C.2 Spatial Panel Data Models

J. Paul Elhorst

C.2.1 Introduction

In recent years, 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 spatial units (zip codes, municipalities, regions, states, jurisdictions, countries, etc.). This interest can be explained 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. Panel data are generally more informative, and they contain more variation and 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 Hsiao 2005 for more details).

Elhorst (2003) has provided a review of issues arising in the estimation of four panel data models commonly used in applied research extended to include spatial error autocorrelation or a spatially lagged dependent variable: fixed effects, random effects, fixed coefficients, and random coefficients models. In addition, Matlab routines to estimate the fixed effects and random effects models have been provided at his website, see <www.regroningen.nl/elhorst> or <http://www.rug.nl/staff/j.p.elhorst/projects>. Many studies have applied these routines by now to estimate regional labor market models, economic growth models, public expenditures or tax setting models, and agricultural models. These applications have led to new insights, developments and extensions, but also to new questions and misunderstandings. This chapter reviews and organizes these recent methodologies. It deals with the possibility to test for spatial interaction effects in standard panel data models, the estimation of fixed effects and the determination of their significance levels, the possibility to test the fixed effects specification against the random effects specification of panel data models extended to include spatial error autocorrelation or a spatially lagged dependent variable using Hausman's specifi-

M.M. Fischer and A. Getis (eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Application*s, DOI 10.1007/978-3-642-03647-7_19,

cation test, the determination of the variance-covariance matrix of the parameter estimates of these extended models, the determination of goodness-of-fit measures and the best linear unbiased predictor when using these models for prediction purposes. For reasons of space, attention is limited to models with spatial fixed effects or spatial random effects. The concluding section also briefly discusses the possibility to test for endogeneity of one or more of the explanatory variables and the possibility to include dynamic effects.

C.2.2 Standard models for spatial panels

First, a simple pooled linear regression model with spatial specific effects is considered, but without spatial interaction effects

$$y_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + \mu_i + \varepsilon_{it} \tag{C.2.1}$$

where i is an index for the cross-sectional dimension (spatial units), with $i = 1, ..., N$, and t is an index for the time dimension (time periods), with $t = 1, ..., T$. y_{it} is an observation on the dependent variable at i and t, $\boldsymbol{X}_{it}$ an 1-by-K row vector of observations on the independent variables, and $\boldsymbol{\beta}$ a matching K-by-1 vector of fixed but unknown parameters. ε_{it} is an independently and identically distributed error term for i and t with zero mean and variance σ^2, while μ_i denotes a spatial specific effect. The standard reasoning behind spatial specific effects is that they control for all space-specific time-invariant variables whose omission could bias the estimates in a typical cross-sectional study.

When specifying interaction between spatial units, the model may contain a spatially lagged dependent variable or a spatial autoregressive process in the error term, known as the spatial lag and the spatial error model, respectively. The spatial lag model posits that the dependent variable depends on the dependent variable observed in neighboring units and on a set of observed local characteristics

$$y_{it} = \delta \sum_{j=1}^{N} W_{ij} \boldsymbol{y}_{jt} + \boldsymbol{X}_{it}\boldsymbol{\beta} + \mu_i + \varepsilon_{it} \tag{C.2.2}$$

where δ is called the spatial autoregressive coefficient and W_{ij} is an element of a spatial weights matrix $\boldsymbol{W}$ describing the spatial arrangement of the units in the sample. It is assumed that $\boldsymbol{W}$ is a pre-specified non-negative matrix of order N.[1] According to Anselin et al. (2006, p.6), the spatial lag model is 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 the neighboring agents. In the empirical literature on strategic interaction among local governments, for example, the spatial lag model is 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 spatial error model, on the other hand, posits that the dependent variable depends on a set of observed local characteristics and that the error terms are correlated across space

$$y_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + \mu_i + \phi_{it} \tag{C.2.3a}$$

$$\phi_{it} = \rho \sum_{j=1}^{N} W_{ij}\,\phi_{jt} + \varepsilon_{it} \tag{C.2.3b}$$

where ϕ_{it} reflects the spatially autocorrelated error term and ρ is called the spatial autocorrelation coefficient. According to Anselin et al. (2006, p.7), a spatial error specification does not require a theoretical model for a spatial or social interaction process, but, instead, is a special case of a non-spherical error covariance matrix. In the empirical literature on strategic interaction among local governments, the spatial error model is consistent with a situation where determinants of taxation or expenditures on public services omitted from the model are spatially autocorrelated, and with a situation where unobserved shocks follow a spatial pattern. A spatially autocorrelated error term may also be interpreted to reflect a mechanism to correct rent-seeking politicians for unanticipated fiscal policy changes (Allers and Elhorst 2005).

In both the spatial lag and the spatial error model, stationarity requires that $1/\omega_{min} < \delta < 1/\omega_{max}$ and $1/\omega_{min} < \rho < 1/\omega_{max}$, where ω_{min} and ω_{max} denote the smallest (i.e., most negative) and largest characteristic roots of the matrix $\boldsymbol{W}$. While it is often suggested in the literature to constrain δ or ρ to the interval (–1, +1), this may be unnecessarily restrictive. For row-normalized spatial weights, the largest characteristic root is indeed +1, but no general result holds for the smallest characteristic root, and the lower bound is typically less than –1. See also the lively dis-

[1] The regularity conditions $\boldsymbol{W}$ should satisfy in a cross-sectional setting have been derived by Lee (2004), but some of these regularity conditions may change in a panel data setting (Yu et al. 2007).

cussion at GeoDa's Openspace mailing list about the bounds on the spatial lag coefficient.

As an alternative to row-normalization, $\boldsymbol{W}$ might be normalized such that the elements of each column sum to one. This type of normalization is sometimes used in the social economics literature (Leenders 2002). Note that the row elements of a spatial weights matrix display the impact *on* a particular unit *by* all other units, while the column elements of a spatial weights matrix display the impact *of* a particular unit *on* all other units (see Chapter C.1 for a more detailed discussion of this issue). 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.

If $\boldsymbol{W_0}$ denotes the spatial weights matrix before normalization, one may also divide the elements of $\boldsymbol{W_0}$ by its largest characteristic root $\omega_{0,max}$ to get $\boldsymbol{W} = (1/\omega_{0,max})\boldsymbol{W_0}$, or 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 may be labeled matrix normalization, since it has the effect that the characteristic roots of $\boldsymbol{W_0}$ are also divided by $\omega_{0,max}$, as a result of which ω_{max}=1, just like the largest characteristic root of a row- or column-normalized matrix. 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 alternative normalizations. This is an important property when $\boldsymbol{W}$ represents an inverse distance matrix, since scaling the rows or columns of an inverse distance matrix so that the weights sum to one would cause this matrix to lose its economic interpretation for this decay (Anselin 1988, pp.23-24). One concomitant advantage of spatial weights matrices that do not lose their property of symmetry as a result of normalization is that notation, in some cases, is considerably simplified and that computation time will speed up (Elhorst 2001, 2005a).

Two main approaches have been suggested in the literature to estimate models that include spatial interaction effects. One is based on the maximum likelihood (ML) principle and the other on instrumental variables or generalized method of moments (IV/GMM) techniques. Although IV/GMM estimators are different from ML estimators in that they do not rely on the assumption of normality of the errors, both estimators assume that the disturbance terms ε_{it} are independently and identically distributed for all i and t with zero mean and variance σ^2. The Jarque-Bera (1980) test may be used to investigate the normality assumption when applying ML estimators.[2] One disadvantage of IV/GMM estimators is the possibility of

[2] This test has a chi-squared distribution with one degree of freedom. In addition, the Jarque-Bera test may be used to test for serial independence and homoskedasticity of the regression residuals. These tests have a chi-squared distribution with p degrees of freedom when testing for p-order serial autocorrelation, and q degrees of freedom when testing for homoskedasticity, one degree for every variable that might explain heteroskedasticity. Although informative, it should be noted that these tests were not developed in the context of a model with spatial interaction effects.

ending up with a coefficient estimate for δ or for ρ outside its parameter space $(1/\omega_{min},1/\omega_{max})$. Whereas this coefficient is restricted to its parameter space by the Jacobian term in the log-likelihood function of ML estimators, it is unrestricted using IV/GMM since these estimators ignore the Jacobian term.

