsilon}_{ct}$ and $\boldsymbol{\mu}_{ct}$ are written in vector form for each cross-section of regions in country c at time t, and $\boldsymbol{\mu}_{ct} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{Rc})$. In addition, $\boldsymbol{W}_c (c = 1, \ldots, N)$ is an $R_c \times R_c$ non-negative matrix with zeros on the diagonal to describe the spatial arrangement of the regions in country c, and λ_c is the spatial autocorrelation coefficient, which is assumed to be fixed but different for various countries. Consequently, the covariance matrix of the error terms in Eq. (3.74c) becomes

$$E(\boldsymbol{\varepsilon}_{ct}) = 0, \quad Var(\boldsymbol{\varepsilon}_{ct}) = \sigma_c^2 \left[(\boldsymbol{I}_{R_c} - \lambda_c \boldsymbol{W}_c)^T (\boldsymbol{I}_{R_c} - \lambda_c \boldsymbol{W}_c) \right]^{-1} = \sigma_c^2 \boldsymbol{\Omega_c} \tag{3.75}$$

Summing up, to account for spatial dependence among regions within countries, the model controls for country-specific common spatial interaction effects to all regions, for spatial autocorrelation and for time-specific effects. Especially if the underlying theoretical model does not suggest endogenous interaction effects among the dependent variable observed in different regions, this model represents a strong alternative. Arbia and Fingleton (2008) confirm that the justification of an interaction effect in the dependent variable is a problem for spatial econometrics.

The estimation procedure of this model largely follows the procedures set out in the previous section. A more detailed description can also be found in Elhorst and Zeilstra (2007). A Matlab routine of this estimation procedure is downloadable for free at www.regroningen.nl.

3.9 Spatial SUR Models

Anselin (1988, pp. 137–150) derives the log-likelihood function for a fixed coefficient model that includes spatial error autocorrelation or a spatially lagged dependent variable, but his case considers response coefficients that vary over time rather than across space, as in Sects. 3.7 and 3.8. This model is called spatial SUR. A full spatial SUR model with all types of interaction effects takes the form

$$\boldsymbol{Y}_t = \delta_t \boldsymbol{W} \boldsymbol{Y}_t + \alpha_t \boldsymbol{\iota}_N + \boldsymbol{X}_t \boldsymbol{\beta}_t + \boldsymbol{W} \boldsymbol{X}_t \boldsymbol{\theta}_t + \boldsymbol{u}_t, \quad \boldsymbol{u}_t = \lambda_t \boldsymbol{W} \boldsymbol{u}_t + \boldsymbol{\varepsilon}_t, \; t = 1, .., T \tag{3.76}$$

with $E(\boldsymbol{\varepsilon}_t) = 0$, $E(\boldsymbol{\varepsilon}_{t1}\boldsymbol{\varepsilon}_{t2}^T) = \sigma_{t1,t2}\boldsymbol{I}_T$ for $t, t1, t2 = 1, \ldots, T$. The ML estimation procedure of this model is described in Anselin (ibid) and Mur et al. (2010, Appendix). Alternatively, one may use this setup to estimate different types of dependent variables. Generally, the dependent variable is then not indexed by t running from 1 to T but by m running from 1 to M. Allers and Elhorst (2011) use this setup to estimate a simultaneous model of fiscal policy interactions. The regression coefficients in their model do not vary over time but over the different types of public services. Similarly, Kelejian and Prucha (2004) set out the GMM estimation procedure of a spatial SUR model with different dependent variables. Baltagi and Bresson (2011) describe the ML estimation procedure of a spatial SUR model with different dependent variables when having data in panel rather than in cross-section. This exercise is repeated in Baltagi and Pirotte (2011), but then with model equations further extended to include spatial random effects (though without spatial lag).

Several of these studies also derive LM or robust LM tests for a spatial lag or for a spatial error, jointly or in each single equation, among which Hepple (1997), Mur et al. (2010), and Baltagi and Bresson (2011). Applications of spatial SUR models can further be found in Rey and Montouri (1999), Fingleton (2001, 2007), Egger and Pfaffermayr (2004), Moscone et al. (2007), Wang and Kockelman (2007), LeGallo and Chasco (2008), and Lauridsen et al. (2010).

Below the ML estimation procedure of a spatial SUR model is described if the parameters need to satisfy adding-up restrictions. This procedure is a simplified version of the model presented in Allers and Elhorst (2011). In this study it is shown that a theoretical model of local expenditures on public services and taxation with fiscal policy interactions among local governments that face a budget constraint can be written as a linear expenditure system (LES) by adopting a Stone-Geary social welfare function. This social welfare function reads as

$$V(E_1, \ldots, E_M) = \sum_{m=1}^{M} \beta^m \log(E_m - \alpha^m) \tag{3.77a}$$

$$\text{where} \quad \sum_{m=1}^{M} \beta^m = 1 \tag{3.77b}$$

The welfare V derived from expenditures on a particular public service m, E_m, is a function of the service level in excess of the committed or subsistence level α^m, and of preferences, which determine β^m (the functional forms of α^m and β^m will be introduced below). Note that the Stone-Geary function can only usefully be applied in cases where all public services are normal and all pairs of public services are net substitutes. As long as local public services are categorized into a limited number of broad groups, these conditions are likely to be met. Another limitation is that the Stone-Geary function is only defined if $E_m > \alpha^m$ for all expenditure categories, known as the limited domain problem.

Maximizing this social welfare function subject to the budget constraint

$$\sum_{\mathrm{m}=1}^{\mathrm{M}} E_m = \Pi \tag{3.78}$$

yields the LES which denoted in terms of individual observations takes the form

$$E_{im} = \delta_m \sum_{j=1}^{N} w_{ij} E_{jm} + \alpha^m + \beta^m \left[\Pi_i - \sum_{n=1}^{M} \left(\delta_n \sum_{j=1}^{N} w_{ij} E_{jn} - \alpha^n \right) \right] + \varepsilon_{im} \tag{3.79}$$

where E_{im} represents the expenditures of jurisdiction i on public service m ($m = 1,.., M$). These expenditures are equal to committed expenditure on this service α^m, including expenditures due to policy interaction effects, plus a fraction β^m of discretionary income, which is the income that remains after all committed expenditures have been financed. Note that a different index (n instead of m) is used to compute the sum of these committed expenditures. Further note that tax revenues can also be taken up in this model by treating them as negative expenditures.

The parameter δ_m represents the importance of interaction effects and can be either negative or positive. In the spillover model (Brueckner 2003), δ_m is likely to be negative. Here, $\delta_m \Sigma w_{ij} E_{jm}$ represents the contribution of service levels in other jurisdictions to the locally available service level, as a result of which, subsistence levels provided by the local government can be lower than without spillovers. In the yardstick competition model (Brueckner 2003), δ_m is likely to be positive. Here, $\delta_m \Sigma w_{ij} E_{jm}$ describes the service level that citizens take for granted because it concerns services that inhabitants of other jurisdictions also enjoy.

One advantage of the LES is that it is not necessary to distinguish prices and quantities. Existing empirical work generally studies interactions in expenditure levels not service levels, since a difficult issue is the lack of adequate output measures for public services and the difficulties in deriving unit costs for public services from factor input prices (see Aaberge and Langørgen 2003, for an extensive discussion). Although other demand systems may be more flexible, the LES is one of the few systems where prices (p) and quantities (q) are not separated from each other, as a result of which expenditure data suffices ($E = pq$). This increases the empirical applicability of the LES relative to other systems.

Following Pollak and Wales (1981), the system can be extended by "translating" α^m and β^m. Discretionary income spending can be made dependent on local preferences

$$\beta_i^m = \beta_{m0} + \sum_{g=1}^{G} \beta_{mg} X_{ig} \tag{3.80}$$

where X_{ig} are exogenous variables that determine the share of discretionary income that is spent on public service m in jurisdiction i, and β_{mg} ($g = 0,\ldots, G$) are unknown coefficients to be estimated. The share of discretionary income now depends on variables that are different from one jurisdiction to the other (X_{ig}). For this reason β_i^m should contain the subscript i. Further note that the nature and the number of exogenous variables is the same for each service sector, but that the coefficients are normally different for different public services.

Although regulations and public pressure set minimum service standards across governments, implying that α^m is the same for every jurisdiction, expenditures associated with those uniform minimum requirements α^m may, just as in Jackman and Papadachi (1981) and Aaberge and Langørgen (2003), taken to depend on exogenous variables $S_{imh}(h = 1, \ldots, H_m)$. The symbol S instead of X is used here to distinguish these two different sets of variables. For example, a community with a large share of schoolchildren in its population needs to spend more per capita to attain a certain educational service level than other communities. This yields

$$\alpha_i^m = \alpha_{m0} + \sum_{h=1}^{H_m} \alpha_{mh} S_{imh} \tag{3.81}$$

where $\alpha_{mh}(h = 0, \ldots, H_m)$ are unknown coefficients to be estimated. Note that the nature and the number of the exogenous variables S_{mh}, in contrast to the variables X_g, may be different for different public services. The full model for every single government i involving a system of M equations can then be written as

$$E_{i1} = \sum_{j=1}^{N} \delta_1 w_{ij} E_{j1} + \alpha_{10} + \sum_{h=1}^{H_1} \alpha_{1h} S_{i1h} + (\beta_{10} + \sum_{g=1}^{G} \beta_{1g} X_{ig}) \Phi_i + \varepsilon_{i1} \tag{3.82a}$$

$$. \quad . \quad . \quad . \quad . \quad .$$

$$\begin{aligned} E_{iM} = &\sum_{j=1}^{N} \delta_M w_{ij} E_{jM} + \alpha_{M0} + \sum_{h=1}^{H_M} \alpha_{Mh} S_{iMh} \\ &+ (\beta_{M0} + \sum_{g=1}^{G} \beta_{Mg} X_{ig}) \Phi_i + \varepsilon_{iM} \end{aligned} \tag{3.82b}$$

$$\sum_{m=1}^{M} \beta_{m0} = 1, \ \sum_{m=1}^{M} \beta_{mg} = 0 \, \text{for} \, g = 1, \ldots, G, \tag{3.82c}$$

$$E(\varepsilon_{im}) = 0, \ E(\varepsilon_{im} \varepsilon_{in}^T) = \sigma_{mn} \, \text{for} \, m, n = 1, \ldots, M, \tag{3.82d}$$

where the term $\Phi_i \equiv \Pi_i - \sum_{n=1}^{M} \left\{ \sum_{j=1}^{N} \delta_n w_{ij} E_{jn} + \alpha_{n0} + \sum_{h=1}^{H_n} \alpha_{nh} S_{inh} \right\}$ denotes discretionary income. The adding-up restrictions in Eq. (3.82c) ensure that expenditures of government i sum to Π_i. The error terms in the different equations are assumed to be correlated, which except for SUR is also known as contemporaneous error correlation. Such a specification is reasonable when the error terms for different expenditure categories are likely to reflect some common immeasurable or omitted factors.

The system of equations in (3.82) demonstrates the differences with many previous studies of fiscal policy interaction based on the single equation spatial lag model. They differ in three respects since they set

$$\alpha_{nh} = 0 \,\text{as part of}\, \Phi_i,\ \text{for}\, n = 1,\ldots,M; h = 0,\ldots,H_n \tag{3.83a}$$

$$\delta_n = 0 \,\text{as part of}\, \Phi_i,\ \text{for}\, n = 1,\ldots,M \tag{3.83b}$$

$$\sigma_{\text{mn}} = 0, \text{for}\, m \neq n; m, n = 1,\ldots,M \tag{3.83c}$$

The first restriction shows why the LES, even though it is based on the assumption that all public services are normal and to be substitutes of each other, is more general than a set of different single equation studies. In a single equation model, expenditures on a particular public service are seen as depending on their own cost variables S_{mh} only, whereas in the LES they also appear to depend on the cost variables of other public services. This is because the LES explicitly takes account of the local government's budget constraint. The second restriction shows that the expenditures on a particular public service not only interact with the same expenditures of other governments, but also with the expenditures of those governments on other public services. This result can again be attributed to the local government's budget constraint. The third restriction demonstrates that a set of single equation models imposes zero correlations between the error terms of the different equations of the model, leading to a loss of efficiency: the parameter estimates will be correct, but their confidence intervals will increase.

Kapteyn et al. (1997) already dealt with the problem of interdependent preferences within a linear expenditure system, but they use a simpler model, and were only able to estimate the reduced form parameters. Anselin (1988, pp. 138–145 and pp. 157–162) extensively describes how to estimate a spatial SUR model by ML, but does not deal with the problem that the linear expenditure system is nonlinear in both the α, β and δ parameters and the explanatory variables,[14] that the system of equations has cross-equation restrictions, since the same α and δ parameters enter into all of the equations, and the likelihood function contains a Jacobian term that as a result is far more complicated.

