

# Advances in Spatial Econometrics: Parametric vs. Semiparametric Spatial Autoregressive Models

Roberto Basile<sup>1</sup>(✉) and Román Mínguez<sup>2</sup>

<sup>1</sup> Department of Economics, Università della Campania “Luigi Vanvitelli”,  
Corso Gran Priorato di Malta, 1, 81043 Capua, CE, Italy  
roberto.basile@unina2.it

<sup>2</sup> Universidad de Castilla-La Mancha, Cuenca, Spain  
roman.minguez@uclm.es

**Abstract.** In this Chapter we provide a critical review of parametric and semi-parametric spatial econometric approaches. We focus on the capability of each class of models to fit the main features of spatial data (such as strong and weak cross-sectional dependence, spatial heterogeneity, nonlinearities, and time persistence), leaving aside the technicalities related to the estimation methods. We also provide a brief discussion of the existent software developed to estimate most of the econometric models exposed in this Chapter.

**Keywords:** Spatial econometrics · Semiparametric models

## 1 Introduction and Motivation

Nowadays, the dominant paradigm in spatial econometrics is still a parametric one. The *first generation* of spatial econometric models (essentially developed to handle cross-sectional data) focused on modeling spatial dependence (or spatial spillover effects) through different alternative linear specifications, such as the Spatial Lag or Spatial Autoregressive Model (SAR), the Spatial Error Model (SEM), the Spatial Durbin Model (SDM), the Spatial Autoregressive in  $X$ -variables Model (SLX), and a mix of SAR and SEM (SAC or SARAR) (Anselin 1988; LeSage and Pace 2009). We may call this collection of econometric tools as “econometrics of interaction”, since they can be applied to any kind of network relationship among different sample units.

During the last decade, these models have been extended to handle spatial panel data (or spatio-temporal data), that is data containing time series observations of a number of geographical units. Elhorst (2014b) defines them *second generation* spatial econometric models. By including a regional specific fixed or random effect, these models prove to be particularly useful to control for unobserved spatial heterogeneity, that is a fundamental task in empirical economic analyses, as failing to do so can introduce omitted-variable biases and preclude causal inference. Moreover, spatial dependence may simply be the consequence of (spatially correlated) omitted variables rather than being the result of spillovers. Thus, controlling both for spatial dependence (through

spatial lag terms) and spatial heterogeneity (through fixed or random effects) is a primary task when dealing with spatial data. More recent developments concern dynamic spatial panel data models and spatial VAR (Vector-Autoregressive) models, which allow to control for time persistence and reverse causality problems.

Notwithstanding these important advances in the literature, it is worth noting that any parametric model is limited to specific forms of spatial variation of the parameters, such as spatial regimes. They are not suitable for more general forms of spatial heterogeneity of model parameters, i.e. when the variation of parameters is continuous (smooth) over space and depends on coordinates, and when the functional form of the relationship between the dependent variable and the regressor is unknown (potentially non-monotonic). Moving away from the parametric approach, another strand of the spatial econometric literature has proposed semiparametric methods as more flexible estimation frameworks, thus following the recommendations of McMillen (2012) of using smoother techniques in order to remove spatial heterogeneity while considering other potential nonlinearities.

First, following Brunson et al. (1996); Cho et al. (2010) have proposed an approach that combines geographically weighted regression (GWR) and spatial autoregression (SEM) methods, called GWR-SEM. The spatial autoregressive error term should allay spatial dependency, while GWR addresses spatial heterogeneity by allowing the coefficients to vary across observations. In the same vein, Páez et al. (2002) propose an estimation method for cross-sectional data in which the covariance is locally varying and that can handle spatial autocorrelation of the error terms. Another notable contribution accounting for both spatial autocorrelation and nonstationarity of the parameters has been made by Pace and LeSage (2004): they propose a spatial autoregressive local estimation based on a recursive approach for maximum-likelihood estimation of SAR that implies estimates on subsamples related to a neighboring of each observation. More recently, combining kernel smoothing methods and standard spatial lag models, Geniaux and Martinetti (2017) have introduced a new class of data generating processes, called MGWR-SAR (Mixed Geographically Weighted Regression Simultaneous Autoregressive Model), in which the regression parameters and the spatial dependence coefficient can vary over space. The advantage of the last class of models is that it allows to consider the mixed case in which some parameters are constant over space and others are spatially varying.