Franzese and Hays (2007) compare the performance of the IV estimator and the ML estimator of panel data models with a spatially lagged dependent variable in terms of unbiasedness and efficiency, but unfortunately without considering spatial fixed or random effects. They find that the ML estimator offers weakly dominant efficiency and generally solid performance in unbiasedness, although it sometimes falls a little short of IV on unbiasedness grounds at lower values of δ.

The main focus in this chapter will be on ML estimation, because the number of studies considering IV/GMM estimators of spatial panel data models is still relatively sparse. One exception is Kelejian et al. (2006), who considered IV estimation of a spatial lag model with time period fixed effects. They point out that this model cannot be combined with a spatial weights matrix whose non-diagonal elements are all equal to $1/(N–1)$. In this situation, the spatially lagged dependent variable can be written in vector form as

$$\left\{\tfrac{1}{N-1}\sum_j y_{j1}-\tfrac{y_{11}}{N-1},\ldots,\tfrac{1}{N-1}\sum_j y_{j1}-\tfrac{y_{N1}}{N-1},\ldots,\tfrac{1}{N-1}\sum_j y_{jT}-\tfrac{y_{1T}}{N-1},\ldots,\tfrac{1}{N-1}\sum_j y_{jT}-\tfrac{y_{NT}}{N-1}\right\}^{\mathrm{T}} \tag{C.2.4}$$

which is asymptotically proportional and thus collinear with the time period fixed effects as N goes to infinity. Another exception is Kapoor et al. (2007), who considered the GMM estimator of a spatial error model and time period random effects. However, neither of these studies considered spatial fixed or random effects, while just these effects often appear to be important in panel data studies.

One shortcoming of the spatial lag model and the spatial error model is that spatial patterns in the data may be explained not only by either endogenous interaction effects or correlated error terms, but also by endogenous interaction effects, exogenous interaction effects and correlated error terms at the same time (Manski 1993). The best strategy would, therefore, seem to be to include the spatially lagged dependent variable, the K spatially lagged independent variables, and the spatially autocorrelated error term simultaneously.[3] However, Manski (1993) has also pointed out that at least one of these $2+K$ spatial interaction effects must be excluded, because otherwise their interaction parameters are not identified. In ad-

[3] In his keynote speech at the First World Conference of the Spatial Econometrics Association 2007, Harry Kelejian advocated models that include both a spatially lagged dependent variable and a spatially autocorrelated error term, while James LeSage in his Presidential Address at the 54th North American Meeting of the Regional Science Association International 2007 advocated models that include both a spatially lagged dependent variable and spatially lagged independent variables.

dition to this, the spatial weights matrix of the spatially lagged dependent variable must be different from the spatial weights matrix of the spatially autocorrelated error term, an additional requirement for identification when applying ML estimators (Anselin and Bera 1998). One ostensible advantage of IV/GMM estimators is that the same spatial weights matrix can be used to estimate a model extended to include a spatially lagged dependent variable and a spatially autocorrelated error term (Kelejian and Prucha 1998; Lee 2003). However, these estimators on their turn are unable to estimate models with spatially lagged independent variables, since they use these variables as instruments.

Alternatively, one may first test whether spatially lagged independent variables must be included and then whether the model should be extended to include a spatially lagged dependent variable or a spatially autocorrelated error term (Florax and Folmer 1992; Elhorst and Freret 2009) or adopt an unconstrained spatial Durbin model and then test whether this model can be simplified (Elhorst et al. 2006; Ertur and Koch 2007). An unconstrained spatial Durbin model with spatial fixed effects takes the form

$$y_{it} = \delta \sum_{j=1}^{N} W_{ij} y_{jt} + \boldsymbol{X}_{it} \boldsymbol{\beta} + \sum_{j=1}^{N} W_{ij} \boldsymbol{X}_{jt} \boldsymbol{\gamma} + \mu_i + \varepsilon_{it} \tag{C.2.5}$$

where $\boldsymbol{\gamma}$, just as $\boldsymbol{\beta}$, is a K-by-1 vector of fixed but unknown parameters. The hypothesis H_0: $\boldsymbol{\gamma} = 0$ can be tested to investigate whether this model can be simplified to the spatial lag model and the hypothesis H_0: $\boldsymbol{\gamma} + \delta\boldsymbol{\beta} = 0$ whether it can be simplified to the spatial error model. A simulation study by Florax et al. (2003) showed that the specific-to-general approach outperforms the general-to-specific approach when using cross-sectional data. However, one objection to this study is that the comparison between the two approaches is invalid because the null rejection frequencies have not been standardized (Hendry 2006). Another objection is that the model that has been used as point of departure did not include spatially lagged independent variables. Hence, a more careful elaboration of the relative merits of both approaches when using spatial panel data remains a topic of further research.

C.2.3 Estimation of panel data models

The spatial 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, while in the random effects model, μ_i is treated as a random variable that is independently and identically distributed with zero mean and variance σ_μ^2. Furthermore, it is assumed that the random variables μ_i and ε_{it} are independent of each other.

Throughout this chapter it is assumed that the data are sorted first by time and then by spatial units, whereas the classic panel data literature tends to sort the data first by spatial units and then by time. When y_{it} and $\boldsymbol{X}_{it}$ of these T successive cross-sections of N observations are stacked, we obtain an NT-by-1 vector for $\boldsymbol{y}$ and an NT-by-K matrix for $\boldsymbol{X}$.

Fixed effects model

If the spatial specific effects are treated as fixed effects, the model in Eq. (C.2.1) can be estimated in three steps. First, the spatial fixed effects μ_i are eliminated from the regression equation by demeaning the dependent and independent variables. This transformation takes the form

$$y_{it}^* = y_{it} - \tfrac{1}{T}\sum_{t=1}^{T} y_{it} \quad \text{and} \quad \boldsymbol{X}_{it}^* = \boldsymbol{X}_{it} - \tfrac{1}{T}\sum_{t=1}^{T} \boldsymbol{X}_{it}\,. \tag{C.2.6}$$

Second, the transformed regression equation $y_{it}^* = \boldsymbol{X}_{it}^* \boldsymbol{\beta} + \varepsilon_{it}^*$ is estimated by OLS: $\boldsymbol{\beta} = (\boldsymbol{X}^{*\mathrm{T}} \boldsymbol{X}^*)^{-1} \boldsymbol{X}^{*\mathrm{T}} \boldsymbol{y}^*$ and $\sigma^2 = (\boldsymbol{y}^* - \boldsymbol{X}^* \boldsymbol{\beta})^{\mathrm{T}} (\boldsymbol{y}^* - \boldsymbol{X}^* \boldsymbol{\beta}) / (NT-N-K)$. 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-by-K matrix rather than $(K+N)$-by-$(K+N)$ as in Eq. (C.2.1). 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

$$\ln L = -\tfrac{NT}{2}\ln(2\pi\sigma^2) - \tfrac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}^* - \boldsymbol{X}_{it}^* \boldsymbol{\beta})^2 \tag{C.2.7}$$

the ML estimators of $\boldsymbol{\beta}$ and σ^2 are $\boldsymbol{\beta} = (\boldsymbol{X}^{*\mathrm{T}} \boldsymbol{X}^*)^{-1} \boldsymbol{X}^{*\mathrm{T}} \boldsymbol{y}^*$ and $\sigma^2 = (\boldsymbol{y}^* - \boldsymbol{X}^* \boldsymbol{\beta})^{\mathrm{T}} (\boldsymbol{y}^* - \boldsymbol{X}^* \boldsymbol{\beta}) / NT$, 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)

$$AsyVar\,(\boldsymbol{\beta}, \sigma^2) = \begin{bmatrix} \dfrac{1}{\sigma^2}\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{X}^* & 0 \\ 0 & \dfrac{NT}{2\sigma^4} \end{bmatrix}^{-1}. \tag{C.2.8}$$

Finally, the spatial fixed effects may be recovered by

$$\mu_i = \tfrac{1}{T}\sum_{t=1}^{T}(y_{it} - \boldsymbol{X}_{it}\boldsymbol{\beta}) \qquad i = 1, \ldots, N. \tag{C.2.9}$$

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. Also note that 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, note 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)

$$AsyVar(\hat{\mu}_i) = \frac{\hat{\sigma}^2}{T} + \hat{\sigma}^2(\tfrac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{it})(\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{X}^{*})^{-1}(\tfrac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{it})^{\mathrm{T}}. \tag{C.2.10}$$

An alternative and equivalent formulation of Eq. (C.2.1) is to introduce a mean intercept α, provided that $\Sigma_i \mu_i = 0$. Then the spatial fixed effect μ_i represents the deviation of the ith spatial unit from the individual mean (see Hsaio 2003, p.33).