The log-likelihood function of the model in (3.82) is

$$\text{Log}\, L = -\frac{\text{N}}{2}\ln|\boldsymbol{\Omega}| - \frac{1}{2}\boldsymbol{e}^T(\boldsymbol{\Omega}^{-1}\otimes \boldsymbol{I}_N)\boldsymbol{e} + \ln|\boldsymbol{J}| \tag{3.84}$$

Usually, one equation is eliminated to avoid any singularity caused by the adding-up restrictions. Theoretically, the results are invariant no matter which

[14] The linear expenditure system in its basic form is linear in the variables but nonlinear in the parameters. However, Barnum and Squire (1979) have shown that it can be rewritten in such a way that linear estimation techniques can still be used to estimate the parameters. Since the linear expenditure system extended to include interaction effects is also nonlinear in its variables, linear estimation techniques can no longer be used. The same applies to the techniques spelled out in Anselin (1988), which are partially linear.

equation is eliminated. If the Mth equation is eliminated, $\boldsymbol{\Omega}$ is a symmetric $(M-1) \times (M-1)$ matrix, $\boldsymbol{\Omega} = [\sigma_{ij}]$ $(i, j = 1, \ldots, M-1)$, and $\boldsymbol{e}$ is an $N(M-1) \times 1$ vector containing the residuals of the model. These residuals are assumed to be sorted first by equation (i.e., type of public services) and then by spatial unit. $\boldsymbol{J}$ denotes the Jacobian term of the transformation from the vector of error terms $\boldsymbol{\varepsilon}$ to the vector of the dependent variables $\boldsymbol{E}$. This Jacobian term should be calculated over all M equations. Consequently, $\boldsymbol{J}$ is an $MN \times MN$ matrix that takes the following form

$$\boldsymbol{J} = \boldsymbol{I}_{MN} - \begin{bmatrix} \delta_1 \boldsymbol{W} & . & 0 \\ . & . & . \\ 0 & . & \delta_M \boldsymbol{W} \end{bmatrix} + \begin{bmatrix} \delta_1 \boldsymbol{B}_1 \circ \boldsymbol{W} & . & \delta_M \boldsymbol{B}_1 \circ \boldsymbol{W} \\ . & . & . \\ \delta_1 \boldsymbol{B}_{M-1} \circ \boldsymbol{W} & . & \delta_M \boldsymbol{B}_{M-1} \circ \boldsymbol{W} \\ \delta_1 \boldsymbol{B}_{M'} \circ \boldsymbol{W} & . & \delta_M \boldsymbol{B}_{M'} \circ \boldsymbol{W} \end{bmatrix} \tag{3.85}$$

where the symbol $\circ$ denotes the element-by-element product of two vectors or matrices (also known as the Hadamard product). Furthermore,

$$\boldsymbol{B}_m = \begin{bmatrix} \boldsymbol{\beta}_1^m & . & \boldsymbol{\beta}_1^m \\ . & . & . \\ \boldsymbol{\beta}_N^m & . & \boldsymbol{\beta}_N^m \end{bmatrix} \text{ for } m = 1, \ldots, M \tag{3.86}$$

where β_i^m ($i = 1, \ldots, N$; $m = 1, \ldots, M$-1) is defined by Eq. (3.80). Finally, the elements β_i^m of the matrix $\boldsymbol{B}_M$ ($m = M$) should be calculated as

$$\beta_i^M = \beta_{M0} + \sum\nolimits_{g=1}^{G} \beta_{Mg} X_{ig} = 1 - \sum\nolimits_{m=1}^{M-1} \beta_{m0} - \sum\nolimits_{g=1}^{G} \left(\sum\nolimits_{m=0}^{M-1} \beta_{mg} X_{ig} \right) \tag{3.87}$$

to account for the adding-up restrictions in (3.82c). $\boldsymbol{\Omega} = [\sigma_{ij}]$ can be estimated by its first-order maximizing condition

$$\hat{\boldsymbol{\Omega}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{e}_i \boldsymbol{e}_i^T \tag{3.88}$$

where $\boldsymbol{e}_i$ is a $M \times 1$ vector of residuals of the M equations in the system of jurisdiction i. Upon inserting the estimated variance–covariance matrix in the log-likelihood function, the concentrated log-likelihood function of the α, β and δ parameters is obtained

$$\mathrm{Log} L_C = C - \frac{\mathrm{N}}{2} \ln \left(\frac{1}{N} \sum_{i=1}^{N} \boldsymbol{e}_i \boldsymbol{e}_i^T \right) + \ln |\boldsymbol{J}| \tag{3.89}$$

where C is a constant not depending on α, β and δ. Neither of these parameters can be solved analytically from the first-order maximizing conditions. Moreover, the maximum for one of these parameters cannot be found in isolation from the others. This implies that a numerical procedure must be used to find these parameters simultaneously, as well as to approach the asymptotic variance matrix of the parameters by the inverse of the Hessian matrix needed for inference (standard errors, t-values). A Matlab routine is available at www.regroningen.nl.

3.10 Conclusion

This chapter gives a systematic overview of panel data models extended to include spatial error autocorrelation or a spatially lagged dependent variable. In addition, it is shown that these two models can be extended to SDEM and the SDM models by changing the set of explanatory variables $\boldsymbol{X}$ into $\boldsymbol{X} = [\boldsymbol{X}\ \boldsymbol{WX}]$. Each spatial panel data has its own specific problems models, which can be summarized as follows.

Estimation of the *spatial fixed effects model* can be carried out with standard techniques developed by Anselin (1988, pp. 181–182), and Anselin and Hudak (1992), but the regression equation must first be demeaned. This model is relatively simple. One methodological shortcoming is the incidental parameters problem. For short panels, where T is fixed and $N \to \infty$, the coefficients of the spatial fixed effects cannot be estimated consistently. However, this problem does not matter when $\boldsymbol{\beta}$ are the coefficients of interest while the spatial fixed effects are not. Moreover, the problem disappears in panels where N is fixed and $T \to \infty$.

Lee and Yu (2010a) have shown that the parameter estimate of σ^2 in the spatial fixed effects model will be biased in short panels (T is fixed and $N \to \infty$), provided that time-period fixed effects are not included. If time-period fixed effects are also included, the parameter estimates of all parameters will be biased. Bias correction procedures under these circumstances have been formulated for the SAR, SEM, SDM and SDEM models.

Estimation of the *spatial random effects model* can be carried out by maximum likelihood, although it requires a specific approach. The iterative two-stage procedure needed to maximize the log-likelihood function of the random effects spatial lag model appears to be simpler than the procedure for the random effects spatial error model. The parameters of the random effects spatial error and spatial lag model can be consistently estimated when $N \to \infty$, $T \to \infty$, or both N, $T \to \infty$. A major problem is that the random effects model is not an appropriate specification when space-time data of adjacent spatial units located in unbroken study areas are used. In addition, the assumption of zero correlation between the random effects and the explanatory variables is particularly restrictive. Hence, the fixed effects model is compelling, even when N is large, and T is small. A Hausman specification test may be used to test the random effects against the fixed effects model.

A *fixed coefficients spatial error* model with varying coefficients for different spatial units is equivalent to a seemingly unrelated regressions model. Although the estimation of this model is standard, the number of equations allowed in commercial econometric software packages is often limited. A fixed coefficients spatial lag model with different coefficients for different spatial units is almost equivalent to a simultaneous linear equation model. Estimation of this model by maximum likelihood is complicated by the fact that the Jacobian term cannot be expressed in function of the characteristic roots of the spatial weight matrix. As a result, the Jacobi iteration method has to be used, but this method that is available only in a limited number of commercial software packages. As an alternative, one can resort to the use of 2SLS, but this method does not account for restrictions on the coefficients within the coefficient matrices. A formidable problem of fixed coefficients models is the large number of parameters causing the estimators to be infeasible. Furthermore, even if the estimators are made feasible by introducing restrictions on the parameters, the quality of the asymptotic approximation used to justify the approach remains rather suspect, unless the ratio N/T tends to zero. The latter can eventually be achieved by joining spatial units within groups, or by considering separate equations for each group.

Maximum likelihood estimation of the *random coefficients model* extended to spatial error autocorrelation or to spatially lagged dependent variables is possible, although it is laborious. It is simpler to use feasible GLS to estimate the random coefficients model with spatial error autocorrelation, and to use feasible GLS in combination with instrumental variables to estimate the random coefficients model comprising spatially lagged dependent variables. These estimators may still be difficult to compute, because they require matrix inversions of large orders, depending on the number of spatial units and the number of explanatory variables. In the random coefficients model containing spatially lagged dependent variables a random element in the coefficients of the spatially lagged dependent variables should be avoided, because it creates intractabilities with respect to both identification and estimation. Although the number of parameters in the random coefficients spatial error and spatial lag models are smaller than in their fixed coefficients counterparts, N may still be too large relative to T in typical spatial panel datasets. This may cause the estimators to be infeasible or asymptotically suspect. Finally, just as the random effects model, the random coefficients model may not be an appropriate specification when space-time data of adjacent spatial units located in unbroken study areas are used.

Multilevel models recognize that data may be grouped at two different levels, among which regions across multiple countries. To account for spatial dependence among regions within countries, a model has been developed that controls for country-specific common spatial interaction effects to all regions, for spatial autocorrelation and for time-specific effects.

The interest for spatial SUR models is increasing. They are different from fixed and random coefficient models in that the coefficients do not vary over space but over time or over different dependent variables. Generally, the estimation procedure of these models is a straightforward extension of the estimation procedure of single equation spatial econometric models based on cross-sectional data Eqs. (2.5a, b) or on panel data Eq. (3.13) or Eqs. (3.14a, b), but there are exceptions. One such exception is the linear expenditure system of fiscal policy interactions obtained from the Stone-Geary social welfare function. The number of empirical applications of simultaneous equations models compared to single equation studies is still limited.

At his Web site www.regroningen.nl, Elhorst has provided Matlab software to estimate spatial panel data models, among which the spatial lag model, the spatial error model, and the spatial Durbin model extended to include spatial and/or time-period fixed effects or extended to include spatial random effects. These routines are documented in Elhorst (2012) and feature:

1. A generalization of the classic and the robust LM tests to a spatial panel data setting;
2. The bias correction procedure proposed by Lee and Yu (2010a) if the spatial panel data model contain spatial and/or time-period fixed effects;
3. The direct and indirect effects estimates of the explanatory variables proposed by LeSage and Pace (2009);
4. A framework to test the spatial Durbin model against the spatial lag and the spatial error model;
5. A framework to choose among fixed effects, random effects or a model without fixed/random effects.

Similar routines written in R have been developed by Millo and Piras (2012). According to Anselin (2010), spatial econometrics has reached a stage of maturity through general acceptance of spatial econometrics as a mainstream methodology; the number of applied empirical researchers who use econometric techniques in their work also indicates nearly exponential growth. The availability of more and better software, not only for cross-sectional data but also for spatial panels and not only written in Matlab or R but recently also in easier accessible packages such as Stata, might encourage even more researchers to enter this field.

References

Aaberge R, Langørgen A (2003) Fiscal and spending behavior of local governments: identification of price effects when prices are not observed. Public Choice 117:125–61

Allers MA, Elhorst JP (2011) A simultaneous equations model of fiscal policy interactions. J Reg Sci 51:271–291

Anselin L (1988) Spatial econometrics: Methods and models. Kluwer, Dordrecht

Anselin L (2010) Thirty years of spatial econometrics. Papers Reg Sci 89:3–25

Anselin L, Bera A (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles D (eds) Handbook of applied economics statistics. Marcel Dekker, New York, pp 237–289

Anselin L, Hudak S (1992) Spatial econometrics in practice: a review of software options. Reg Sci Urban Econ 22(3):509–536

Anselin L, Bera AK, Florax R, Yoon MJ (1996) Simple diagnostic tests for spatial dependence. Reg Sci Urban Econ 26(1):77–104

Anselin L, Le Gallo J, Jayet H (2006) Spatial panel econometrics. In: Matyas L, Sevestre P (eds) The econometrics of panel data, fundamentals and recent developments in theory and practice, 3rd edn. Kluwer, Dordrecht, pp 901–969

Arbia G, Fingleton B (2008) New spatial econometric techniques and applications in regional science. Papers in Regional Science 87:311–317

Arrelano M (2003) Panel data econometrics. Oxford University Press, Oxford

Balestra P, Negassi S (1992) A random coefficient simultaneous equation system with an application to direct foreign investment by French firms. Empir Econ 17:205–220

Balestra P, Nerlove M (1966) Pooling cross-section and time-series data in the estimation of a dynamic model: the demand for natural gas. Econometrica 34(3):585–612

Baltagi BH (2005) Econometric analysis of panel data, 3rd edn. Wiley, Chichester

Baltagi BH (2006) Random effects and spatial autocorrelation with equal weights. Econom Theory 22(5):973–984

Baltagi BH, Bresson G (2011) Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors. An application to hedonic housing prices in Paris. J Urban Econ 69:24–42

Baltagi BH, Levin D (1986) Estimating dynamic demand for cigarettes using panel data: the effects of bootlegging, taxation and advertising reconsidered. The Review of Economics and Statistics 48:148–155

Baltagi BH, Levin D (1992) Cigarette taxation: raising revenues and reducing consumption. Struct Change Econ Dyn 3(2):321–335

Baltagi BH, Li D (2004) Prediction in the panel data model with spatial autocorrelation. In: Anselin L, Florax RJGM, Rey SJ (eds) Advances in spatial econometrics: Methodology, tools, and applications. Springer, Berlin Heidelberg New York, pp 283–295

Baltagi BH, Liu L (2011) Instrumental variable estimation of a spatial autoregressive panel model with random effects. Econ Lett 111:135–137

Baltagi BH, Pirotte A (2010) Panel data inference under spatial dependence. Econ Model 27:1368–1381

Baltagi BH, Pirotte A (2011) Seemingly unrelated regressions with spatial error components. Empir Econ 40:5–49

Baltagi BH, Griffin JM, Xiong W (2000) To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. Rev Econ Stat 82:117–126

Baltagi BH, Song SH, Koh W (2003) Testing panel data models with spatial error correlation. J Econom 117(1):123–150

Baltagi BH, Song SH, Jung BC, Koh W (2007) Testing for serial correlation, spatial autocorrelation and random effects using panel data. J Econom 140(1):5–51

Baltagi BH, Egger P, Pfaffermayr M (2012) A generalized spatial panel data model with random effects. CESifo Working Paper Series No. 3930. Available at SSRN: http://ssrn.com/abstract=2145816

Barnum HW, Squire L (1979) An econometric application of the theory of the farm-household. J Dev Econ 6:79–102

Beck N (2001) Time-series-cross-section data: What have we learned in the past few years? Ann Rev Polit Sci 4:271–293

Beenstock M, Felsenstein D (2007) Spatial vector autoregressions. Spat Econ Anal 2(2):167–196

Bowden RJ, Turkington DA (1984) Instrumental variables. Cambridge University Press, Cambridge

Breusch TS (1987) Maximum likelihood estimation of random effects models. J Econom 36(3):383–389

Brueckner JK (2003) Strategic interaction among local governments: An overview of empirical studies. International Regional Science Review 26(2):175–188

Burridge P (1980) On the Cliff-Ord test for spatial autocorrelation. J R Stat Soc B 42:107–108

Burridge P (1981) Testing for a common factor in a spatial autoregression model. Environ Plann A 13(7):795–400

Chasco C, López AM (2009) Multilevel models: an application to the beta-convergence model. Région et Développement 30:35–58

Corrado L, Fingleton B (2012) Where is the economics in spatial econometrics? J Reg Sci 52(2):210–239

Cressie NAC (1993) Statistics for spatial data. Wiley, New York

Debarsy N, Ertur C (2010) Testing for spatial autocorrelation in a fixed effects panel data model. Reg Sci Urban Econ 40:453–470