To test for spatial interaction effects in a cross-sectional setting, Anselin et al. (1996) developed Lagrange multiplier (LM) tests for a spatially lagged dependent variable, for spatial error correlation, and their counterparts robustified against the alternative of the other form.[4] These tests have become very popular in empirical research. Recently, Anselin et al. (2006) also specified the first two LM tests for a spatial panel

[4] 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 by Donald Lacombe at ≤http://oak.cats.ohiou.edu/~lacombe/research.html≥.

$$\mathrm{LM}_\delta = \frac{[\boldsymbol{e}^{\mathrm{T}}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{y}\,\hat{\sigma}^{-2}]^2}{J} \text{ and } \mathrm{LM}_p = \frac{[\boldsymbol{e}^{\mathrm{T}}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{e}\,\hat{\sigma}^{-2}]^2}{T\,T_W} \tag{C.2.11}$$

where the symbol ⊗ denotes the Kronecker product, $\boldsymbol{I}_T$ denotes the identity matrix and its subscript the order of this matrix, and $\boldsymbol{e}$ 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

$$J = \frac{1}{\hat{\sigma}^2}\left\{\left((\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}[\boldsymbol{I}_{NT} - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}](\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{X}\hat{\boldsymbol{\beta}} + TT_W\,\hat{\sigma}^2\right\} \tag{C.2.12}$$

$$T_W = trace(\boldsymbol{W}\boldsymbol{W} + \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}) \tag{C.2.13}$$

where *trace* denotes the trace of a matrix. In view of these formulas, the robust counterparts of these LM tests for a spatial panel will take the form

$$\mathrm{Robust\,LM}_\rho = \frac{[\boldsymbol{e}^{\mathrm{T}}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{y}\,\hat{\sigma}^{-2} - \boldsymbol{e}^{\mathrm{T}}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{e}\,\hat{\sigma}^{-2}]^2}{J - TT_W} \tag{C.2.14}$$

$$\mathrm{Robust\,LM}_\rho = \frac{[\boldsymbol{e}^{\mathrm{T}}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{e}\,\hat{\sigma}^{-2} - [TT_W/J]\,\boldsymbol{e}^{\mathrm{T}}(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{y}\,\hat{\sigma}^{-2}]^2}{TT_W\,[1 - TT_W/J]^{-1}}. \tag{C.2.15}$$

Note that the performance of these tests when having panel data instead of cross-sectional data and when having a model extended to include spatially lagged independent variables must still be investigated.

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. The labor force participation rate and its evolution over the business cycle is one of the best examples (Elhorst 2008a). 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

$$y_{it}-\sum_{j=1}^{N}W_{ij}\,y_{jt}=y_{it}-\tfrac{1}{N}\sum_{j=1}^{N}y_{jt} \text{ and } \boldsymbol{X}_{it}-\sum_{j=1}^{N}W_{ij}\,\boldsymbol{X}_{jt}=\boldsymbol{X}_{it}-\tfrac{1}{N}\sum_{j=1}^{N}\boldsymbol{X}_{jt} \qquad \text{(C.2.16)}$$

which is equivalent to the demeaning procedure of Eq. (C.2.6) 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.

Applied researchers also often find significant differences among the coefficient estimates from models with and without spatial fixed effects. These models are different in that they utilize different parts of the variation between observations. Models with controls for spatial fixed effects utilize the time-series component of the data, whereas models without controls for spatial fixed effects utilize the cross-sectional component of the data. As a result, some studies argue that models with controls for spatial fixed effects tend to give short-term estimates and models without controls for spatial fixed effects tend to give long-term estimates (Baltagi 2005, pp.200-201; Partridge 2005). A related problem of controlling for spatial fixed effects is that any variable that does not change over time or only varies a little cannot be estimated, because it is wiped out by the demeaning transformation. This is the main reason for many studies not controlling for spatial fixed effects.

On the other hand, 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 is 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 $-2s$, 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, the LR test can be carried out instead of, or in addition to, the classical F-test spelled out in Baltagi (2005, p.13). It is another advantage of estimating models by ML.

Random effects model

A compromise solution to the all or nothing way of utilizing the cross-sectional component of the data is the random effects model. This model avoids the loss of degrees of freedom incurred in the fixed effects model associated with a relatively large N and the problem that the coefficients of time-invariant variables cannot be estimated. However, whether the random effects model is an appropriate specification in spatial research remains controversial. When the random effects model is implemented, the units of observation should be representative of a larger population, and the number of units should potentially be able to go to infinity. 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, it is questionable whether they are 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), and '*the individual spatial units have characteristics that actually set them apart from a larger population*' (Anselin 1988, p.51). According to 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*'. In addition, the traditional assumption of zero correlation between μ_i in the random effects model and the explanatory variables, which also needs to be made, is particularly restrictive.

An iterative two-stage estimation procedure may be used to obtain the ML estimates of the random effects model (Breusch 1987). Note that the random effects model also includes a constant term, as a result of which the number of independent variables is K+1. The log-likelihood of the random effects model in Eq. (C.2.1) is

$$\ln L = -\tfrac{NT}{2}\ln(2\pi\sigma^2) + \tfrac{N}{2}\ln\theta^2 - \tfrac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}^{\bullet} - \boldsymbol{X}_{it}^{\bullet}\boldsymbol{\beta})^2 \tag{C.2.17}$$

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

$$y_{it}^{\bullet} = y_{it} - (1-\theta)\tfrac{1}{T}\sum_{t=1}^{T} y_{it} \quad \text{and} \quad \boldsymbol{X}_{it}^{\bullet} = \boldsymbol{X}_{it} - (1-\theta)\tfrac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{it}\,. \tag{C.2.18}$$

If $\theta = 0$, this transformation simplifies to the demeaning procedure of Eq. (C.2.6) and hence the random effects model to the fixed effects model.

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

$$\ln L = -\tfrac{NT}{2}\ln\left\{\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(y_{it} - (1-\theta)\tfrac{1}{T}\sum_{t'=1}^{T} y_{it'}\right] - \left[\boldsymbol{X}_{it} - (1-\theta)\tfrac{1}{T}\sum_{t'=1}^{T}\boldsymbol{X}_{it'}\right]\boldsymbol{\beta})^2\right\} + \tfrac{N}{2}\ln\theta^2. \tag{C.2.19}$$

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

$$AsyVar(\boldsymbol{\beta}, \theta, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{\bullet \mathrm{T}}\boldsymbol{X}^{\bullet} & 0 & 0 \\ 0 & N\left(1+\frac{1}{\theta^2}\right) & -\frac{N}{\sigma^2} \\ 0 & -\frac{N}{\sigma^2} & \frac{NT}{2\sigma^4} \end{bmatrix}^{-1}. \tag{C.2.20}$$

One can test whether the spatial random effects are significant by performing a LR test of the hypothesis H_0: $\theta = 1$.[5] This test statistic has a chi-squared distribution with one degree of freedom. If the hypothesis is rejected, the spatial random effects are significant.

C.2.4 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. It is assumed that $\boldsymbol{W}$ is constant over time and that the panel is balanced. Although the estimators can be modified for a spatial weights matrix that changes over time, as well as for an unbalanced panel, their asymptotic properties, in the event of an unbalanced panel, may become problematic if the reason why data are missing is not known.

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 $E\,[(\Sigma_j W_{ij} y_{jt})\, \varepsilon_{it}] = 0$. 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, we derive the ML estimator to account for the endogeneity of $\Sigma_j W_{ij} y_{jt}$. The log-likelihood function of the model in Eq. (C.2.2) if the spatial specific effects are assumed to be fixed is

$$\ln L = -\tfrac{NT}{2}\ln(2\pi\sigma^2) + T\ln|I_n - \delta \boldsymbol{W}| - \tfrac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it} - \delta\sum_{j=1}^{N} W_{ij} y_{jt} - \boldsymbol{X}_{it}\boldsymbol{\beta} - \mu_i \right]^2 \tag{C.2.21}$$

where the second term on the right-hand side represents the Jacobian term of the transformation from $\boldsymbol{\varepsilon}$ to $\boldsymbol{y}$ 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

[5] $\theta = 1$ implies $\sigma_\mu^2 = 0$, since σ_μ^2 may be calculated from θ by $[(1-\theta^2)/\theta^2]\,[\sigma^2/T]$.

$$\frac{\partial \ln L}{\partial \mu_i} = \frac{1}{\sigma^2} \sum_{t=1}^{T} \left[y_{it} - \delta \sum_{j=1}^{N} W_{ij} y_{jt} - \boldsymbol{X}_{it} \boldsymbol{\beta} - \mu_i \right] = 0 \qquad i = 1, \ldots, N. \tag{C.2.22}$$

When solving μ_i from Eq. (C.2.22), one obtains

$$\mu_i = \frac{1}{T} \sum_{t=1}^{T} \left[y_{it} - \delta \sum_{j=1}^{N} W_{ij} y_{jt} - \boldsymbol{X}_{it} \boldsymbol{\beta}) \right] \qquad i = 1, \ldots, N. \tag{C.2.23}$$

This equation shows that the standard formula for calculating the spatial fixed effects, Eq. (C.2.9), applies to the fixed effects spatial lag model in a straightforward manner. Corrections for the spatial dependence among the observations at each point in time, other than the addition of the spatially lagged dependent variable to these formulas, are not necessary.[6]

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

$$\ln L = -\frac{NT}{2} \ln(2\pi\sigma^2) + T \ln|I_N - \delta \boldsymbol{W}| - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[y_{it}^* - \delta \left(\sum_{j=1}^{N} W_{ij} y_{jt} \right)^* - \boldsymbol{X}_{it}^* \boldsymbol{\beta} \right]^2 \tag{C.2.24}$$

where the asterisk denotes the demeaning procedure introduced in Eq. (C.2.6).

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 log-likelihood function in Eq. (C.2.24) 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 NT 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-by-1 vectors for $\boldsymbol{y}^*$ and $(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{y}^*$, and an NT-by-K matrix for $\boldsymbol{X}^*$ of the demeaned variables. Note that these calculations have to be performed only once and that the NT-by-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 $\boldsymbol{y}^*$

[6] Anselin et al. (2006) asked for a more careful elaboration of this.

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

$$\ln L = C - \frac{NT}{2}\ln\left[(\boldsymbol{e}_0^* - \delta\boldsymbol{e}_1^*)^{\mathrm{T}}(\boldsymbol{e}_0^* - \delta\boldsymbol{e}_1^*)\right] + T\ln|\boldsymbol{I}_N - \delta\boldsymbol{W}| \qquad \text{(C.2.25)}$$

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). To speed up computation time and to overcome numerical difficulties one might face in evaluating $\ln|\boldsymbol{I}_N - \delta\boldsymbol{W}|$, Pace and Barry (1997) propose to compute this determinant once over a grid of values for the parameter δ ranging from $1/\omega_{min}$ to one prior to estimation, provided that $\boldsymbol{W}$ is normalized. This only requires the determination of the smallest 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 log determinant of $(\boldsymbol{I}_N - \delta\boldsymbol{W})$, they point out that one can quickly evaluate the concentrated log-likelihood function for all values of δ in the grid and determine the optimal value of δ as that which maximizes the concentrated log-likelihood function over this grid.[7]

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

$$\boldsymbol{\beta} = \boldsymbol{b}_0 - \delta\boldsymbol{b}_1 = (\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\mathrm{T}}[\boldsymbol{y}^* - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W})\boldsymbol{y}^*] \qquad \text{(C.2.26a)}$$

$$\sigma^2 = \frac{1}{NT}(\boldsymbol{e}_0^* - \delta\boldsymbol{e}_1^*)^{\mathrm{T}}(\boldsymbol{e}_0^* - \delta\boldsymbol{e}_1^*)\,. \qquad \text{(C.2.26b)}$$

Instead of the demeaned variables, one may also use the original variables $\boldsymbol{y}$ and $\boldsymbol{X}$, since $\boldsymbol{y}^* = \boldsymbol{Q}\boldsymbol{y}$, $(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{y}^* = \boldsymbol{Q}\,(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{y}$, and $\boldsymbol{X}^* = \boldsymbol{Q}\boldsymbol{X}$, where $\boldsymbol{Q}$ denotes the demeaning operator in matrix form

$$\boldsymbol{Q} = \boldsymbol{I}_{NT} - \frac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^{\mathrm{T}} \otimes \boldsymbol{I}_N \qquad \text{(C.2.27)}$$

[7] The computation of the log determinant may be carried out using the Matlab routine ‘lndet’ from LeSage's website <www.spatial-econometrics.com> (LeSage 1999).

and $\boldsymbol{\iota}_T$ is a vector of ones whose subscript denotes the length of this vector. Since $\boldsymbol{Q}$ is a symmetric idempotent matrix, the estimator of $\boldsymbol{\beta}$ starting with the original variables may also be written as

$$\boldsymbol{\beta} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\,\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\,[\boldsymbol{y} - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{y}] = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\,\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Q}\,[\boldsymbol{y} - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{y}]. \tag{C.2.28}$$

Anselin et al. (2006) have pointed out that this estimator may also be seen as the GLS estimator of a linear regression model with disturbance covariance matrix $\sigma^2\boldsymbol{Q}$, but the difficulty of this interpretation is that $\boldsymbol{Q}$ is singular. Their conclusion that the singularity of $\boldsymbol{Q}$ also limits the practicality of this model has been contradicted by Hsaio (2003, p.320), Magnus and Neudecker (1988, pp.271-273) and Baltagi (1989) in that $\boldsymbol{Q}$ may be replaced by its general inverse,[8] which again produces (C.2.28).

Finally, the asymptotic variance matrix of the parameters is computed for inference (standard errors, t-values). This matrix has been derived by Elhorst and Freret (2009) and takes the form (since this matrix is symmetric the upper diagonal elements are left aside)

$$AsyVar(\boldsymbol{\beta}, \delta, \sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{X}^* & & \\ \frac{1}{\sigma^2}\boldsymbol{X}^{*\mathrm{T}}(\boldsymbol{I}_T \otimes \tilde{\boldsymbol{W}})\boldsymbol{X}^*\boldsymbol{\beta} & T\,trace(\tilde{\boldsymbol{W}}\tilde{\boldsymbol{W}} + \tilde{\boldsymbol{W}}^{\mathrm{T}}\tilde{\boldsymbol{W}}) + \frac{1}{\sigma^2}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}^{*\mathrm{T}}(\boldsymbol{I}_T \otimes \tilde{\boldsymbol{W}}^{\mathrm{T}}\tilde{\boldsymbol{W}})\boldsymbol{X}^*\boldsymbol{\beta} & \\ 0 & \frac{T}{\sigma^2}trace(\tilde{\boldsymbol{W}}) & \frac{NT}{2\sigma^4} \end{bmatrix}^{-1} \tag{C.2.29}$$

where $\tilde{\boldsymbol{W}} = \boldsymbol{W}(I_N - \delta\boldsymbol{W})^{-1}$. 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 NT observations and the summation over T cross-sections involving manipulations of the N-by-N spa-

[8] $\boldsymbol{Q}^+$ is called the generalized (Moore-Penrose) inverse of $\boldsymbol{Q}$ if it satisfies the conditions: $\boldsymbol{Q}\,\boldsymbol{Q}^+\boldsymbol{Q} = \boldsymbol{Q}$, $\boldsymbol{Q}^+\boldsymbol{Q}\,\boldsymbol{Q}^+ = \boldsymbol{Q}^+$, $(\boldsymbol{Q}^+\boldsymbol{Q})^{\mathrm{T}} = \boldsymbol{Q}^+\boldsymbol{Q}$ and $(\boldsymbol{Q}\,\boldsymbol{Q}^+)^{\mathrm{T}} = \boldsymbol{Q}\,\boldsymbol{Q}^+$ (Magnus and Neudecker 1988, p.32).

tial 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.