Debarsy N, Ertur C, LeSage JP (2012) Interpreting dynamic space-time panel data models. Stat Methodol 9(1–2):158–171
Driscoll JC, Kraay AC (1998) Consistent covariance matrix estimation with spatially dependent panel data. Rev Econ Stat 80:549–560
Egger P, Pfaffermayr M (2004) Distance, trade and FDI: a Hausman-Taylor SUR approach. J Appl Econom 16:227–246
Elhorst JP (2003) Specification and estimation of spatial panel data models. Int Reg Sci Rev 26(3):244–268
Elhorst JP (2005) Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. Geograph Anal 37(1):62–83
Elhorst JP (2008a) Serial and spatial autocorrelation. Econ Lett 100(3):422–424
Elhorst JP (2008b) A spatiotemporal analysis of aggregate labour force behaviour by sex and age across the European Union. J Geogr Syst 10(2):167–190
Elhorst JP (2010a) Applied spatial econometrics: raising the bar. Spat Econ Anal 5(1):9–28
Elhorst JP (2010b) Spatial panel data models. In: Fischer MM, Getis A (eds) Handbook of applied spatial analysis. Springer, Berlin, pp 377–407
Elhorst JP (2012) Matlab software for spatial panels. Int Reg Sci Rev. doi:10.1177/0160017612452429
Elhorst JP (2013) Spatial panel models. In: Handbook of Regional Science, Ch. 82. Springer, Berlin (Forthcoming)
Elhorst JP, Zeilstra AS (2007) Labour force participation rates at the regional and national levels of the European Union: an integrated analysis. Papers Reg Sci 86:525–549
Frees EW (2004) Longitudinal and panel data. Cambridge, Cambridge University Press
Fiebig DG (2001) Seemingly unrelated regression. In: Baltagi BH (ed) A companion to theoretical econometrics. Blackwell, Malden, pp 101–121
Fingleton B (2001) Theoretical economic geography and spatial econometrics: dynamic perspectives. J Econ Geogr 1:201–225
Fingleton B (2007) Multi-equation spatial econometric model, with application to EU manufacturing productivity growth. J Geogr Syst 9:119–144
Florax RJGM, Folmer H, Rey SJ (2003) Specification searches in spatial econometrics: the relevance of Hendry's methodology. Reg Sci Urban Econ 33(5):557–579
Froot KA (1989) Consistent covariance matrix estimation with cross-sectional dependence and heteroskedasticity in financial data. J Fin Anal 24:333–355
Goldstein H (1995) Multilevel statistical models, 2nd edn. Arnold (Oxford University Press), London
Gould MI, Fieldhouse E (1997) Using the 1991 census SAR in a multilevel analysis of male unemployment. Environ Plan A 29:611–628
Greene WH (2008) Econometric analysis, 6th edn. Pearson, New Jersey
Griffith DA (1988) Advanced spatial statistics. Kluwer, Dordrecht
Griffith DA, Lagona F (1998) On the quality of likelihood-based estimators in spatial autoregressive models when the data dependence structure is misspecified. J Stat Plan Infer 69(1):153–174
Halleck Vega S, Elhorst JP (2012) On spatial econometric models, spillover effects, and W. University of Groningen, Working paper
Hausman JA (1975) An instrumental variables approach to full information estimators for linear and certain nonlinear econometric models. Econometrica 43:727–738
Hausman JA (1983) Specification and estimation of simultaneous equation models. In: Griliches Z, Intriligator MD (eds) Handbook of econometrics, vol 1. Elsevier, Amsterdam, pp 392–448
Hayashi F (2000) Econometrics. Princeton University Press, Princeton
Hepple LW (1997) Testing for spatial autocorrelation in simultaneous equation models. Computational, Environmental and Urban Systems 21(5):307–315
Hsiao C (1996) Random coefficients models. In Mátyás L, Sevestre P (eds) The econometrics of panel data, 2nd revised edn. Kluwer, Dordrecht, pp 77–99
Hsiao C (2003) Analysis of panel data, 2nd edn. Cambridge University Press, Cambridge

Hsiao C, Tahmiscioglu AK (1997) A panel analysis of liquidity constraints and firm investment. J Am Stat Assoc 92:455–465
Hunneman A, Bijmolt T, Elhorst JP (2007) Store location evaluation based on geographical consumer information. In: Paper presented at the marketing science conference, Singapore, 28–30 June 2007
Jackman R, Papadachi J (1981) Local authority education expenditure in England and Wales: why standards differ and the impact of government grants. Public Choice 36:425–439
Jenrich RI, Schluchter MD (1986) Unbalanced repeated-measures models with structured covariance matrices. Biometrics 42:805–820
Jones K (1991) Multi-level models for geographical research. In: Concepts and techniques in modern geography, vol 54. University of East Anglia, Norwich.
Kapoor M, Kelejian HH, Prucha IR (2007) Panel data models with spatially correlated error components. Journal of Econometrics 140(1):97–130
Kapteyn A, Van de Geer S, Van de Stadt H, Wansbeek T (1997) Interdependent preferences: an econometric analysis. J Appl Econom 12:665–686
Kelejian HH (1974) Random parameters in simultaneous equations framework: Identification and estimation. Econometrica 42:517–527
Kelejian HH, Piras G (2012) Estimation of spatial models with endogenous weighting matrices and an application to a demand model for cigarettes. In: Paper presented at the 59th North American meetings of the RSAI, 2012, Ottawa, Canada
Kelejian HH, Prucha IR (1998) A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J Real Estate Fin Econ 17(1):99–121
Kelejian HH, Prucha IR (2002) 2SLS and OLS in a spatial autoregressive model with equal spatial weights. Reg Sci Urban Econ 32(6):691–707
Kelejian HH, Prucha IR (2004) Estimation of simultaneous systems of spatially interrelated cross-sectional equations. J Econom 118:27–50
Kelejian HH, Prucha IR, Yuzefovich Y (2006) Estimation problems in models with spatial weighting matrices which have blocks of equal elements. J Reg Sci 46(3):507–515
Lahiri SN (2003) Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. Sankhya 65:356–388
Lauridsen J, Bech M, López F, Maté M (2010) A spatiotemporal analysis of public pharmaceutical expenditures. Ann Reg Sci 44(2):299–314
Lee LF (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econom Rev 22(4):307–335
Lee LF (2004) Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72(6):1899–1925
Lee LF, Yu J (2010a) Estimation of spatial autoregressive panel data models with fixed effects. J Econom 154(2):165–185
Lee LF, Yu J (2010b) Some recent developments in spatial panel data models. Reg Sci Urban Econ 40:255–271
Lee LF, Yu J (2012a) QML estimation of spatial dynamic panel data models with time varying spatial weights matrices. Spat Econ Anal 7(1):31–74
Lee LF, Yu J (2012b) Spatial panels: Random components versus fixed effects. Int Econ Rev 53:1369–1388
Lee LF, Liu X, Lin X (2010) Specification and estimation of social interaction models with network structures. Econom J 13(2):145–176
LeGallo J, Chasco C (2008) Spatial analysis of urban growth in Spain, 1900-2001. Empir Econ 34:59–80
LeSage JP (1999) Spatial econometrics. www.spatial-econometrics.com/html/sbook.pdf
LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press Taylor & Francis Group, Boca Raton
Lindstrom MJ, Bates DM (1988) Newton-Raphson and EM algorithms for linear mixed-effects model for repeated-measures data. J Am Stat Assoc 83:1014–1022

Longford NT (1993) Random coefficient models. Clarendon Press, Oxford
Magnus JR (1982) Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. J Econom 19(2):239–285
McCall L (1998) Spatial routes to gender wage (in) equality: regional restructuring and wage differentials by gender and education. Econ Geogr 74:379–404
Millo G, Piras G (2012) Splm: spatial panel data models in R. J Stat Softw 47(1):1–38
Millo G (2013) Maximum likelihood estimation of spatially and serially correlated panels with random effects. Comput Stat Data Anal http://dx.doi.org/10.1016/j.bbr.2011.03.031, Forthcoming in print
Montes-Rojas GV (2010) Testing for random effects and serial correlation in spatial autoregressive model. Journal of Statistical Planning and Inference 140:1013–1020
Mood AM, Graybill F, Boes DC (1974) Introduction to the theory of statistics, 3rd edn. McGraw-Hill, Tokyo
Moscone F, Tosetti E, Knapp M (2007) SUR model with spatial effects: an application to mental health expenditure. Health Econ 16:1403–1408
Mur J, Angulo A (2009) Model selection strategies in a spatial setting: Some additional results. Reg Sci Urban Econ 39:200–213
Mur J, López F, Herrera M (2010) Testing for spatial effects in seemingly unrelated regressions. Spat Econ Anal 5(4):399–440
Murphy KJ, Hofler RA (1984) Determinants of geographic unemployment rates: A selectively pooled-simultaneous model. Rev Econ Stat 66:216–223
Mutl J, Pfaffermayr M (2011) The Hausman test in a Cliff and Ord panel model. The Econometrics Journal 14:48–76
Nerlove M, Balestra P (1996) Formulation and estimation of econometric models for panel data. In Mátyás L, Sevestre P (eds) The econometrics of panel data, 2nd revised edn. Kluwer, Dordrecht, pp 3–22
Pace RK, Barry R (1997) Quick computation of spatial autoregressive estimators. Geographical Analysis 29(3):232–246
Parent O, LeSage JP (2010) A spatial dynamic panel model with random effects applied to commuting times. Transp Res Part B 44:633–645
Parent O, LeSage JP (2011) A space-time filter for panel data models containing random effects. Comput Stat Data Anal 55:475–490
Pfaffermayr M (2009) Maximum likelihood estimation of a general unbalanced spatial random effects model: a Monte Carlo study. Spat Econ Anal 4(4):467–483
Pollak RA, Wales TJ (1981) Demographic variables in demand analysis. Econometrica 49:1533–1551
Rey S, Montouri B (1999) US regional income convergence: a spatial econometrics perspective. Reg Stud 33:143–156
Sampson RJ, Morenoff JD, Earles F (1999) Beyond social capital: spatial dynamics of collective efficacy for children. American Sociological Review 64:633–660
Schubert U (1982) REMO – An interregional labor market model of Austria. Environ Plan A 14:1233–1249
Shiba T, Tsurumi H (1988) Bayesian and non-Bayesian tests of independence in seemingly unrelated regressions. Int Econ Rev 29:377–395
Shin C, Amemiya Y (1997) Algorithms for the likelihood-based estimation of the random coefficient model. Stat Probab Lett 32:189–199
Swamy PAVB (1970) Efficient inference in a random coefficient regression model. Econometrica 38:311–323
Swamy PAVB (1974) Linear models with random coefficients. In Zarembka P (ed) Frontiers in econometrics.Academic Press, New York, pp 143-168
Verbeek M (2000) A guide to modern econometrics. Wiley, Chichester
Wang X, Kockelman KM (2007) Specification and estimation of a spatially and temporally autocorrelated seemingly unrelated regression model: application to crash rates in China. Transportation 34:281–300

Wang W, Lee LF (2013) Estimation of spatial panel data models with randomly missing data in the dependent variable. Reg Sci Urban Econ. doi:10.1016/j.regsciurbeco.2013.02.001

Ward C, Dale A (1992) Geographical variation in female labour force participation: An application of multilevel modelling. Reg Stud 26:243–255

White EN, Hewings GJD (1982) Space-time employment modeling: Some results using seemingly unrelated regression estimators. J Reg Sci 22:283–302

Yang Z, Li C, Tse YK (2006) Functional form and spatial dependence in spatial panels. Econ Lett 91(1):138–145

Zeilstra AS, Elhorst JP (2012) An integrated analysis of regional and national unemployment differentials in the European Union. Reg Stud (Forthcoming). http://www.tandfonline.com/loi/cres20

Chapter 4
Dynamic Spatial Panels: Models, Methods and Inferences

Abstract This chapter provides a survey of the existing literature on the specification and estimation of dynamic spatial panel data models, a collection of models for spatial panels extended to include one or more of the following variables and/or error terms: a dependent variable lagged in time, a dependent variable lagged in space, a dependent variable lagged in both space and time, independent variables lagged in time, independent variables lagged in space, serial error autocorrelation, spatial error autocorrelation, spatial-specific and time-period specific effects. The well-known Baltagi and Li (2004) panel dataset, explaining cigarette demand for 46 US states over the period 1963 to 1992, is used to investigate whether the extension of a non-dynamic to a dynamic spatial panel data specification increases the explanatory power of the model.

Keywords Dynamic Effects · Estimation methods · Stationarity conditions · Endogeneity · Non-stability · Spatial spillover effects · Cigarette demand

4.1 Introduction

This chapter provides a survey of the existing literature on the specification and estimation of dynamic spatial panel data models. Ideally, a dynamic model in space and time should be able to deal with (i) serial dependence between the observations on each spatial unit over time, (ii) spatial dependence among the observations at each point in time, (iii) unobservable spatial and/or time-period specific effects, and (iv) endogeneity of one or more of the regressors other than dependent variables lagged in space and/or time. The first problem is the domain of the voluminous time-series econometrics literature (Hamilton 1994; Enders 1995; Hendry 1995), the second problem of the spatial econometrics literature (Anselin 1988; Anselin et al. 2008; LeSage and Pace 2009), and the last two

J. P. Elhorst, *Spatial Econometrics*, SpringerBriefs in Regional Science,
DOI: 10.1007/978-3-642-40340-8_4,

problems of the panel data econometrics literature (Hsiao 2003; Arrelano 2003; Baltagi 2005), to mention just a few well-known textbooks in these fields.

At the turn of this century there was no straightforward estimation procedure for dynamic spatial panel data models. This was because methods developed for dynamic but non-spatial and for spatial but non-dynamic panel data models produced biased estimates when these methods/models were put together. The literature to be reviewed in this chapter includes the main methodological studies that have attempted to solve this shortcoming. The survey also examines the reasoning behind different model specifications and the purposes for which they can be used, which should be useful for practitioners.

4.2 A Generalized Dynamic Model in Space and Time

This section initially focuses on a dynamic model in space and time that generalizes several simpler models that have been considered in the literature. It should be stressed that this generalized model suffers from identification problems and thus is not useful for empirical research. However, when these econometric models are arranged in a framework and their mutual relationships exemplified, it may help to identify which models are the most likely candidates to study space–time data, dependent on the purpose of a particular empirical study.

The most general model when written in vector form for a cross-section of observations at time t reads as

$$\boldsymbol{Y}_t = \tau \boldsymbol{Y}_{t-1} + \delta \boldsymbol{W}\boldsymbol{Y}_t + \eta \boldsymbol{W}\boldsymbol{Y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta}_1 + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\beta}_2 + \boldsymbol{X}_{t-1}\boldsymbol{\beta}_3 + \boldsymbol{W}\boldsymbol{X}_{t-1}\boldsymbol{\beta}_4 + \boldsymbol{Z}_t\boldsymbol{\pi} + \boldsymbol{v}_t \tag{4.1a}$$

$$\boldsymbol{v}_t = \rho \boldsymbol{v}_{t-1} + \lambda \boldsymbol{W}\boldsymbol{v}_t + \boldsymbol{\mu} + \xi_t \boldsymbol{\iota}_N + \boldsymbol{\varepsilon}_t \tag{4.1b}$$

$$\boldsymbol{\mu} = \kappa \boldsymbol{W}\boldsymbol{\mu} + \boldsymbol{\zeta} \tag{4.1c}$$

where $\boldsymbol{Y}_t$ denotes an $N \times 1$ vector consisting of one observation of the dependent variable for every spatial unit ($i = 1, \ldots, N$) in the sample at time t ($t = 1, \ldots, T$), $\boldsymbol{X}_t$ is an $N \times K$ matrix of exogenous explanatory variables, and $\boldsymbol{Z}_t$ is an $N \times L$ matrix of endogenous explanatory variables. A vector or a matrix with subscript $t-1$ denotes its serially lagged value, while a vector or a matrix premultiplied by $\boldsymbol{W}$ denotes its spatially lagged value. The $N \times N$ matrix $\boldsymbol{W}$ is a non-negative matrix of known constants describing the spatial arrangement of the units in the sample. Its diagonal elements are set to zero by assumption, since no spatial unit can be viewed as its own neighbor. The parameters τ, δ and η are the response parameters of successively the dependent variable lagged in time, $\boldsymbol{Y}_{t-1}$, the dependent variable lagged in space, $\boldsymbol{W}\boldsymbol{Y}_t$, and the dependent variable lagged in both space and time, $\boldsymbol{W}\boldsymbol{Y}_{t-1}$. The restrictions that need to be imposed on these parameters and on $\boldsymbol{W}$ to obtain a stationary model are set out in the next section. The $K \times 1$ vectors $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$ and $\boldsymbol{\beta}_4$ represent response parameters of the exogenous explanatory variables, and $\boldsymbol{\pi}$ is an $L \times 1$ vector of response parameters of the endogenous explanatory variables in the model.

The $N \times 1$ vector $\boldsymbol{v}_t$ reflects the error term specification of the model, which is assumed to be serially correlated and to be spatially correlated; ρ is the serial autocorrelation coefficient and λ is the spatial autocorrelation coefficient. In contrast to Eq. (4.1a), an error term lagged in both space and time, $\boldsymbol{W}\boldsymbol{v}_{t-1}$, is not included in Eq. (4.1b), since it is uncommon in the literature. The $N \times 1$ vector $\boldsymbol{\mu} = (\mu,\ldots, \mu_N)^{T}$ contains spatial specific effects, μ_i, and are meant to control for all spatial specific, time-invariant variables whose omission could bias the estimates in a typical cross-sectional study (Baltagi 2005). Similarly, ξ_t (t = 1, …, T) denote time-period specific effects, where $\boldsymbol{\iota}_N$ is an $N \times 1$ vector of ones, meant to control for all time-specific, unit-invariant variables whose omission could bias the estimates in a typical time-series study. These spatial and time-period specific effects may be treated as fixed or as random effects. In addition to this, the spatial specific effects are assumed to be spatially autocorrelated with spatial autocorrelation coefficient κ. Finally, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^{T}$ and $\boldsymbol{\zeta}$ are vectors of i.i.d. disturbance terms, whose elements have zero mean and finite variance σ^2 and σ_{ζ}^2, respectively.

4.3 Stationarity

To achieve stationarity in a dynamic spatial panel data model, restrictions need to be imposed on the parameters of the model and on the spatial weights matrix $\boldsymbol{W}$. The restrictions that need to be imposed on κ and $\boldsymbol{W}$ in a cross-sectional equation like (4.1c) have been shown and extensively discussed in Sect. 2.3.

Elhorst (2008a) demonstrates that the characteristic roots of the matrix $\rho(\boldsymbol{I}_N\text{-}\lambda\boldsymbol{W})^{-1}$ in a space–time equation like (4.1b) should lie within the unit circle. Since the smallest and largest characteristic roots of this matrix take the form $\rho/(1-\lambda\omega_{min})$ and $\rho/(1\text{-}\lambda\omega_{\max})$, or vice versa (dependent on whether ρ is positive or negative), stationarity in time requires the conditions

$$|\rho| < 1 - \lambda\omega_{\max} \quad \text{if} \quad \lambda \geq 0 \tag{4.2c}$$

$$|\rho| < 1 - \lambda\omega_{\min} \quad \text{if} \quad \lambda < 0 \tag{4.2c}$$

These stationarity conditions are graphed in Fig. 4.1a and show that a trade-off exists between the serial and spatial autocorrelation coefficients.

Finally, Elhorst (2001) derives that the characteristic roots of the matrix $(\tau\mathbf{I}_N + \eta\mathbf{W})(I_N\text{-}\delta W)^{-1}$ in a space–time equation like (4.1c) should lie within the unit circle (re-derived in Parent and LeSage 2011), which is the case when

$$\tau > 1 - (\delta + \eta)\omega_{\max} \quad \textit{if} \;\; \delta + \eta \geq 0 \tag{4.3a}$$

$$\tau > 1 - (\delta + \eta)\omega_{\min} \quad \text{if} \;\; \delta + \eta < 0 \tag{4.3b}$$

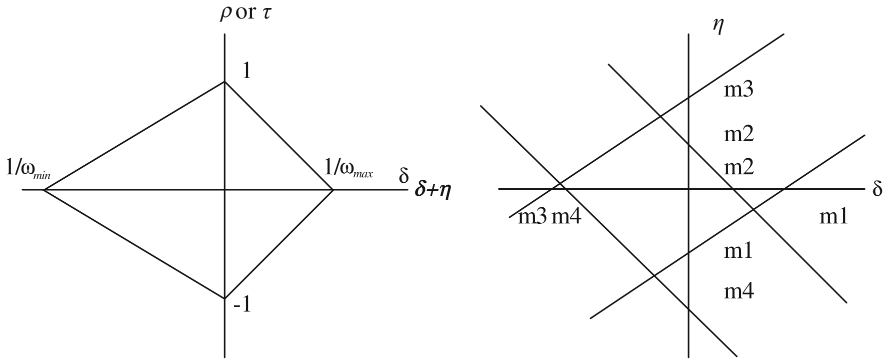

Fig. 4.1 Stationary regions of different model equations. **a** Stationarity region of $\delta + \eta$ and τ in Eq. (4.1a), and of λ and ρ in Eq. (4.1b), **b** Stationarity region of δ and η, given τ, in Eq. (4.1a). m1, m2 m3, and m4 denote points of intersection with the horizontal or vertical axes, where $m1 = \frac{1+\tau}{\omega_{\max}} > 0$, $m2 = \frac{1-\tau}{\omega_{\max}} > 0$, $m3 = \frac{1+\tau}{\omega_{\min}} < 0$, and $m4 = \frac{1-\tau}{\omega_{\min}} < 0$

$$-1 + (\delta - \eta)\omega_{\max} < \tau \quad \text{if } \delta - \eta \geq 0 \tag{4.3c}$$

$$-1 + (\delta - \eta)\omega_{\min} < \tau \quad \text{if } \delta - \eta < 0 \tag{4.3d}$$

If λ is replaced by $\delta + \eta$ and ρ by τ, then the stationarity condition between $\delta + \eta$ on the one hand and τ on the other are similar to those graphed for λ and ρ in Fig. 4.1a. This implies that there exists a trade-off between the serial autoregressive coefficient and the sum of the two spatial autoregressive coefficients. The stationarity conditions between δ and η, given τ, are graphed in Fig. 4.1b. This figure shows that the stationarity region of the two spatial autoregressive coefficients takes the form of a rhombus. The location and the size of this rhombus depend on τ and the smallest and largest characteristic roots of the spatial weights matrix.

The graphs in Fig. 4.1 make clear that the stationarity region implied by the restriction $|\tau| + |\delta| + |\eta| < 1$, put forward in Yu et al. (2008), is too restrictive. For example, whereas the combination of values $(\tau, \delta, \eta) = (0.1,0.9,-0.1)$ based on the restriction $|\tau| + |\delta| + |\eta| < 1$ should be rejected, it is not based on the results presented here. This is because the largest characteristic root of the matrix $(\tau \boldsymbol{I}_N + \eta \boldsymbol{W})(I_N\text{-}\delta \boldsymbol{W})^{-1}$ is smaller than one for this combination of values. By contrast, the stationarity region implied by the restriction $|\tau| + |\delta| + |\eta| < 1$, put forward in Lee and Yu (2010a), is not restrictive enough. For example, whereas the combination of values $(\tau, \delta, \eta) = (1.1,-0.2,0.0)$ is permitted by the restriction $|\tau| + |\delta| + |\eta| < 1$, it should be rejected based on the results presented here. This is because the largest characteristic root of the matrix $(\tau \boldsymbol{I}_N + \eta \boldsymbol{W})(I_N\text{-}\delta \boldsymbol{W})^{-1}$ is greater than one for this combination of values, indicating that the model would explode under these circumstances.

If a model appears to be unstable, that is, if the parameter estimates do not satisfy one of the stationarity conditions, Lee and Yu (2010a) propose to take every variable in Eq. (4.1a) in deviation of its spatially lagged value. Mathematically, this is equivalent with multiplying Eq. (4.1a) by the matrix $(\boldsymbol{I}_N\text{-}\boldsymbol{W})$. The largest characteristic root $\omega_{\max}$ in the stationarity conditions (4.3a) and (4.3c) may then be replaced by $\omega_{\max-1}$, the second largest characteristic root of the spatial weights matrix $\boldsymbol{W}$. Since these newly obtained restrictions are less restrictive than the original ones, the spatial first-differenced model might be stable as a result.

In conclusion, we can say that the stationarity conditions on the spatial and temporal parameters shown in (4.2) and (4.3) go beyond the standard condition $|\tau| < 1$ in serial models and the standard condition $1/\omega_{\min} < \delta < 1/\omega_{\max}$ in spatial models, and that they are considerably more difficult to work with.

The stationarity conditions that need to be imposed on the $N \times N$ spatial weights matrix $\boldsymbol{W}$ in a panel data setting are set forth in Yu et al. (2008). The matrix $I_N\text{-}p\mathbf{W}$ for $p = \delta, \lambda$ should be nonsingular, and the row and column sums of the matrices $\boldsymbol{W}$ and $(I_N\text{-}p\mathbf{W})^{-1}$ should be uniformly bounded in absolute value as N goes to infinity. In addition,

$$\sum_{h=1}^{\infty} abs\left\{\left[(\boldsymbol{I}_N - \delta\boldsymbol{W})^{-1}(\tau\boldsymbol{I}_N + \eta\boldsymbol{W})\right]^h\right\} \tag{4.4}$$

should be uniformly bounded. Let ω_i (i = 1, …, N) denote the characteristic roots of $\boldsymbol{W}$ and $\boldsymbol{R}_N$ the corresponding $N \times N$ matrix of normalized characteristic vectors, then this formula may be rewritten as $\Sigma_h \text{abs}\left\{\left[\boldsymbol{R}_N\boldsymbol{D}_N(\boldsymbol{R}_N)^{-1}\right]^h\right\}$, where $\boldsymbol{D}_N$ is a diagonal matrix whose diagonal elements are $(\tau + \eta\omega_i)/(1-\delta\omega_i)$ (i = 1, …, N). This expression represents the stationarity region of the parameters τ, η, and δ shown in Fig. 4.1. Just as in the previous chapter, the assumption that the row and column sums of $\boldsymbol{W}$ before row-normalization should not diverge to infinity at a rate equal to or faster than the rate of the sample size N, which is made in the cross-sectional setting, is not explicitly made in a panel data setting, unless time-period fixed effects are also considered.

4.4 Feasible Models

Figure 4.2 presents two regressions equations that have extensively been discussed in the econometric literature: the dynamic model without spatial interaction effects and the spatial model without dynamic effects. After a brief discussion of these two types of models, we will consider a taxonomy of dynamic models in both space and time.

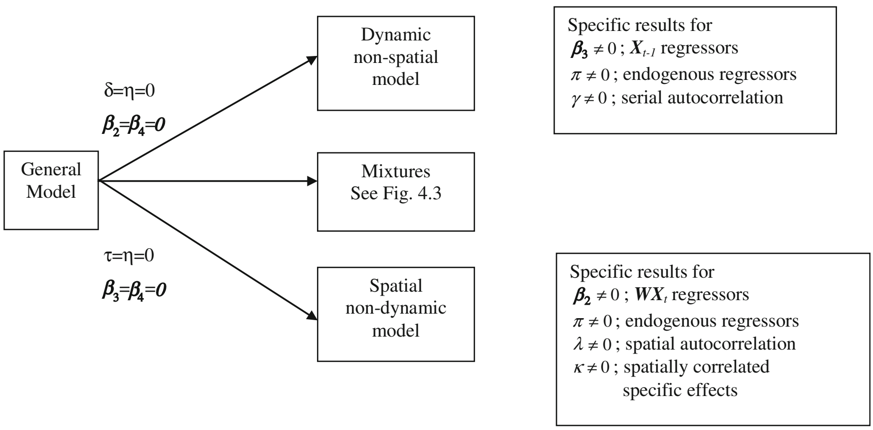

Fig. 4.2 Two feasible models that ignore either spatial or dynamic effects

4.4.1 Dynamic but Non-Spatial Panel Data Models

A panel data model without spatial interaction effects and without dynamic effects (i.e., a static panel data model) can be estimated by the least-squares dummy variables (LSDV) estimator if the spatial specific effects μ are treated as fixed effects, and by the generalized least-squares (GLS) estimator if the spatial specific effects μ are treated as random effects (Hsiao 2003, Chap. 3; Baltagi 2005,Chap. 2). The most serious estimation problem caused by the extension of this model with a dependent variable lagged in time, $\boldsymbol{Y}_{t-1}$, is that these two estimators become inconsistent *if T is fixed, regardless of the size of N* (Hsiao 2003, Chap. 4; Arrelano 2003; Baltagi 2005, Chap. 4). This is because the right-hand variable $\boldsymbol{Y}_{t-1}$ is correlated with the spatial specific effect μ, and uncorrelatedness of regressors and disturbances is one of the basic conditions that needs to be satisfied in regression analysis. Three procedures to remove this inconsistency if T is fixed have been developed.