Fixed effects spatial error model

Anselin and Hudak (1992) have also spelled out how the parameters $\boldsymbol{\beta}$, ρ 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 NT observations. The log-likelihood function of model in Eq. (C.2.3) if the spatial specific effects are assumed to be fixed is

$$\ln L = -\tfrac{NT}{2}\ln(2\pi\sigma^2) + T\ln|\boldsymbol{I}_N - \rho\boldsymbol{W}|$$

$$-\tfrac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\boldsymbol{y}_{it}^* - \rho\left[\sum_{j=1}^{N} W_{ij}\boldsymbol{y}_{jt}\right]^* - \left[\boldsymbol{X}_{it}^* - \rho\left(\sum_{j=1}^{N} W_{ij}\boldsymbol{X}_{jt}\right)^*\right]\boldsymbol{\beta}\right\}^2. \tag{C.2.30}$$

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

$$\boldsymbol{\beta} = \{[\boldsymbol{X}^* - \rho\,(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{X}^*]^{\mathrm{T}}\,[\boldsymbol{X}^* - \rho\,(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{X}^*]\}^{-1}$$

$$[\boldsymbol{X}^* - \rho\,(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{X}^*]^{\mathrm{T}}\,[\boldsymbol{y}^* - \rho\,(\boldsymbol{I}_T \otimes \boldsymbol{W})\,\boldsymbol{y}^*] \tag{C.2.31a}$$

$$\sigma^2 = \frac{\boldsymbol{e}(\rho)^{\mathrm{T}}\,\boldsymbol{e}(\rho)}{NT} \tag{C.2.31b}$$

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

$$\ln L = -\tfrac{NT}{2}\ln[\boldsymbol{e}(\rho)^{\mathrm{T}}\boldsymbol{e}(\rho)] + T\ln|\boldsymbol{I}_N - \rho\boldsymbol{W}|. \tag{C.2.32}$$

Maximizing this function with respect to ρ yields the ML estimator of ρ, 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 ρ are alternately estimated until convergence occurs. The asymptotic variance matrix of the parameters takes the form

$$AsyVar(\boldsymbol{\beta},\rho,\sigma^2) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{X} & & \\ 0 & T\ trace(\tilde{\tilde{\boldsymbol{W}}}\tilde{\tilde{\boldsymbol{W}}}+\tilde{\tilde{\boldsymbol{W}}}^{\mathrm{T}}\tilde{\tilde{\boldsymbol{W}}}) & \\ 0 & \frac{T}{\sigma^2}trace(\tilde{\tilde{\boldsymbol{W}}}) & \frac{NT}{2\sigma^4} \end{bmatrix}^{-1} \quad \text{(C.2.33)}$$

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

$$\mu_i = \tfrac{1}{T}\sum_{t=1}^{T}(y_{it} - \boldsymbol{X}_{it}\boldsymbol{\beta}) \qquad i=1,...,N. \quad \text{(C.2.34)}$$

Random effects spatial lag model

The log-likelihood of model in Eq. (C.2.2) if the spatial effects are assumed to be random is

$$\ln L = -\tfrac{NT}{2}\ln(2\pi\sigma^2) + T\ln|\boldsymbol{I}_N - \delta\boldsymbol{W}| - \tfrac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it}^{\bullet} - \delta\left(\sum_{j=1}^{N}W_{ij}y_{jt}\right)^{\bullet} - \boldsymbol{X}_{it}^{\bullet}\boldsymbol{\beta}\right]^2 \quad \text{(C.2.35)}$$

where the symbol $^\bullet$ denotes the transformation introduced in Eq. (C.2.18) dependent on θ. Given θ, this log-likelihood function is identical to the log-likelihood function of the fixed effects spatial lag model in Eq. (C.2.24). This implies that the same procedure can be used to estimate $\boldsymbol{\beta}$, δ and σ^2 as described above [Eqs. (C.2.25), (C.2.26a) and (C.2.26b)], but that the superscript * must be replaced by $^\bullet$. Given $\boldsymbol{\beta}$, δ and σ^2, θ can be estimated by maximizing the concentrated log-likelihood function with respect to θ

$$\ln L = -\frac{NT}{2}\ln[\boldsymbol{e}(\theta)^T \boldsymbol{e}(\theta)] + \frac{N}{2}\ln\theta^2 \tag{C.2.36}$$

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

$$\begin{aligned}\boldsymbol{e}(\theta)_{it} = y_{it} - (1-\theta)\frac{1}{T}\sum_{t=1}^{T} y_{it} - \delta\left[\sum_{j=1}^{N} W_{ij}\, y_{jt} - (1-\theta)\frac{1}{T}\sum_{t=1}^{T} W_{ij}\, y_{jt}\right] \\ - \left[\boldsymbol{X}_{it} - (1-\theta)\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{it}\right]\boldsymbol{\beta}.\end{aligned} \tag{C.2.37}$$

Again an iterative procedure may be used where the set of parameters $\boldsymbol{\beta}$, δ and σ^2 and the parameter θ 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

$$AsyVar(\boldsymbol{\beta}, \delta, \theta, \sigma^2) =$$

$$\begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{\bullet T}\boldsymbol{X}^{\bullet} & & & \\ \frac{1}{\sigma^2}\boldsymbol{X}^{\bullet T}(\boldsymbol{I}_T \otimes \widetilde{\boldsymbol{W}})\boldsymbol{X}^{\bullet}\boldsymbol{\beta} & T\, trace(\widetilde{\boldsymbol{W}}^T\boldsymbol{W} + \widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}}) + \frac{1}{\sigma^2}\boldsymbol{\beta}^T\boldsymbol{X}^{\bullet T}(\boldsymbol{I}_T \otimes \widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}})\boldsymbol{X}^{\bullet}\boldsymbol{\beta} & & \\ 0 & -\frac{1}{\sigma^2}\, trace(\widetilde{\boldsymbol{W}}) & N(1+\frac{1}{\theta^2}) & \\ 0 & \frac{T}{\sigma^2}\, trace(\widetilde{\boldsymbol{W}}) & -\frac{N}{\sigma^2} & \frac{NT}{2\sigma^4} \end{bmatrix}^{-1}$$

(C.2.38)

Random effects spatial error model

The log-likelihood of model in Eq. (C.2.3) if the spatial effects are assumed to be random is (Anselin 1988; Elhorst 2003; Baltagi 2005)

$$\ln L = -\tfrac{NT}{2}\ln(2\pi\sigma^2) - \tfrac{1}{2}\ln|\boldsymbol{V}| + (T-1)\sum_{i=1}^{N}\ln|\boldsymbol{B}|$$

$$-\tfrac{1}{2\sigma^2}\boldsymbol{e}^{\mathrm{T}}(\tfrac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^{\mathrm{T}} \otimes \boldsymbol{V}^{-1})\boldsymbol{e} - \tfrac{1}{2\sigma^2}\boldsymbol{e}^{\mathrm{T}}(\boldsymbol{I}_T - \tfrac{1}{T}\boldsymbol{\iota}_T\boldsymbol{\iota}_T^{\mathrm{T}})\otimes(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})\boldsymbol{e} \qquad \text{(C.2.39)}$$

where $\boldsymbol{V}=T\,\varphi\,\boldsymbol{I}_N + (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}$ with $\varphi = \sigma_\mu^2/\sigma^2$,[9] $\boldsymbol{B} = \boldsymbol{I}_N - \rho\,\boldsymbol{W}$ and $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}$. It is the matrix $\boldsymbol{V}$ that complicates the estimation of this model considerably. First, the Pace and Barry (1997) procedure to overcome numerical difficulties one might face in evaluating $\ln|\boldsymbol{B}| = \ln|\boldsymbol{I}_N - \rho\,\boldsymbol{W}|$ cannot be used to calculate $\ln|\boldsymbol{V}| = \ln|T\,\varphi\,\boldsymbol{I}_N + (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}|$. Second, there is no simple mathematical expression for the inverse of $\boldsymbol{V}$. Baltagi (2006) solves these problems by considering a random effects spatial error model with equal weights, i.e., a spatial weights matrix $\boldsymbol{W}$ whose non-diagonal elements are all equal to $1/(N-1)$. Due to this setup, the inverse of $\boldsymbol{V}$ and a feasible GLS estimator of $\boldsymbol{\beta}$ can be determined mathematically. Furthermore, by considering a GLS estimator the term $\ln|\boldsymbol{V}|$ in the log-likelihood function does not have to be calculated.