The first and most popular procedure is generalized method-of-moments (GMM). By defining and solving a set of moment conditions that need to be satisfied at the true values of the parameters to be estimated, one obtains a set of exogenous variables correlated with $\boldsymbol{Y}_{t-1}$ but orthogonal to the errors, which as a results can be used to instrument $\boldsymbol{Y}_{t-1}$. The Arrelano and Bond (1991) *difference* GMM estimator is based on moment conditions after taking first differences to eliminate the spatial specific effects. Typically, this GMM estimator instruments $\Delta\boldsymbol{Y}_{t-1}$ by the variables $\boldsymbol{Y}_1$ up to $\boldsymbol{Y}_{t-2}$ and $\boldsymbol{X}_1$ up to $\boldsymbol{X}_{t-1}$ ($t \geq 3$). In practice, the difference GMM estimator has been shown to perform poorly on data with persistent series. The explanation is that, under these circumstances, the lagged levels of variables tend to have only weak correlation with the first-differenced lagged dependent variable. The Blundell and Bond (1998) *system* GMM estimator has been shown to offer much increased efficiency and less finite sample bias compared to the difference GMM estimator, since this estimator also utilizes lagged first differences for the equation in levels. Typically, this GMM estimator also instruments $\boldsymbol{Y}_{t-1}$ by the variables $\Delta \boldsymbol{Y}_1$ up to $\Delta \boldsymbol{Y}_{t-2}$ and $\Delta \boldsymbol{X}_2$ up to $\Delta \boldsymbol{X}_{t-1}$ ($t \geq 3$). See Baltagi (2005) for a summary of the main results and the main references, and Kukenova and Monteiro (2009) for a more detailed mathematical exposition.

The second procedure applies maximum likelihood (ML) based on the *unconditional* likelihood function of the model. Regression equations that include variables lagged one period in time are often estimated conditional upon the first observations. Nerlove (1999, p. 139), however, points out that conditioning on those initial values is an undesirable feature, especially when the time dimension of the panel is short. If the process generating the data in the sample period is stationary, the initial values convey a great deal of information about this process since they reflect how it has operated in the past. By taking account of the density function of the first observation of each time-series of observations, the unconditional likelihood function is obtained. This procedure has been applied successfully to random effects dynamic panel data models formulated in levels (Bhargava and Sargan 1983). Unfortunately, the unconditional likelihood function does not exist when applying this procedure to the fixed effects model, even without exogenous explanatory variables. The reason is that the coefficients of the fixed effects cannot be estimated consistently, since the number of these coefficients increases as N increases. The standard solution to eliminate these fixed effects from the regression equation by demeaning the Y and X variables also does not work, because this technique creates a correlation of order ($1/T$) between the serial lagged dependent variable and the demeaned error terms, known as the Nickell (1981) bias, as a result of which the common parameter τ cannot be estimated consistently. Only when T tends to infinity, does this inconsistency disappear. More recently, Hsiao et al. (2002) have suggested an alternative procedure for the fixed effects dynamic panel data model. This procedure first-differences the model to eliminate the spatial fixed effects and then considers the unconditional likelihood function of the first-differenced model taking into account the density function of the first first-differenced observations on each cross-sectional unit. They find that this likelihood function is well defined, depends on a fixed number of parameters and satisfies the usual regularity conditions. Thereupon, they conclude that the ML estimator is consistent and asymptotically normally distributed when N tends to infinity, regardless of the size of T. They also find that the ML estimator is asymptotically more efficient than the GMM estimator.

The third procedure is to bias-correct the LSDV estimator. Kiviet (1995), Hahn and Kuersteiner (2002) and Bun and Carree (2005) develop bias correction procedures when both the number of cross-sectional units (N) and the number of time points (T) in the sample go to infinity such that the limit of the ratio of N and T exists and is bounded between zero and infinity ($0 < \lim(N/T) < \infty$). The only problem is that in most empirical studies based on space–time data the most relevant asymptotics are believed to be N tends to infinity and T is fixed. When T is fixed, the spatial-specific effects must be eliminated by first-differencing, whereas first-differencing is not necessary when T tends to infinity.

Specific problems occur when the dynamic panel data model is also extended to include a serially autocorrelated error term, $\rho \neq 0$, regressors lagged in time, $\boldsymbol{X}_{t-1}$, or endogenous regressors, $\boldsymbol{Z}_t$. The consistency of the difference and system GMM estimators relies on the assumptions that there is no first-order serial autocorrelation in the errors of the level equation, and that the instruments are truly exogenous and therefore valid to define the moment conditions. The Arrelano and Bond (1991) test for serial autocorrelation tests the hypothesis that there is no second-order serial correlation in the first-differenced residuals, which in turn implies that the errors from the level equation are serially uncorrelated. However, correlation coefficients of observations on variables made in single spatial units one year, two years, up to T-1 years apart tend to be large and to diminish only slightly over time (Elhorst 2008b). Consequently, the null hypothesis of no serial autocorrelation of the error terms must often be rejected. One remedial reaction could be to re-estimate the model using methods that assume that the errors are generated by a first-order serial autoregressive process, but this approach has been severely criticized. Rather than improving an initial model when it appears to be unsatisfactory, Hendry (1995, Chap. 4) argues that it is better to start with a more general model containing a series of simpler models nested within it as special cases. The general model Hendry recommends as a generalization of the first-order serial autocorrelation model for time-series data is the first-order serial autoregressive distributed lag model, a linear dynamic regression model in which the dependent variable $\boldsymbol{Y}_t$ is regressed on $\boldsymbol{Y}_{t-1}$ and the explanatory variables $\boldsymbol{X}_t$ and $\boldsymbol{X}_{t-1}$. For this reason, dynamic panel data models extended to include explanatory variables $\boldsymbol{X}_{t-1}$ are more popular than dynamic panel data models extended to include serial autocorrelation. Another reason is that the econometric literature has paid much attention to estimators of the covariance matrix that are robust to serial autocorrelation and heteroskedasticity, affecting inferences regarding the statistical significance of the explanatory variables in the model (Newey and West 1987; Greene 2008).

If one or more of the explanatory variables are endogenous ($\boldsymbol{Z}_t$), they need to be instrumented too. Since the GMM estimator already instruments $\boldsymbol{Y}_{t-1}$, this estimator can easily be extended to include additional endogenous explanatory variables. See Kukenova and Monteiro (2009) how to adjust the GMM estimator when having both endogenous and exogenous explanatory variables ($\boldsymbol{Z}_t$ and $\boldsymbol{X}_t$).

4.4.2 Taxonomy of Dynamic Models in Space and Time

When imposing the parameter restrictions $\tau = \eta = 0$ and $\boldsymbol{\beta}_3 = \boldsymbol{\beta}_4 = \boldsymbol{0}$ on Eq. (4.1), as shown in Fig. 4.2, one obtains the spatial Durbin model. The ins and outs of this model have been extensively discussed in the previous chapter and therefore will not be repeated here.

Fig. 4.3 Dynamic spatial panel data models that have been considered in the literature

1. $\varepsilon_{t-1} + \boldsymbol{W}\varepsilon_t$

2. $\boldsymbol{Y}_{t-1} + \boldsymbol{W}\varepsilon_t$

3. $\boldsymbol{Y}_{t-1} + \boldsymbol{WY}_t + \boldsymbol{WY}_{t-1} + \boldsymbol{X}_t + \boldsymbol{WX}_t$

4. $\boldsymbol{Y}_{t-1} + \boldsymbol{WY}_t + \boldsymbol{WY}_{t-1} + \boldsymbol{X}_t$, no $\boldsymbol{WX}_t$

5. $\boldsymbol{Y}_{t-1} + \boldsymbol{WY}_{t-1} + \boldsymbol{X}_t + \boldsymbol{WX}_t$, no $\boldsymbol{WY}_t$

6. $\boldsymbol{Y}_{t-1} + \boldsymbol{WY}_t + \boldsymbol{WY}_{t-1} + \boldsymbol{X}_t + \boldsymbol{WX}_t$
   Restriction on coefficient of $\boldsymbol{WY}_{t-1}$

7. $\boldsymbol{Y}_{t-1} + \boldsymbol{WY}_t + \boldsymbol{X}_t + \boldsymbol{WX}_t$, no $\boldsymbol{WY}_{t-1}$

Figure 4.3 presents seven different models that have mixed dynamics in both space and time. A first set of studies (model 1 in Fig. 4.3) have mixed space and time in the error term specification. The parameters in Eqs. (4.1b) and (4.1c) that are allowed to vary and those that have been not been included in these studies are reported below in parentheses. Baltagi et al. (2003) consider the testing of spatial error correlation in a model with spatial random effects (λ, $\boldsymbol{\mu}$; but ρ, ξ, κ not included). Baltagi et al. (2007) extend this study to include serial autocorrelation (ρ, λ, $\boldsymbol{\mu}$; but ξ,κ not included). Elhorst (2008a) considers ML estimation of a model with serial and spatial autocorrelation (ρ,λ; but $\boldsymbol{\mu}$, ξ, κ not included). Kapoor et al. (2007) consider GMM estimation of a spatial error model with time-period random effects (λ,ξ; but ρ, $\boldsymbol{\mu}$, κ not included). Baltagi et al. (2012) consider the testing of spatial autocorrelation in both the remainder error term and the spatial random effects (λ, $\boldsymbol{\mu}$, κ; but ρ, ξ not included). Finally, Montes-Rojas (2010) considers the testing of serial error correlation and spatial random effects in a spatial lag model (δ, ρ, $\boldsymbol{\mu}$; but ξ, κ not included). This short overview shows that not every model combination has been considered yet. It is questionable, however, whether more research is needed in this direction. First, the fixed effects model is often more appropriate than the random effects model when modeling spatial panel data (see the discussion in Sect. 3.4). Second, Lee and Yu (2010a) argue that the fixed effects model is robust to and also computationally simpler than the random effects model. Equation (4.1c) can be rewritten as $\boldsymbol{\zeta} = (\boldsymbol{I}_N - \kappa\boldsymbol{W})^{-1}\boldsymbol{\mu}$. Consequently, if $\boldsymbol{\mu}$ is treated as a vector of fixed effects for every spatial unit in the sample, so can $\boldsymbol{\zeta}$ without having to estimate the parameter κ. Likewise, if $\boldsymbol{\mu}$ is treated as a vector of random effects for every spatial unit in the sample, a vector of fixed effects for every spatial unit in the sample $\boldsymbol{\zeta}$ can replace $\boldsymbol{\mu}$ without having to estimate the parameter κ. In other words, by controlling for spatial fixed effects, spatial autocorrelation among the spatial specific effects is automatically accounted for, no matter whether these effects are fixed or random and without having to estimate the magnitude of this form of spatial error autocorrelation. Third, spatial interaction effects among the dependent variable $\boldsymbol{Y}$ and/or the independent variables $\boldsymbol{X}$ are more important than spatial interaction effects among the error terms; when ignoring $\boldsymbol{WY}$ and/or $\boldsymbol{WX}$ variables, the estimator of the remaining parameter estimates will lose its property of being consistent. By contrast, when ignoring spatial interaction effects among the error term, $\boldsymbol{Wv}_t$, the estimator of the remaining parameter estimates will 'only' lose its property of being efficient. Fourth, these type of models cannot be used to determine short-term effects and indirect (spatial spillover) effects (see model 1 in Table 4.1), which are often the main purpose of the analysis.

Perhaps more important is the development of estimators of the covariance matrix that are robust to serial autocorrelation, spatial autocorrelation, and heteroskedasticity, as already pointed out above. Newey and West (1987) derive a consistent estimator of the covariance matrix robust to serial autocorrelation and heteroskedasticity. Similarly, Kelejian and Prucha (2010) derive a consistent estimator of the covariance matrix robust to spatial autocorrelation and heteroskedasticity. Whether these two estimators can be combined and be used in a panel data setting still needs to be investigated. Pesaran and Tosetti (2011) are among the first to consider such an estimator. This study, however, estimates one equation for every spatial unit in the sample, which requires T to be large while T in most space–time studies tends to be small, and does not consider $\boldsymbol{WY}_t$ and $\boldsymbol{WX}_t$ variables in the deterministic regression equation. Whether this approach is still practicable and whether the parameters are identified if this model is extended to include $\boldsymbol{WY}_t$ and $\boldsymbol{WX}_t$ variables is an interesting topic for further research.

A second set of studies (model 2 in Fig. 4.3) have mixed space and time by specifying the deterministic regression equation as a dynamic panel data model and the stochastic error term specification as a spatial error model. Elhorst (2005) considers ML estimation of this model extended to include spatial and time-period fixed effects, and Yang et al. (2006) ML estimation of this model extended to include spatial random effects (but no time-specific effects). Practice has shown that this setup of separating deterministic dynamic effects in time and stochastic interaction effects among different units across space is beneficial. First, it offers the opportunity to control for independent variables lagged in time, $\boldsymbol{X}_{t-1}$. When interaction effects among different units across space are taken up in the regression equation rather than the error term specification, identification of the parameters requires the elimination of the variables $\boldsymbol{X}_{t-1}$ (Anselin et al. 2008). Besides, it also offers the opportunity to adjust the error specification such that endogeneous $\boldsymbol{Z}_t$ variables can be controlled for, also when the model is estimated by ML (see Elhorst 2008b). Second, the forecast performance of these models tends to be much better than that of dynamic panel data model that do not control for spatial autocorrelation (Elhorst 2005; Kholodilin et al. 2008). The disadvantage of this

Table 4.1 Short-term, long-term, direct and indirect (spatial spillover) effects of different models

Type of model	Short-term direct effect	Short-term indirect effect	Long-term direct effect	Long-term indirect effect	Shortcoming
0. Static spatial Durbin model	—	—	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\bar{d}}$	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\overline{rsum}}$	No short-term effects
1. Error terms lagged in space and/or in time	—	—	β_{1k}	—	No short-term and indirect effects No indirect effects
2. Dynamic model + spatial error corr.	β_{1k}	—	$\beta_{1k}/(1-\tau)$	—	No indirect effects
3. Dynamic spatial Durbin model	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\bar{d}}$	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\overline{rsum}}$	$[((1-\tau)\boldsymbol{I}-(\delta+\eta)\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\bar{d}}$	$[((1-\tau)\boldsymbol{I}-(\delta+\eta)\boldsymbol{W})^{-1}(\beta_{1k}\mathbf{I}_N+\beta_{2k}\boldsymbol{W})]^{\overline{rsum}}$	—
4. $\boldsymbol{\beta}_2 = 0$	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N)]^{\bar{d}}$	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N)]^{\overline{rsum}}$	$[((1-\tau)\boldsymbol{I}-(\delta+\eta)\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N)]^{\bar{d}}$	$[((1-\tau)\boldsymbol{I}-(\delta+\eta)\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N)]^{\overline{rsum}}$	Ratio ind./dir effects the same for every X
5. $\delta = 0$	$[(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\bar{d}}$	$[(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\overline{rsum}}$	$[((1-\tau)\boldsymbol{I}-\eta\boldsymbol{W}^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\bar{d}}$	$[((1-\tau)\boldsymbol{I}-\eta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\overline{rsum}}$	No short-term global indirect effects
6. $\eta = -\tau\delta$	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\bar{d}}$	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\overline{rsum}}$	$[\frac{1}{1-\tau}(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\bar{d}}$	$[\frac{1}{1-\tau}(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\overline{rsum}}$	Ratio ind./dir effects constant over time
7. $\eta = 0$	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\bar{d}}$	$[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\overline{rsum}}$	$[((1-\tau)\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\bar{d}}$	$[((1-\tau)\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})]^{\overline{rsum}}$	—

The superscript $\bar{d}$ denotes the operator that calculates the mean diagonal element of a matrix and the superscript $\overline{rsum}$ denotes the operator that calculates the mean row sum of the non-diagonal elements (see Sect. 2.7 for an explanation) type of models, however, is that they cannot be used to determine indirect (spatial spillover) effects (see model 2 in Table 4.1).

A third set of studies (model 3 in Fig. 4.3) have considered a spatial Durbin model extended to include dynamic effects. These studies mainly deal with growth and convergence among countries or regions (Ertur and Koch 2007; Elhorst et al. 2010). Typically, these studies regress economic growth on economic growth in neighboring economies, on the initial income level in the own and in neighboring economies, and on the rates of saving, population growth, technological change and depreciation in the own and in neighboring economies. Elhorst et al. (2010) demonstrate that this economic growth model can be represented by the dynamic regression equation

$$\boldsymbol{Y}_t = \tau \boldsymbol{Y}_{t-1} + \delta \boldsymbol{W}\boldsymbol{Y}_t + \eta \boldsymbol{W}\boldsymbol{Y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta}_1 + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\beta}_2 + \boldsymbol{v}_t \tag{4.5}$$

which may be labeled a dynamic spatial Durbin model. By rewriting this model as

$$\begin{aligned}\boldsymbol{Y}_t = {} & (\boldsymbol{I} - \delta \boldsymbol{W})^{-1}(\tau \boldsymbol{I} + \eta \boldsymbol{W})\boldsymbol{Y}_{t-1} + (\boldsymbol{I} - \delta \boldsymbol{W})^{-1}(\boldsymbol{X}_t\boldsymbol{\beta}_1 + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\beta}_2) \\ & + (\boldsymbol{I} - \delta \boldsymbol{W})^{-1}\boldsymbol{v}_t\end{aligned} \tag{4.6}$$

the matrix of partial derivatives of the expected value of $\boldsymbol{Y}$ with respect to the kth explanatory variable of $\boldsymbol{X}$ in unit 1 up to unit N at a particular point in time can be seen to be

$$\begin{bmatrix} \frac{\partial E(\boldsymbol{Y})}{\partial x_{1k}} & \cdots & \frac{\partial E(\boldsymbol{Y})}{\partial x_{Nk}} \end{bmatrix}_t = (\boldsymbol{I} - \delta \boldsymbol{W})^{-1}[\beta_{1k}\boldsymbol{I}_N + \beta_{2k}\boldsymbol{W}]. \tag{4.7}$$

These partial derivatives denote the effect of a change of a particular explanatory variable in a particular spatial unit on the dependent variable of all other units in the *short term*. Similarly, the *long-term* effects can be seen to be

$$\begin{bmatrix} \frac{\partial E(\boldsymbol{Y})}{\partial x_{1k}} & \cdots & \frac{\partial E(\boldsymbol{Y})}{\partial x_{Nk}} \end{bmatrix} = [(1 - \tau)\boldsymbol{I} - (\delta + \eta)\boldsymbol{W}]^{-1}[\beta_{1k}\boldsymbol{I}_N + \beta_{2k}\boldsymbol{W}]. \tag{4.8}$$

The expressions in (4.7) and (4.8) show that short-term indirect effects do not occur if both $\delta = 0$ and $\beta_{2k} = 0$, while long-term indirect effects do not occur if both $\delta = -\eta$ and $\beta_{2k} = 0$. Debarsy et al. (2012) found similar expressions and also derive formulas for the path along which an economy moves to its long-term equilibrium.

The results reported in Table 4.1 show that this dynamic spatial Durbin model (model 3) can be used to determine short-term and long-term direct effects, and short-term and long-term indirect (spatial spillover) effects. Using the expressions in (4.7) and (4.8), it is also possible to indicate the disadvantages of certain parameter restrictions put forward in previous studies.

The first restriction is $\boldsymbol{\beta}_2 = \boldsymbol{0}$ (model 4 in Fig. 4.3 and Table 4.1). This model is considered in Yu et al. (2008), Lee and Yu (2010b), and Bouayad-Agha and Védrine (2010). The disadvantage of this restriction is that local indirect (spatial spillover) effects are set to zero by construction, as a result of which the indirect effects in relation to the direct effects become the same for every explanatory variable, both in the short term and in the long term. If this ratio happens to be p percent for one variable, it is also p percent for any other variable. This is because β_{1k} in the numerator and β_{1k} in the denominator of this ratio cancel each other out. For example, the ratio for the kth explanatory variable in the short term takes the form

$$\begin{aligned}&\left[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N)\right]^{\overline{rsum}}/\left[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N)\right]^{\overline{d}}=\\&\left[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}\right]^{\overline{rsum}}/\left[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}\right]^{\overline{d}},\end{aligned} \tag{4.9}$$

which shows that it is independent of β_{1k} and thus the same for every explanatory variable. A similar result is obtained when considering this ratio in the long term.

The second restriction that might be imposed is $\delta = 0$ (model 5 in Fig. 4.3 and Table 4.1). This model is considered in LeSage and Pace (2009, Chap. 7) and Korniotis (2010). The disadvantage of this restriction is that the matrix $(\boldsymbol{I}\text{-}\delta\boldsymbol{W})^{-1}$ degenerates to the identity matrix and thus the global short-term indirect (spatial spillover) effect of every explanatory variable to zero. In other words, this model is less suitable if the analysis focuses on spatial spillover effects in the short term.

The third restriction that might be imposed is $\eta = -\tau\delta$ (model 6 in Fig. 4.3 and Table 4.1). This restriction is put forward in Parent and LeSage (2010, 2011). The advantage of this restriction is that the impact of a change in one of the explanatory variables on the dependent variable can be decomposed into a spatial effect and a time effect; the impact over space falls by the factor $\delta\boldsymbol{W}$ for every higher-order neighbor, and over time by the factor τ for every next time period (see Elhorst 2010a for a mathematical derivation). The disadvantage is that the indirect (spatial spillover) effects in relation to the direct effects remain constant over time for every explanatory variable. The ratio of the kth explanatory variable takes the form

$$\left[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})\right]^{\overline{rsum}}/\left[(\boldsymbol{I}-\delta\boldsymbol{W})^{-1}(\beta_{1k}\boldsymbol{I}_N+\beta_{2k}\boldsymbol{W})\right]^{\overline{d}} \tag{4.10}$$

both in the short term and the long term. In other words, if it is p percent for one variable in the short term, it is also p percent for that variable in the long term.

The fourth restriction that might be imposed is $\eta = 0$ (model 7 in Fig. 4.3 and Table 4.1). This model restriction is considered in Franzese and Hays (2007), Kukenova and Monteiro (2009), Elhorst (2010b), Jacobs et al. (2009), and Brady (2011). Although this model also limits the flexibility of the ratio between indirect and direct effects, it seems to be the least restrictive model. More empirical research is needed to find out whether this is really the case.

4.5 Methods of Estimation

Three methods have been developed in the literature to estimate models that have mixed dynamics in both space and time. One method is to bias-correct the maximum likelihood (ML) or quasi-maximum likelihood (QML) estimator, one method is based on instrumental variables or generalized method of moments (IV/GMM), and one method utilizes the Bayesian Markov Chain Monte Carlo (MCMC) approach. These methods are (partly) based on previous studies discussed in the previous section.

Yu et al. (2008) construct a bias corrected estimator for a dynamic model ($\boldsymbol{Y}_{t-1}$, $\boldsymbol{WY}_t$ and $\boldsymbol{WY}_{t-1}$) with spatial fixed effects. Lee and Yu (2010c) extend this study to include time-period fixed effects. They first estimate the model by the ML estimator for the spatial lag model with spatial (and time-period) fixed effects, conditional upon the first observation of every spatial unit in the sample due to the regressors $\boldsymbol{Y}_{t-1}$ and $\boldsymbol{WY}_{t-1}$. Next, they provide a rigorous asymptotic theory for their ML estimator and suggest a bias corrected ML estimator when both the number of spatial units (N) and the number of time points (T) in the sample go to infinity such that the limit between N and T exists and is bounded between zero and infinity ($0 < \lim(N/T) < \infty$). In the words of Lee and Yu (2010b, p. 2), this condition implies that '$T \rightarrow \infty$ where T cannot be too small relative to N'. The bias correction is derived for both normally distributed error terms (ML) and for error terms that do not rely on the normality assumption. In the latter case the first four moments are required (QML). Finally, it is to be noted that this bias corrected ML estimator can also be used when either the variable $\boldsymbol{Y}_{t-1}$ or the variable $\boldsymbol{WY}_{t-1}$ is eliminated from the model.

Elhorst (2010b) investigates the small sample properties of the bias corrected ML estimator. For this purpose, he extends the unconditional ML estimator proposed by Hsiao et al. (2002) with the variable $\boldsymbol{WY}_t$, as well as the Bhargava and Sargan (1983) approximation that is used to determine the expected value and the variance of the first first-differenced observations in the sample. One of his conclusions is that the parameter estimate δ of the variable $\boldsymbol{WY}_t$ is still considerably biased when using this unconditional ML estimator. However, if the parameter estimate δ is based on the bias corrected ML estimator and the other parameters, given δ, on the unconditional ML estimator, then this so-called mixed estimator outperforms the bias corrected estimator of Yu et al. (2008) for small values of T ($T = 5$).

Korniotis (2010) constructs a bias corrected LSDV estimator for a dynamic panel data model ($\boldsymbol{Y}_{t-1}, \boldsymbol{WY}_{t-1}$) with spatial fixed effects, also assuming $0 < \lim(N/T) < \infty$. The bias correction in this study is different from that in Yu et al. (2008), since the LSDV estimator does not have to account for endogenous interaction effects $\boldsymbol{WY}_t$.

A couple of studies have considered IV/GMM estimators, building on previous work of Arrelano and Bond (1991), and Blundell and Bond (1998). Elhorst (2010b) extends the Arrelano and Bond difference GMM estimator to include endogenous interaction effects and finds that this estimator can still be severely biased, especially with respect to the parameter estimate δ of the variable $\boldsymbol{WY}_t$. He notes a bias of 0.061. The explanation for this can be found in Lee and Yu (2010b). They find that a 2SLS estimator like the Arrelano and Bond GMM estimator which is based on lagged values of $\boldsymbol{Y}_{t-1}$, $\boldsymbol{WY}_{t-1}$, $\boldsymbol{X}_t$ and $\boldsymbol{WX}_t$ is not consistent due to too many moments, and that the dominant bias is caused by the endogeneity of the variable $\boldsymbol{WY}_t$ rather than the variable $\boldsymbol{Y}_{t-1}$. To avoid these problems, they propose an optimal GMM estimator based on linear moment conditions, which are standard, and quadratic moment conditions, which are implied by the variable $\boldsymbol{WY}_t$, and therefore not standard in dynamic panel data models. They prove that this GMM estimator is consistent, also when T is small relative to N.

Both Kukenova and Monteiro (2009), and Jacobs et al. (2009) consider a dynamic panel data model ($\boldsymbol{Y}_{t-1}$, $\boldsymbol{WY}_t$) and extend the system GMM estimator of Blundell and Bond (1998) to account for endogenous interaction effects ($\boldsymbol{WY}_t$). The former study also considers endogenous explanatory variables $\boldsymbol{Z_t}$, and the latter spatially autocorrelated error terms $\boldsymbol{W\varepsilon}_t$. The main argument of applying GMM estimators rather than traditional spatial maximum likelihood estimators is that the former can also be used to instrument endogenous explanatory variables (other than the variables $\boldsymbol{Y}_{t-1}$ and $\boldsymbol{WY}_t$).

Both studies find that the system GMM estimator substantially reduces the bias in the parameter estimate of the $\boldsymbol{WY}_t$ variable, and that the system GMM estimator outperforms the Arrelano and Bond difference GMM estimator. The main message of these studies seems to be that the bias Lee and Yu (2010b) have recently found to occur in theory may reduce so strongly that they become acceptable in practice. In Jacobs et al. (2009), the bias in δ of the variable $\boldsymbol{WY}_t$ amounts to 0.50 % of the true parameter value, on average. On the other hand, Monte Carlo simulation experiments can only cover a limited number of situations and therefore do not prove that these results hold in general. Kukenova and Monteiro (2009), for example, only consider positive values for the spatial autoregressive coefficient τ of the variable $\boldsymbol{WY}_t$. Furthermore, in some cases both studies also find biases that are rather large. For $T = 10$, $N = 50$, and $\delta = 0.3$, for example, Kukenova and Monteiro (2009, appendix 6.C) find a bias of -0.0219, or 7.3 % of the true parameter value. Comparably, Jacobs et al. (2009, Table A.1) find an increasing bias, up to 6.1 % of the true parameter value, in the spatial autoregressive coefficient τ of the variable $\boldsymbol{WY}_\mathrm{t}$, provided that spatial autocorrelation in the error terms is not accounted for. When correcting for spatial error correlation, this bias considerably diminishes, but then the bias in the spatial autocorrelation coefficient λ increases Table 4.2.

Parent and LeSage (2010, 2011) point out that the Bayesian MCMC approach considers conditional distributions of each parameter of interest conditional on the others, which leads to some computational simplification. Just as in Elhorst (2001, 2005, 2010b), they treat the first period cross-section as endogenous, using the Bhargava and Sargan (1983) approximation. They find that the correct treatment of the initial observations (endogenous instead of exogenous) is important, especially in cases when T is small. Since Yu et al. (2008) and Elhorst (2010b) find that maximizing the log-likelihood function leads to biased estimates of the spatial autoregressive parameter δ of the variable $\boldsymbol{WY}_t$, the former when considering the log-likelihood conditional upon the first cross-section of observations and the latter when considering the unconditional log-likelihood, the question arises whether the Bayesian MCMC estimator is not also subject to a bias. For $T = 5$, $N = 50$, and $\delta = 0.7$, for example, Parent and LeSage (2011, Table 4.3) find a bias of 0.0149, or 2.13 % of the true parameter value.