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

$$\ln|\boldsymbol{V}| = \ln|T\varphi\boldsymbol{I}_N + (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}| = \sum_{i=1}^{N}\ln\left[T\varphi + \frac{1}{(1-\rho\omega_i)^2}\right]. \qquad \text{(C.2.40)}$$

Furthermore, he suggests to adopt the transformation

$$y_{it}^{o} = y_{it} - \rho\sum_{j=1}^{N}\boldsymbol{W}_{ij}y_{jt} + \sum_{j=1}^{N}\left\{\left[p_{ij} - (1-\rho\boldsymbol{W}_{ij})\right]\frac{1}{T}\sum_{t=1}^{T}y_{jt}\right\} \qquad \text{(C.2.41)}$$

and the same for the variables $\boldsymbol{X}_{it}$, where p_{ij} is an element of an N-by-N matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{V}^{-1}$. $\boldsymbol{P}$ can be the spectral decomposition of $\boldsymbol{V}^{-1}$, $\boldsymbol{P} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{R}$, where $\boldsymbol{R}$ is an N-by-N matrix of which the ith column is the characteristic vector $\boldsymbol{r}_i$ of $\boldsymbol{V}$, which is the same as the characteristic vector of the spatial weights matrix $\boldsymbol{W}$ (see Griffith 1988, Table 3.1), $\boldsymbol{R} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$, and $\boldsymbol{\Lambda}$ an N-by-N diagonal matrix with the ith diagonal element the corresponding characteristic root, $c_i = T\varphi + (1-\rho\omega_i)^{-2}$.

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

A similar procedure has been adopted by Yang et al. (2006). It is clear that for large N the numerical determination of $\boldsymbol{P}$ can be problematic. However, Hunneman et al. (2007) find that if $\boldsymbol{W}$ is kept symmetric by using one of the alternative normalizations discussed in Section C.2.2, this procedure works well within a reasonable amount of time for values of N up to 4,000.

As a result of Eqs. (C.2.40) and (C.2.41), the log-likelihood function simplifies to

$$\ln L = -\tfrac{NT}{2}\ln(2\pi\sigma^2) - \tfrac{1}{2}\sum_{i=1}^{N}\ln\left[1+T\phi(1-\rho\omega_i)^2\right] + T\sum_{i=1}^{N}\ln(1-\rho\omega_i) - \tfrac{1}{2\sigma_2}\boldsymbol{e}^{\mathrm{oT}}\boldsymbol{e}^{\mathrm{o}} \tag{C.2.42}$$

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

$$\ln L = C - \tfrac{NT}{2}\ln\left[\boldsymbol{e}(\rho,\phi)^{\mathrm{T}}e(\rho,\phi)\right] - \tfrac{1}{2}\sum_{i=1}^{N}\ln\left[1+T\phi(1-\rho\omega_i)^2\right] + T\sum_{i=1}^{N}\ln(1-\rho\omega_i) \tag{C.2.43}$$

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

$$\begin{aligned}\boldsymbol{e}(\rho,\theta)_{it} &= y_{it} - \rho\sum_{j=1}^{N}W_{ij}y_{jt} + \sum_{j=1}^{N}\left\{\left[p(\rho,\varphi)_{ij} - (1-\rho W_{ij})\right]\frac{1}{T}\sum_{t=1}^{T}y_{jt}\right\} \\ &- \left\{\boldsymbol{X}_{it} - \rho\sum_{j=1}^{N}W_{ij}\boldsymbol{X}_{jt} + \sum_{j=1}^{N}\left\{\left[p(\rho,\varphi)_{ij} - (1-\rho W_{ij})\right]\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{jt}\right\}\right\}\boldsymbol{\beta}.\end{aligned} \tag{C.2.44}$$

The notation $p_{ij} = p(\rho,\varphi)_{ij}$ is used to indicate that the elements of the matrix $\boldsymbol{P}$ depend on ρ and φ. One can iterate between $\boldsymbol{\beta}$ and σ^2 on the one hand, and ρ and φ on the other, until convergence. The estimators of $\boldsymbol{\beta}$ and σ^2, given ρ and φ, can be obtained by OLS regression of the transformed variable $\boldsymbol{y}^{\mathrm{o}}$ on the transformed variables $\boldsymbol{X}^{\mathrm{o}}$. However, the estimators of ρ 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). 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 (2008b) 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 $\boldsymbol{\beta}$, ρ, φ and σ^2, but of $\boldsymbol{\beta}$, ρ, σ_μ^2 and σ^2. This matrix takes the following form[10]

$$AsyVar(\boldsymbol{\beta}, \rho, \sigma_\mu^2, \sigma^2) =$$

$$\begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}^{\mathrm{oT}}\boldsymbol{X}^{\mathrm{o}} & & & \\ 0 & \frac{T-1}{2}trace(\boldsymbol{\Gamma})^2+\frac{1}{2}trace(\boldsymbol{\Sigma\Gamma})^2 & & \\ 0 & \frac{T}{2\sigma^2}trace(\boldsymbol{\Sigma\Gamma V}^{-1}) & \frac{T^2}{2\sigma 4}trace(\boldsymbol{V}^{-1})^2 & \\ 0 & \frac{T-1}{2\sigma^2}trace(\boldsymbol{\Gamma})+\frac{1}{2\sigma^2}trace(\boldsymbol{\Sigma\Gamma\Sigma}) & \frac{T}{2\sigma^4}trace(\boldsymbol{\Sigma V}^{-1}) & \frac{1}{2\sigma^4}[(T-1)N+trace(\boldsymbol{\Sigma})^2] \end{bmatrix}^{-1}$$

(C.2.45)

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

$$var(\varphi)=\varphi^2\left[\frac{var(\sigma_\mu^2)}{(\varphi\sigma^2)^2}+\frac{var(\sigma^2)}{(\sigma^2)^2}-2\frac{var(\sigma_\mu^2,\sigma^2)}{(\varphi\sigma^2)\sigma^2}\right]. \quad \text{(C.2.46)}$$

[10] Note that the matrix $\boldsymbol{Z}_0$ in Baltagi et al. (2007, pp.39-40) has been replaced by $\boldsymbol{Z}_0=[T\sigma_\mu^2\boldsymbol{I}_N+\sigma^2(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}]^{-1}=\frac{1}{\sigma^2}[T\varphi\boldsymbol{I}_N+(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}]^{-1}=\frac{1}{\sigma^2}\boldsymbol{V}^{-1}$.

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. Since a spatial error specification also does not require a theoretical model for a spatial or social interaction process, but is a special case of a non-spherical error covariance matrix, and the random effects models in spatial research is controversial, the random effects spatial error model will probably be of limited value in empirical research.

C.2.5 Model comparison and prediction

This section sets forth Hausman's specification test for statistically significant differences between random effects models and fixed effects models, two goodness-of-fit measures, one that includes the impact of spatial fixed or random effects and the impact of a spatial lag and one that does not, and the best linear unbiased predictor of the different models.

Random effects versus fixed effects

The random effects model can be tested against the fixed effects model using Hausman's specification test (Baltagi 2005, pp.66-68). The hypothesis being tested is H_0: $h = 0$, where

$$h = \boldsymbol{d}^{\mathrm{T}}[var(\boldsymbol{d})]^{-1}\boldsymbol{d}\,,\ \boldsymbol{d} = \hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE} \text{ and } var(\boldsymbol{d}) = \hat{\sigma}^2_{RE}(\boldsymbol{X}^{\bullet\mathrm{T}}\boldsymbol{X}^{\bullet})^{-1} - \hat{\sigma}^2_{FE}(\boldsymbol{X}^{*\mathrm{T}}\boldsymbol{X}^{*})^{-1}. \tag{C.2.47}$$

Note the reversed 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, one might calculate $\boldsymbol{d}$ by $\boldsymbol{d}$ = $[\hat{\boldsymbol{\beta}}^{\mathrm{T}}\ \hat{\delta}]^{\mathrm{T}}_{FE} - [\hat{\boldsymbol{\beta}}^{\mathrm{T}}\ \hat{\delta}]^{\mathrm{T}}_{RE}$ to obtain a test statistic that 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 Eqs. (C.2.29) and (C.2.38). If the hypothesis is rejected, the random effects models must be rejected in favor of the fixed effects model.

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 people use

$$R^2(\boldsymbol{e},\boldsymbol{\Omega}) = 1 - \frac{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\Omega}\,\boldsymbol{e}}{(\boldsymbol{y}-\overline{\boldsymbol{y}})^{\mathrm{T}}(\boldsymbol{y}-\overline{\boldsymbol{y}})} \quad \text{or} \quad R^2(\tilde{\boldsymbol{e}}) = 1 - \frac{\tilde{\boldsymbol{e}}^{\mathrm{T}}\tilde{\boldsymbol{e}}}{(\boldsymbol{y}-\overline{\boldsymbol{y}})^{\mathrm{T}}(\boldsymbol{y}-\overline{\boldsymbol{y}})} \tag{C.2.48}$$

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

One objection to the measures in Eq. (C.2.48) 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 ρ, 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, even though it may be seen as a linear regression model with disturbance covariance matrix $\sigma^2\boldsymbol{Q}$. 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)

$$corr^2(\boldsymbol{y},\hat{\boldsymbol{y}}) = \frac{[(\boldsymbol{y}-\overline{\boldsymbol{y}})^{\mathrm{T}}(\hat{\boldsymbol{y}}-\overline{\boldsymbol{y}})]^2}{[(\boldsymbol{y}-\overline{\boldsymbol{y}})^{\mathrm{T}}(\boldsymbol{y}-\overline{\boldsymbol{y}})][(\hat{\boldsymbol{y}}-\overline{\boldsymbol{y}})^{\mathrm{T}}(\hat{\boldsymbol{y}}-\overline{\boldsymbol{y}})]} \tag{C.2.49}$$

where $\hat{\boldsymbol{y}}$ is an NT-by-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.[11] In vector notation, the reduced form of the spatial lag model in Eq. (C.2.2) is

$$\boldsymbol{y} = [\boldsymbol{I}_{NT} - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W})]^{-1}[\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{\tau}_T \otimes \boldsymbol{I}_N)\boldsymbol{\mu} + \boldsymbol{\varepsilon}] \tag{C.2.50}$$

where $\boldsymbol{\mu}$ is an N-by-1 vector of the spatial specific effects, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^{\mathrm{T}}$. 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 $[\boldsymbol{I}_{NT} - \delta(\boldsymbol{I}_T \otimes \boldsymbol{W})]^{-1}$.

Table C.2.1. Two goodness-of-fit measures of the four spatial panel data models

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

Notes: $R^2(\boldsymbol{e}, \boldsymbol{I}_N)$ and $R^2(\tilde{\boldsymbol{e}})$ are defined by Eq. (C.2.48), $corr^2$ is defined by Eq. (C.2.49)

[11] See the routine ‘sar’ posted at LeSage's website <www.spatial-econometrics.com>

The two measures for the different spatial panel data models are listed in Table C.2.1. 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.

Prediction

Finally, prediction formulas are presented for fixed effects and random effects models with spatial interaction effects. Goldberger (1962) shows that the best linear unbiased predictor (BLUP) for the cross-sectional units in a linear regression model with disturbance covariance matrix $\boldsymbol{\Omega}$ at a future period $T+C$ is given by

$$\hat{\boldsymbol{y}}_{T+C} = \boldsymbol{X}_{T+C}\,\hat{\boldsymbol{\beta}} + \psi^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{e} \tag{C.2.51}$$

where $\psi = E\,(\boldsymbol{\varepsilon}_{T+C}\,\boldsymbol{\varepsilon})$ is the covariance between the future disturbance $\boldsymbol{\varepsilon}_{T+C}$ and the sample disturbances $\boldsymbol{\varepsilon}$, $\boldsymbol{X}$ covers the independent variables of the model, $\hat{\boldsymbol{\beta}}$ is the estimator of $\boldsymbol{\beta}$, and $\boldsymbol{e}$ denotes the residual vector of the model. Baltagi and Li (2004) derive the prediction formulas for the fixed effects and random effects model with spatial autocorrelation. Here, we also present these formulas for the fixed effects and random effects model extended to include a spatially lagged dependent variable based on own derivations. The prediction formulas are listed in Table C.2.2.

Baltagi and Li (2004) point out that $\psi = 0$ in the fixed effects model, provided that error terms are not serially correlated over time. Unlike the fixed effects model, the correction term $\psi^{\mathrm{T}}\boldsymbol{\Omega e}$ in the random effects model is not zero. In the random effects spatial lag model, the correction term $\psi^{\mathrm{T}}\boldsymbol{\Omega}\,\boldsymbol{e}$ is identically equal to its counterpart in a standard random effects model, which has been reported in Baltagi and Li (2004). To calculate this correction term (see Table C.2.2), the residuals of each spatial unit are first averaged over the sample period and then multiplied with $(1-\theta^2)$, a factor that can take values between zero and one.[12] However, in addition to the standard random effects model, both $\boldsymbol{X}_{T+C}\,\hat{\boldsymbol{\beta}}$ and the correction term should also be premultiplied with the N-by-N spatial multiplier matrix $(\boldsymbol{I}_{NT} - \delta\boldsymbol{W})^{-1}$.

[12] Note that $(1-\theta^2) = T\sigma_\mu^2/\,(T\sigma_\mu^2 + \sigma^2)$ (see Baltagi 2005, p. 20, for the second part of this formula).

Table C.2.2. Prediction formula of the four spatial panel data models

Fixed effects spatial lag model
$\hat{\boldsymbol{y}}_{T+C} = (\boldsymbol{I}_N - \hat{\delta}\boldsymbol{W})^{-1}\boldsymbol{X}_{T+C}\hat{\boldsymbol{\beta}} + (\boldsymbol{I}_N - \hat{\delta}\boldsymbol{W})^{-1}\hat{\boldsymbol{u}}$
Fixed effects spatial error model
$\hat{\boldsymbol{y}}_{T+C} = \boldsymbol{X}_{T+C}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}}$
Random effects spatial lag model
$\hat{\boldsymbol{y}}_{T+C} = (\boldsymbol{I}_N - \hat{\delta}\boldsymbol{W})^{-1}\boldsymbol{X}_{T+C}\hat{\boldsymbol{\beta}} + (\boldsymbol{I}_N - \hat{\delta}\boldsymbol{W})^{-1}(1-\hat{\theta}^2)\left\{\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} y_{1t} - \boldsymbol{X}_{1t}\hat{\boldsymbol{\beta}} \\ \vdots \\ y_{Nt} - \boldsymbol{X}_{Nt}\hat{\boldsymbol{\beta}} \end{pmatrix}\right\}$
Random effects spatial error model
$\hat{\boldsymbol{y}}_{T+C} = \boldsymbol{X}_{T+C}\hat{\boldsymbol{\beta}} + \hat{\varphi}\hat{\boldsymbol{V}}^{-1}\sum_{t=1}^{T}\begin{pmatrix} y_{1t} - \boldsymbol{X}_{1t}\hat{\boldsymbol{\beta}} \\ \vdots \\ y_{Nt} - \boldsymbol{X}_{Nt}\hat{\boldsymbol{\beta}} \end{pmatrix}$

Just as in the random effects spatial lag model, the residuals in the random effects spatial error model are first averaged over the sample period (see Table C.2.2). However, the sum of the residuals is not just divided by T, but premultiplied by $\boldsymbol{V}^{-1}=[T\varphi\boldsymbol{I}_N + (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1}]^{-1}$, a matrix that also accounts for the interaction effects among the residuals. Finally, the 'average' residuals are multiplied by φ, which measures the ratio between σ_{μ}^2 and σ^2.

One problem of predictors based on fixed or random effects models is that one has no information on the spatial fixed effects or the averaged residuals of spatial units outside the sample. For this reason, some researchers abandon fixed or random effects models. However, they better stick to the fixed effects or random effects models, provided that these effects appear to be (jointly) significant, and set the spatial fixed effects or the averaged residuals of spatial units outside the sampling region to zero or, alternatively, try to approach them from proximate spatial units within the sample region.

C.2.6 Concluding remarks

The spatial econometrics literature has exhibited a growing interest in the specification and estimation of econometric relationships based on spatial panels. Many empirical studies have found their way to the Matlab routines of the fixed effects and random effects models the author of this chapter has provided at his website.