4.6 Non-Stability

The estimation of a dynamic spatial panel data model gets more complicated if it turns out that the condition $\tau + \delta + \eta < 1$ is not satisfied, i.e., if the model is unstable. To get rid of possible unstable components in $\boldsymbol{Y}_t$, Lee and Yu (2010a), and Yu et al. (2012) propose to transform the model in spatial first-differences, that is, by taking every variable in the dynamic spatial panel data model in deviation of its spatially lagged value. If the following spatial dynamic panel data model

$$\boldsymbol{Y}_t = \tau \boldsymbol{Y}_{t-1} + \delta \boldsymbol{W} \boldsymbol{Y}_t + \eta \boldsymbol{W} \boldsymbol{Y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\mu} + \xi_t \boldsymbol{\iota}_N + \boldsymbol{\varepsilon}_t \tag{4.11}$$

is taken as point of departure, this operation is mathematically equivalent with multiplying (4.11) by the matrix ($\boldsymbol{I}$–$\boldsymbol{W}$)

$$\begin{aligned}(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{Y}_t = \tau(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{Y}_{t-1} + \delta \boldsymbol{W}(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{Y}_t + \eta \boldsymbol{W}(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{Y}_{t-1} \\ + (\boldsymbol{I} - \boldsymbol{W})\boldsymbol{X}_t \boldsymbol{\beta} + (\boldsymbol{I} - \boldsymbol{W})\mu + (\boldsymbol{I} - \boldsymbol{W})\boldsymbol{\varepsilon}_t\end{aligned} \tag{4.12}$$

where we made use of the property ($\boldsymbol{I}$–$\boldsymbol{W}$)$\boldsymbol{W}$ = $\boldsymbol{W}$($\boldsymbol{I}$–$\boldsymbol{W}$). The resulting equation has some important properties, which require further explanation. First, since $\xi_t(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{\iota}_N = \boldsymbol{0}$, all time-period fixed effects are eliminated from the model. Note that these fixed effects do remain effective, since the estimation of a model formulated in levels produces the same parameter estimates as the estimation of that model reformulated in spatial first-differences without time fixed effects. The reason to renounce the first approach is that we also want to remove the inconsistency caused by the possibly unstable character of $\boldsymbol{Y}_t$. Second, just as first differencing in time would reduce the number of observations available for estimation, so does first differencing in space; the former by one for every country and the latter by one for every time period. Third, since we assumed that ε_{it} has zero mean and variance σ^2, the variance of ($\boldsymbol{I}$–$\boldsymbol{W}$)$\boldsymbol{\varepsilon}_t$ is $\sigma^2\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma} = (\boldsymbol{I}-\boldsymbol{W})(\boldsymbol{I}-\boldsymbol{W})^{\mathrm{T}}$. This change of the variance–covariance matrix from $\sigma^2\boldsymbol{I}$ into $\sigma^2\boldsymbol{\Sigma}$ needs to be accounted for when estimating the parameters of Eq. (4.12).

Since the eigenvalues of the matrix ($\boldsymbol{I}$–$\boldsymbol{W}$) are equal to $1-\omega_i$ (i = 1, …, N), where each ω_i denotes a particular eigenvalue of the spatial weights matrix $\boldsymbol{W}$, and the largest eigenvalue of $\boldsymbol{W}$ is one ($\omega_{\max} = 1$), provided that $\boldsymbol{W}$ is normalized, at least one eigenvalue of the matrix ($\boldsymbol{I}$–$\boldsymbol{W}$) will be zero. This implies that the determinant of ($\boldsymbol{I}$–$\boldsymbol{W}$) equals zero and thus that this matrix does not have full rank. If ($\boldsymbol{I}$–$\boldsymbol{W}$) does have rank $N-1$ instead of N, so does $\boldsymbol{\Sigma}$.[1] Let $\boldsymbol{\Lambda}_{N-1}$ denote the (N-1) × (N-1) diagonal matrix of nonzero eigenvalues of $\boldsymbol{\Sigma}$, and $\boldsymbol{F}_{N,N-1}$ the corresponding orthonormal $N \times (N$-1) matrix of eigenvectors. Then we can transform Eq. (4.12) by the matrix $\boldsymbol{P} = \boldsymbol{\Lambda}_{N-1}^{-1/2}\boldsymbol{F}_{N,N-1}^T$, to get

[1] If $\boldsymbol{W}$ has more than just one eigenvalue that is equal to 1, say p, the number $N-1$ must be adjusted to $N-p$.

$$\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{Y}_t = \tau\boldsymbol{P}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{Y}_{t-1} + \delta\boldsymbol{PW}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{Y}_t + \eta\boldsymbol{PW}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{Y}_{t-1} + \boldsymbol{P}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{P}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{\mu} + \boldsymbol{P}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{\varepsilon}_t \tag{4.13}$$

This transformation has three effects. First, since $\boldsymbol{P}$ is a $(N-1) \times N$ matrix, the transformation $\boldsymbol{Y}_t^* = \boldsymbol{P}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{Y}_t$ reduces the length of $\boldsymbol{Y}_t^*$ to $N-1$. The same applies to the length of the transformed matrices or vectors $\boldsymbol{X}_t^*$, $\boldsymbol{\mu}^*$ and $\boldsymbol{\varepsilon}_t^*$. It is the perfect linear combination among the observations due to the multiplication of Eq. (4.11) by the matrix $(\boldsymbol{I}-\boldsymbol{W})$ that causes the number of observations to go down. Note however that this decrease in the number of observations is merely a reduction in the number of degrees of freedom, since the information of all N observations is still implied in the data. Second, since the transformation $\boldsymbol{P}$ reverses the transformation $(\boldsymbol{I}-\boldsymbol{W})$ (except for degrees of freedom), we have $E(\boldsymbol{\varepsilon}_t^*\boldsymbol{\varepsilon}_t^{*T}) = \sigma^2\boldsymbol{I}_{N-1}$. Third, since

$\boldsymbol{W}^* \equiv \boldsymbol{PW}(\boldsymbol{I}-\boldsymbol{W}) = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{F}_{N,N-1}^T\boldsymbol{W}\boldsymbol{F}_{N,N-1}\boldsymbol{\Lambda}^{1/2}$ (see Lee and Yu, 2010a), Eq. (4.13) can be rewritten as

$$\boldsymbol{Y}_t^* = \tau\boldsymbol{Y}_{t-1}^* + \delta\boldsymbol{W}^*\boldsymbol{Y}_t^* + \eta\boldsymbol{W}^*\boldsymbol{Y}_{t-1}^* + \boldsymbol{X}_t^*\boldsymbol{\beta} + \boldsymbol{\mu}^* + \boldsymbol{\varepsilon}_t^* \tag{4.14}$$

whose parameters can be consistently estimated by the same bias corrected (Q)ML estimator that is used to estimate Eq. (4.11). Yu et al. (2012) formally show that the transformed model will be stable if $\tau + \omega_{max-1}(\delta+\eta) < 1$, where $\omega_{\max-1}$ denotes the second largest eigenvalue of the spatial weights matrix $\boldsymbol{W}$. Importantly, the latter restriction is less restrictive than the original restriction $\tau + \delta + \eta < 1$.

In sum, there are three differences between Eqs. (4.14) and (4.11). First, the number of degrees of freedom in each time period is $N-1$ instead of N. Second, time-period fixed effects are wiped out, although their effectiveness has not. Third, the matrix $\boldsymbol{W}$ is replaced by $\boldsymbol{W^*}$. It is important to note that the row elements of $\boldsymbol{W^*}$, in contrast to those in $\boldsymbol{W}$, do not necessarily sum up to one. Nevertheless, the $N-1$ eigenvalues of $\boldsymbol{W^*}$ are identical to those of $\boldsymbol{W}$ that remain after the unit eigenvalue of $\boldsymbol{W}$ is excluded.

Yu et al. (2012) show that the dynamic spatial panel data model in (4.11) also has a revealing error correction model (ECM) representation

$$\Delta\boldsymbol{Y}_t = (\boldsymbol{I}-\delta\boldsymbol{W})^{-1}[(\tau-1)\boldsymbol{I} + (\delta+\eta)\boldsymbol{W}]\boldsymbol{Y}_{t-1} + (I-\delta\boldsymbol{W})^{-1}\boldsymbol{X}_t\beta + (\boldsymbol{I}-\delta\boldsymbol{W})^{-1}[\boldsymbol{\mu} + \xi_{t0}\imath_N + \varepsilon_t] \tag{4.15}$$

where $\Delta\boldsymbol{Y}_t = \boldsymbol{Y}_t - \boldsymbol{Y}_{t-1}$ denotes first differences in time, the change of the dependent variable in time. This equation may be rewritten as

$$\Delta\boldsymbol{Y}_t = \delta\boldsymbol{W}\Delta\boldsymbol{Y}_t + (\tau-1)\boldsymbol{Y}_{t-1} + (\delta+\eta)\boldsymbol{W}\boldsymbol{Y}_{t-1} + \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{\mu} + \xi_{t0}\boldsymbol{\imath}_N + \boldsymbol{\varepsilon}_t \tag{4.16}$$

which in turn can also be obtained from Eq. (4.11) by subtracting $\boldsymbol{Y}_{t-1}$ from and adding $\rho\boldsymbol{W}\boldsymbol{Y}_{t-1}$ to both sides of this equation and rearranging terms. Equation (4.16) shows that the change in the dependent variable in period t is explained not only by the initial level of the dependent variable of the spatial unit itself in the preceding year ($\boldsymbol{Y}_{t-1}$), but also by the initial levels of the dependent variables of other spatial units countries in the preceding period ($\boldsymbol{WY}_{t-1}$). The coefficient of the initial level in the spatial unit itself, ($\tau-1$), is smaller than zero, indicating that an increase of the dependent variable becomes more likely if the initial value of that variable is relatively low. The sign of the coefficient of the initial levels in other spatial units, ($\delta + \eta$), is expected to be positive, indicating that an increase of the dependent variable becomes also more likely if the initial value of that variable in other spatial units is already relatively high.

The mathematical formulas of the direct and indirect effects of a change in one of the explanatory variables on the dependent variable $\boldsymbol{Y}$ in the short term and in the long term have been shown in Eqs. (4.7) and (4.8). If a study focuses on the error correction model representation and, related to that, on the direct and indirect effects estimates of changes in the explanatory variables on the change in the dependent variable, $\Delta \boldsymbol{Y}_t$, these effects estimates at a particular point in time t take the form

$$\begin{bmatrix} \frac{\partial E(\Delta \boldsymbol{Y})}{\partial x_{1k}} & . & \frac{\partial E(\Delta \boldsymbol{Y})}{\partial x_{Nk}} \end{bmatrix}_t = \begin{bmatrix} \frac{\partial \Delta y_1}{\partial x_{1k}} & . & \frac{\partial \Delta y_1}{\partial x_{Nk}} \\ . & . & . \\ \frac{\partial \Delta y_N}{\partial x_{1k}} & . & \frac{\partial \Delta y_N}{\partial x_{Nk}} \end{bmatrix}_t = (\boldsymbol{I} - \delta \boldsymbol{W})^{-1} \beta_k \tag{4.17}$$

Since these effects are independent from the time index, the right-hand side of (4.17) in contrast to the left-hand side no longer contains the symbol t.

In a similar fashion, one can calculate the effects estimates of convergence. Using the ECM presentation in Eq. (4.15), we get

$$\frac{\partial E(\Delta \boldsymbol{Y}_t)}{\partial Y_{t-1}} = (\boldsymbol{I} - \delta \boldsymbol{W})^{-1} [(\tau - 1)\boldsymbol{I} + (\delta + \eta)\boldsymbol{W}] \tag{4.18}$$

The average diagonal element of this matrix measures the strength of the convergence effect of the spatial units themselves, and the average row sum of the off-diagonal elements the convergence effect of other spatial units.

A specific situation occurs if $\tau + \delta + \eta = 1$. Yu et al. (2012) label this situation as spatial cointegration, after conventional cointegration in the time series literature. The cointegration matrix is ($\boldsymbol{I}-\boldsymbol{W}$) and the cointegration rank is the number of eigenvalues of $\boldsymbol{W}$ that are smaller than 1, which is $N-1$ (but note footnote 1). If $\tau + \delta + \eta = 1$, we have

$$\frac{\partial E(\Delta \boldsymbol{Y}_t)}{\partial \boldsymbol{Y}_{t-1}} = (\tau - 1)(\boldsymbol{I} - \delta \boldsymbol{W})^{-1} (\boldsymbol{I} - \boldsymbol{W}) \tag{4.19}$$

If $\boldsymbol{W}$ is normalized, the total effect of this partial derivative equals zero by construction. This is due to the perfect linear combination among the observations caused by the matrix ($\boldsymbol{I}$–$\boldsymbol{W}$). The direct effect of this matrix is 1, since the diagonal elements of $\boldsymbol{I}$ are one and of $\boldsymbol{W}$ are zero. The indirect effect of this matrix is -1, since the off-diagonal elements of $\boldsymbol{I}$ sum up to zero and of $\boldsymbol{W}$ to one, provided that $\boldsymbol{W}$ is row-normalized. In sum, the total effect is $1-1 = 0$. In other words, the change in the dependent variable in different spatial units over time does not converge if this variable turns out to be spatially cointegrated.

It should be noted that the direct and indirect effects on the change in the dependent variable, $\Delta \boldsymbol{Y}_t$, are identical to the short-term direct and indirect effects of changes in the explanatory variables on the level of the dependent variable $\boldsymbol{Y}_t$. This property, however, does not hold for the long-term effects. Mathematically, the long-term effects boil down to $[(1\text{-}\tau)\boldsymbol{I}\text{-}(\delta+\eta)\boldsymbol{W}]^{-1}\boldsymbol{\beta}_{k.}$. If the model is spatially cointegrated, the matrix in square brackets is singular, as a result of which the long-term direct and indirect effects are not defined. This justifies the treatment of $\tau + \delta + \eta = 1$ as a special case and the focus on $\Delta \boldsymbol{Y}_t$ rather than $\boldsymbol{Y}_t$.

In a study on financial liberalization of 62 countries over the period 1976–2005, Elhorst et al. (2013) estimate a dynamic spatial panel data model in levels, as well as a model reformulated in spatial first-differences. They find that the coefficient estimates of the first specification point to unstable components in the dependent variable, whereas the coefficient estimates of the second specification do not. Up to now, this is one of the few empirical studies that have found that taking spatial first-differences is an effective tool to obtain a stable model.

4.7 Empirical Illustration

Baltagi and Li (2004) estimate a demand model for cigarettes based on a panel from 46 U.S. states in which real per capita sales of cigarettes by persons of smoking age (14 years and older) measured in packs of cigarettes per capita (C_{it}) is regressed on the average retail price of a pack of cigarettes measured in real terms (P_{it}) and on real per capita disposable income (Y_{it}). Moreover, all variables are taken in logs. Whereas Baltagi and Li (2004) use the first 25 years for estimation to reserve data for out of sample forecasts, we use the full data set covering the period 1963–1992. This dataset can be downloaded freely from www.wiley.co.uk/baltagi/, while an adapted version is available at www.regroningen.nl/elhorst. More details, as well as reasons to include state-specific effects (μ_i) and time-specific effects (ξ_t), were given in the previous chapter.

Column (1) of Table 4.2 reports the estimation results when adopting a non-dynamic spatial Durbin model with spatial and time-period fixed effects. In the previous chapter it was found that the model specification with spatial and time-period fixed effects outperformed its counterparts without spatial and/or time-period fixed effects, as well as the random effects model.

Table 4.2 Estimation results of cigarette demand using different model specifications

Determinants	(1) Non-dynamic spatial Durbin model with fixed effects	Dynamic spatial Durbin model with fixed effects
$\log(C)_{-1}$		0.865
		(65.04)
W*log(C)	0.264	0.076
	(8.25)	(2.00)
W*$\log(C)_{-1}$		_0.015
		(_0.29)
log(P)	_1.001	_0.266
	(_24.36)	(_13.19)
log(Y)	0.603	0.100
	(10.27)	(4.16)
W*log(P)	0.093	0.170
	(1.13)	(3.66)
W*log(Y)	_0.314	_0.022
	(_3.93)	(_0.87)
R^2	0.902	0.977
logL	1691.4	2623.3

Notes t-values in parentheses

The main shortcoming of a non-dynamic spatial Durbin model is that it cannot be used to calculate short-term effect estimates of the explanatory variables. This is made clear in Table 4.3, which reports the corresponding effects estimates of the models presented in Table 4.2; since a non-dynamic model only produces long-term effects estimates, the cells reporting short-term effects estimates are left empty.

Table 4.3 Effects estimates of cigarette demand using different model specifications

Determinants	(1) Non-dynamic spatial Durbin model with fixed effects	(2) Dynamic spatial Durbin model with fixed effects
Short-term direct effect Log(P)		−0.262 (−11.48)
Short-term indirect effect Log(P)		0.160 (3.49)
Short-term direct effect Log(Y)		0.099 (3.36)
Short-term indirect effect Log(Y)		−0.018 (−0.45)
Long-term direct effect Log(P)	−1.013 (−24.73)	−1.931 (−9.59)
Long-term indirect effect Log(P)	−0.220 (−2.26)	0.610 (0.98)
Long-term direct effect Log(Y)	0.594 (10.45)	0.770 (3.55)
Long-term indirect effect Log(Y)	−0.197 (−2.15)	0.345 (0.48)

The direct effects estimates of the two explanatory variables reported in column (1) of Table 4.3 are significantly different from zero and have the expected signs. Higher prices restrain people from smoking, while higher income levels have a positive effect on cigarette demand. The price elasticity amounts to−1.013 and the income elasticity to 0.594. Note that these direct effects estimates are different from the coefficient estimates of −1.001 and 0.603 reported in column (1) of Table 4.2 due to feedback effects that arise as a result of impacts passing through neighboring states and back to the states themselves.

The spatial spillover effects (indirect effects estimates) of both variables are negative and significant. Own-state price increases will restrain people not only from buying cigarettes in their own state, but to a limited extent also from buying cigarettes in neighboring states (elasticity -0.220). By contrast, whereas an income increase has a positive effect on cigarette consumption in the own state, it has a negative effect in neighboring states. We come back to this result below. Further note that the non-dynamic spatial Durbin model without spatial and time-period effects indicates a positive rather than a negative spatial spillover effect of price increases, and that only a positive outcome would be consistent with Baltagi and Levin (1992), who found that price increases in a particular state—due to tax increases meant to reduce cigarette smoking and to limit the exposure of non-smokers to cigarette smoke—encourage consumers in that state to search for cheaper cigarettes in neighboring states. However, there are two reasons why this comparison is invalid. First, whereas Baltagi and Levin's (1992) model is dynamic, it is not spatial. They do consider the price of cigarettes in neighboring states, but not any other spatial interaction effects. Second, whereas our model contains spatial interaction effects, it is not (yet) dynamic. For these reasons it is interesting to consider the estimation results of our dynamic spatial panel data model.

Column (2) of Table 4.3 reports the direct and indirect effects of the dynamic model, both in the short term and long term. Consistent with microeconomic theory, the short-term direct effects appear to be substantially smaller than the long-term direct effects; -0.262 versus-1.931 for the price variable and 0.099 versus 0.770 for the income variable. This is because it takes time before price and income changes fully settle. The long-term direct effects in the dynamic spatial Durbin model, on their turn, appear to be greater (in absolute value) than their counterparts in the non-dynamic spatial Durbin model; -1.931 versus -1.013 for the price variable and 0.770 versus 0.594 for the income variable. Apparently, the non-dynamic model underestimates the long-term effects. The short-term spatial spillover effect of a price increase turns out to be positive; the elasticity amounts to 0.160 and is highly significant (t-value 3.49). This finding is in line with the original finding of Baltagi and Levin (1992) in that a price increase in one state encourages consumers to search for cheaper cigarettes in neighboring states. The negative spatial spillover effect of a price increase we found earlier for the non-dynamic spatial Durbin model demonstrates that a non-dynamic approach falls short here. Although greater and again positive, we do not find empirical evidence that the long-term spatial spillover effect of a price increase is also significant. A similar result is found by Debarsy et al. (2012). It is to be noted that they estimate the parameters of the model by the Bayesian MCMC estimator developed by Parent and LeSage (2010, 2011), whereas we use the bias corrected ML estimator developed by (Lee and Yu 2010c). Furthermore, the spatial weights matrix used in that study is based on lengths of state borders in common between each state and its neighboring states, whereas we use a binary contiguity matrix.

The long-term spatial spillover effect of the income variable derived from the dynamic spatial panel data model appears to be positive, which suggests that an income increase in a particular state has a positive effect on smoking not only in that state itself, but also in neighboring states. Furthermore, the spatial spillover effect is smaller than the direct effect, which makes sense since the impact of a change will most likely be larger in the place that instigated the change. However, the spatial spillover effect of an income increase is not significant. A similar result is found by Debarsy et al. (2012). Interestingly, the spatial spillover effect of the income variable in the non-dynamic spatial panel data model appeared to be negative and significant. Apparently, the decision whether to adopt a dynamic or a non-dynamic model represents an important issue. Some researchers prefer simpler models to more complex ones (Occam's razor). One problem of complex models is overfitting, the fact that excessively complex models are affected by statistical noise, whereas simpler models may capture the underlying process better and may thus have better predictive performance. However, if one can trade simplicity for increased explanatory power, the complex model is more likely to be the correct one.

To investigate whether the extension of the non-dynamic model to the dynamic spatial panel data model increases the explanatory power of the model, one may test whether the coefficients of the variables $\boldsymbol{Y}_{t-1}$ and $\boldsymbol{WY}_{t-1}$ are jointly significant using an LR-test. The outcome of this test ($2 \times (2623.3-1691.4) = 1863.8$ with 2 df) evidently justifies the extension of the model with dynamic effects.

4.8 Conclusion

At the turn of century there was no straightforward estimation procedure for dynamic spatial panel data models. Today, they can be estimated by bias-corrected ML or QML, IV/GMM, and Bayesian MCMC methods. However, many problems remain. One problem is the bias in the coefficient δ of the variable $\boldsymbol{WY}_t$; not every method is able to tackle that bias sufficiently. Another problem is the performance of some estimators when T is small; treating the initial observations endogenous instead of exogenous maybe beneficial under these circumstances. A third problem is that not every estimator is able to deal with endogenous explanatory variables other than the dependent variables lagged in space and/or time. A final problem is that the stationarity conditions that need to be imposed on the parameters of the model are not always implemented correctly.

A dynamic panel data model can take several forms. In this chapter we presented the most popular ones. Each form appeared to have certain shortcomings. Dependent on the purpose of a particular empirical study, it is the researcher to determine which form is most appropriate.

References

Anselin L (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht

Anselin L, Le Gallo J, Jayet H (2008) Spatial panel econometrics. In: Matyas L, Sevestre P (eds) The econometrics of panel data, fundamentals and recent developments in theory and practice, 3rd edn. Kluwer, Dordrecht, pp 901–969

Arrelano M (2003) Panel data econometrics. Oxford University Press, Oxford

Arrelano M, Bond S (1991) Some tests of specification for panel data: monte carlo evidence and an application to employment equations. Rev Econ Stud 58(2):277–297

Baltagi BH (2005) Econometric analysis of panel data, 3rd edn. Wiley, Chichester

Baltagi BH, Levin D (1992) Cigarette taxation: raising revenues and reducing consumption. Struct Change Econ Dyn 3(2):321–335

Baltagi BH, Li D (2004) Prediction in the panel data model with spatial autocorrelation. In: Anselin L, Florax RJGM, Rey SJ (eds) Advances in spatial econometrics: methodology, tools, and applications. Springer, Berlin, pp 283–295

Baltagi BH, Song SH, Koh W (2003) Testing panel data models with spatial error correlation. J Econometrics 117(1):123–150

Baltagi BH, Song SH, Jung BC, Koh W (2007) Testing for serial correlation, spatial autocorrelation and random effects using panel data. J Econometrics 140(1):5–51

Baltagi BH, Egger P, Pfaffermayr M (2012) A generalized spatial panel data model with random effects. CESifo Working Paper Series No. 3930. Available at SSRN: http://ssrn.com/abstract=2145816

Bhargava A, Sargan JD (1983) Estimating dynamic random effects models from panel data covering short time periods. Econometrica 51(6):1635–1659

Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. J Econometrics 87(1):115–143

Bouayad-Agha S, Védrine L (2010) Estimation strategies for a spatial dynamic panel using GMM; a new approach to the convergence issue of European regions. Spat Econ Anal 5(2):205–227

Brady RR (2011) Measuring the diffusion of housing prices across space and time. J Appl Econometrics 26(2):213–231

Bun M, Carree M (2005) Bias-corrected estimation in dynamic panel data models. J Bus Econ Stat 3(2):200–211

Debarsy N, Ertur C, LeSage JP (2012) Interpreting dynamic space-time panel data models. Stat Methodol 9(1–2):158–171

Elhorst JP (2001) Dynamic models in space and time. Geogr Anal 33(2):119–140

Elhorst JP (2005) Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. Geogr Anal 37(1):62–83

Elhorst JP (2008a) Serial and spatial autocorrelation. Econ Lett 100(3):422–424

Elhorst JP (2008b) A spatiotemporal analysis of aggregate labour force behaviour by sex and age across the European union. J Geogr Syst 10(2):167–190

Elhorst JP (2010a) Applied spatial econometrics: raising the bar. Spat Econ Anal 5(1):9–28

Elhorst JP (2010b) Dynamic panels with endogenous interaction effects when T is small. Reg Sci Urban Econ 40(5):272–282

Elhorst JP, Piras G, Arbia G (2010) Growth and convergence in a multi-regional model with space-time dynamics. Geogr Anal 42(3):338–355

Elhorst JP, Zandberg E, de Haan J (2013) The impact of interaction effects among neighboring countries on financial liberalization and reform; a dynamic spatial panel data approach. Spat Econ Anal. doi:10.1080/17421772.2012.760136

Enders W (1995) Applied econometric time series. Wiley, New York

Ertur C, Koch W (2007) Growth, technological interdependence and spatial externalities: theory and evidence. J Appl Econometrics 22(6):1033–1062

Franzese RJ Jr, Hays JC (2007) Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. Polit Anal 15(2):140–164

Greene WH (2008) Econometric analysis, 6th edn. Pearson, New Jersey

Hahn J, Kuersteiner G (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Econometrica 70(4):1639–1657

Hamilton JD (1994) Time series analysis. Princeton University Press, New Jersey

Hendry DF (1995) Dynamic econometrics. Oxford University Press, Oxford

Hsiao C (2003) Analysis of panel data, 2nd edn. Cambridge University Press, Cambridge

Hsiao C, Pesaran MH, Tahmiscioglu AK (2002) Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. J Econometrics 109(1):107–150

Jacobs JPAM, Ligthart JE, Vrijburg H (2009) Dynamic panel data models featuring endogenous interaction and spatially correlated errors. http://ideas.repec.org/p/ays/ispwps/ paper0915.html

Kapoor M, Kelejian HH, Prucha IR (2007) Panel data models with spatially correlated error components. J Econometrics 140(1):97–130

Kelejian HH, Prucha IR (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. J Econometrics 157(1):53–67

Kholodilin KA, Siliverstovs B, Kooths S (2008) A dynamic panel data approach to the forecasting of the GDP of German Länder. Spat Econ Anal 3(2):195–207

Kiviet JF (1995) On bias, inconsistency and efficiency of some estimators in dynamic panel data models. J Econometrics 68(1):53–78

Korniotis GM (2010) Estimating panel models with internal and external habit formation. J Bus Econ Stat 28(1):145–158

Kukenova M, Monteiro JA (2009) Spatial dynamic panel model and system GMM: A monte carlo investigation. http://ideas.repec.org/p/pra/mprapa/11569.html

Lee LF, Yu J (2010a) Some recent developments in spatial panel data models. Reg Sci Urban Economics 40:255–271

Lee LF, Yu J (2010b) Efficient GMM estimation of spatial dynamic panel data models with fixed effects. http://www.economics.smu.edu.sg/events/Paper/LungfeiLee.pdf

Lee LF, Yu J (2010c) A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theor 26(2):564–597

LeSage JP, Pace RK (2009a) Introduction to spatial econometrics. CRC Press Taylor & Francis Group, Boca Raton

Montes-Rojas GV (2010) Testing for random effects and serial correlation in spatial autoregressive model. J Stat Plan Infer 140:1013–1020

Nerlove M (1999) Properties of alternative estimators of dynamic panel models: an empirical analysis of cross-country data for the study of economic growth. In: Hsiao C, Lahiri K, Lee LF, Pesaran MH (eds) Analysis of panels and limited dependent variable models. Cambridge University Press, Cambridge, pp 136–170

Newey W, West K (1987) A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55(3):703–708

Nickell S (1981) Biases in dynamic models with fixed effects. Econometrica 49(6):1417–1426

Parent O, LeSage JP (2010) A spatial dynamic panel model with random effects applied to commuting times. Transp Res Part B 44:633–645

Parent O, LeSage JP (2011) A space-time filter for panel data models containing random effects. Comput Stat Data Anal 55:475–490

Pesaran MH, Tosetti E (2011) Large panels with common factors and spatial correlation. J Econometrics 161(2):182–202

Yang Z, Li C, Tse YK (2006) Functional form and spatial dependence in spatial panels. Econ Lett 91(1):138–145

Yu J, de Jong R, Lee L (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. J Econometrics 146(1):118–134

Yu J, de Jong R, Lee LF (2012) Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration. J Econometrics 167:16–37