Updated versions have been made available and include the (robust) LM tests, the estimation of fixed effects and the determination of their significance level, the determination of the variance-covariance matrix of the parameters estimates, the determination of good-of-fit measures, Hausman's specification test and the formulas for the best linear unbiased predictor, as discussed in this chapter.

Two other areas where more insight has been gained into the extension of spatial panel data models with spatial interaction effects is the possibility to test for endogeneity of one or more of the explanatory variables and the possibility to include dynamic effects. However, this literature has not yet been crystallized.

Fingleton and LeGallo (2007) consider models including an endogenous spatial lag, additional endogenous variables due to a system feedback and an autoregressive or a moving average error process, and suggest an IV/GMM estimator based on Kelejian and Prucha (1998) and Fingleton (2008). Elhorst et al. (2007) present a framework to determine the best of three estimators (2SLS, fixed effects 2SLS and first-difference 2SLS) in the presence of potential endogeneity using two Hausman type test-statistics. Using this framework, they conclude that the first-difference 2SLS is the preferred estimator of the East German wage curve, since the regional unemployment rate, the main explanatory variable of the wage rate, is not strictly exogenous and the spatial specific effects are not uncorrelated to the explanatory variables. To investigate the possible endogeneity of the regional unemployment rate in combination with time-specific effects, a similar framework is used, except for the first-difference 2SLS estimator. This is because first differencing does not assist in eliminating time specific effects. For this reason, they develop a spatial first-difference 2SLS estimator where the values of y and $\boldsymbol{X}$ in every spatial unit are taken in deviation of y and $\boldsymbol{X}$ in one reference spatial unit.

Finally, Elhorst (2008a) adopts the use of matrix exponentials, a transformation recently introduced by LeSage and Pace (2007). This transformation is different from the spatial lag model in Eq. (C.2.2) or the spatial error model in Eq. (C.2.3) in that its Jacobian term is zero. This zero Jacobian term opens the opportunity to use an estimation method partly based on IV and partly based on ML to control for endogeneity of one or more of the explanatory variables.

There has also been a growing interest in the estimation of dynamic panel data models. Elhorst (2005a) derives the ML estimator and Su and Yang (2007) the corresponding regularity conditions of a dynamic panel data model extended to include spatial error autocorrelation. Elhorst (2005b), Korniotis (2005), Yu et al. (2007) and Vrijburg et al. (2007) consider a dynamic panel data model extended to include a spatially lagged dependent variable. Up to now, the first of these six studies has also been applied successfully in the empirical work of other researchers (Kholodilin et al. 2008).

Acknowledgements. The author of this chapter is grateful to Maarten Allers, Jan Jacobs, Thomas Seyffertitz and the editors of this Handbook for useful suggestions and comments on an earlier draft.

References

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

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

Anselin L, Bera AK (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In Ullah A, Giles DEA (eds) Handbook of applied economic 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, 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 edition). Kluwer, Dordrecht, pp.901-969

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

Baltagi BH (1989) Applications of a necessary and sufficient condition for OLS to be BLUE. Stat Prob Letters 8(5):457-461

Baltagi BH (2005) Econometric analysis of panel data (3rd edition). Wiley, New York, Chichester, Toronto and Brisbane

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

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 and New York, pp.283-295

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

Beck N (2001) Time-series-cross-section data: what have we learned in the past few years? Ann Rev Pol Sci 4(1):271-293

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

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

Cressie NAC (1993) Statistics for spatial data (revised edition). Wiley, New York, Chichester, Toronto and Brisbane

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

Elhorst JP (2003) Specification and estimation of spatial panel data models. Int Reg Sci Rev 26(3):244-268

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

Elhorst J.P. (2005b) Models for dynamic panels in space and time; an application to regional unemployment in the EU. Paper presented at the Spatial Econometrics Workshop, April 8-9, 2005, Kiel

Elhorst JP (2008a) 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 (2008b) Serial and spatial autocorrelation. Econ Letters 100(3):422-424

Elhorst JP, Freret S (2009) Evidence of political yardstick competition in France using a two-regime spatial Dublin model with fixed effects. J Reg Sci. DOI: 10.1111/j.1467-9787.2009.00613.x [forthcoming]

Elhorst JP, Blien U, Wolf K (2007) New evidence on the wage curve: a spatial panel approach. Int Reg Sci Rev 30(2):173-191

Elhorst JP, Piras G, Arbia G (2006) Growth and convergence in a multi-regional model with space-time dynamics. Paper presented at the Spatial Econometric Workshop, May 25-27, 2006, Rome

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

Fingleton B (2008) A generalized method of moments estimator for a spatial panel data model with endogenous spatial lag and spatial moving average errors. Spat Econ Anal 3(1):27-44

Fingleton B, Le Gallo J (2007) Estimating spatial models with endogenous variables, a spatial lag en spatially dependent disturbances: finite sample properties. Paper presented at the First World Conference of the Spatial Econometrics Association, July 11-14, 2007, Cambridge

Florax RJGM, Folmer H (1992) Specification and estimation of spatial linear regression models. Reg Sci Urban Econ 22(3):405-432

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

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

Goldberger AS (1962) Best linear unbiased prediction in the generalized linear regression model. J Am Stat Assoc 57:369-375

Greene WH (2008) Econometric analysis (6th edition). Pearson, Upper Saddle River [NJ]

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 mis-specified. J Stat Plann Inference 69(1):153-174

Hendry DF (2006) A comment on ‘Specification searches in spatial econometrics: The relevance of Hendry's methodology’. Reg Sci Urban Econ 36(2):309-312

Hsiao C (2003) Analysis of Panel Data (2nd edition). Cambridge University Press, Cambridge

Hsiao C (2005) Why panel data? University of Southern California, IEPR Working Paper 05.33

Hunneman A, Bijmolt T, Elhorst JP (2007) Store location evaluation based on geographical consumer information. Paper presented at the Marketing Science Conference, June 28-30, 2007, Singapore

Jarque CM, Bera AK (1980) Efficient tests for normality, homoskedasticity and serial independence of regression residuals. Econ Letters 6(3):255-259

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 (1998) A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J Real Est Fin Econ 17(1):99-121

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
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
Korniotis GM (2005) A dynamic panel estimator with both fixed and spatial effects. Paper presented at the Spatial Econometrics Workshop, April 8-9, 2005, Kiel
Lahiri SN (2003) Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. Sankhya 65(2): 356-388
Lee LF (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econ Rev 22(4):307-335
Lee LF (2004) Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72(6):1899-1925
Leenders RTAJ (2002) Modeling social influence through network autocorrelation: Constructing the weight matrix. Soc Netw 24(1):21-47
LeSage JP (1999) Spatial econometrics. www.spatial-econometrics.com/html/sbook.pdf
LeSage JP, Pace RK (2007) A matrix exponential spatial specification. J Econometrics 140(1):190-214
Magnus JR (1982) Multivariate error components analysis of linear and non-linear regression models by maximum likelihood. J Econometrics 19(2):239-285
Magnus JR, Neudecker H (1988) Matrix differential calculus with applications in statistics and econometrics. Wiley, New York, Chichester, Toronto and Brisbane
Manski CF (1993) Identification of endogenous social effects:the reflection problem. Rev Econ Stud 60:531-542
Mood AM, Graybill F, Boes DC (1974) Introduction to the theory of statistics (3rd edition). McGraw-Hill, Tokyo
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 edition). Kluwer, Dordrecht, pp.3-22
Ord JK (1975) Estimation methods for models of spatial interaction. J Am Stat Assoc 70:120-126
Pace RK, Barry R (1997) Quick computation of spatial autoregressive estimators. Geogr Anal 29(3):232-246
Partridge MD (2005) Does income distribution affect U.S. state economic growth. J Reg Sci 45(2):363-394
Su L, Yang Z (2007) QML Estimation of dynamic panel fata models with spatial errors. Paper presented at the First World Conference of the Spatial Econometrics Association, July 11-14, 2007, Cambridge
Verbeek M (2000) A guide to modern econometrics. Wiley, New York, Chichester, Toronto and Brisbane
Vrijburg H, Jacobs JPAM, Ligthart JE (2007) A spatial econometric approach to commodity tax competition. Paper presented at the NAKE Research Day, October 24, 2007, Utrecht
Yang Z, Li C, Tse YK (2006) Functional form and spatial dependence in spatial panels. Econ Letters 91(1):138-145
Yu J, Jong R de, Lee L (2007) 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