Second, Basile et al. (2014); Montero et al. (2012) have combined penalized regression spline (PS) methods (Eilers et al. 2015) with standard cross-section spatial autoregressive models (such as SAR, SEM, SDM and SLX). An important feature of PS-SAR, PS-SEM, PS-SDM and PS-SLX models is the possibility to include within the same specification (i) spatial autoregressive terms to capture spatial interaction or network effects (thus avoiding *spatial dependence bias*), (ii) parametric and nonparametric (smooth) terms to identify nonlinear relationships between the response variable and the covariates (thus avoiding *functional form bias*), (iii) a geoadditive term, that is a smooth function of the spatial coordinates, to capture a spatial trend effect, that is to capture spatially autocorrelated unobserved heterogeneity (thus avoiding *spatial heterogeneity bias*), and (iv) the interaction between the geoadditive term and a covariate of particular interest to identify spatially varying effects of  $X$ -variables.

Third, Mínguez et al. (2017) have proposed an extension of the PS-SAR to spatio-temporal data when both a large cross-section and a large time series dimensions are available. With this kind of data it is possible to estimate not only spatial trends, but also spatio-temporal trends in a nonparametric way (Lee and Durbán 2011), so as to capture region-specific nonlinear time trends net of the effect of spatial autocorrelation. In other words, this approach allows to answer questions like: How do unobserved time-related factors (i.e. common factors), such as economic-wide technological or demand shocks, heterogeneously affect long term dynamics of all units in the sample? And how does their inclusion in the model affect the estimation of spatial interaction effects? In this sense, the PS-SAR model with spatio-temporal trend represents an alternative to parametric methods aimed at disentangling *common factors* effects (such as common business cycle effects) and *spatial dependence* effects (local interactions between spatial units generating spillover effects), where the former is sometimes regarded as ‘strong’ cross-sectional dependence, and the latter as ‘weak’ cross-sectional dependence (Chudik et al. 2011).

In this paper, we propose a critical review of parametric and semiparametric spatial econometric approaches trying to highlight their pros and cons. We will focus on the capability of each class of models to fit the main features of spatial data (such as strong and weak spatial dependence, spatial heterogeneity, nonlinearities, and time persistence) leaving the estimation techniques on backstage. The plan of the paper is as follows. Section 2 summarizes the huge literature on parametric spatial autoregressive models. Section 3 is dedicated to the broad category of semiparametric spatial autoregressive models, disentangling GWR (or MGWR) models based on kernel methods and models based on penalized spline smoothers. Section 4 provides a brief discussion of the software available for the practitioners to apply all these models. Finally, Sect. 5 concludes.

## 2 Parametric Spatial Autoregressive Models

### 2.1 Modeling Spatial Interaction Effects: Spatial Autoregressive Models for Cross-Sectional Data

Unlike time dependence, spatial dependence is a difficult concept to grasp, some people find. Let us start from a generic notion of “*interdependence*” and, then, return to the specific concept of spatial dependence. To introduce the concept of “*interdependence*”, let us consider a simple example. Imagine we want to model the scientific productivity (*SP*) of a sample of researchers connected among each other in a network of co-authorships. *SP* can be measured, for example, in terms of number of publications or better in terms of a continuous outcome variable such as an evaluation score whose distribution is assumed to be normal. For simplicity, we assume that this score depends only on investments in human capital (such as number of books read, number of new courses attended, number and length of academic visits abroad, and so on). To model  $y_i = SP_i$  for each individual researcher  $i$ , we start from the classical linear regression model:

$$y_i = \alpha + \sum_k \beta_k x_{ik} + \varepsilon_i \quad i = 1, \dots, N \quad \varepsilon_i \sim iid \mathcal{N}(0, \sigma_\varepsilon^2) \quad (1)$$

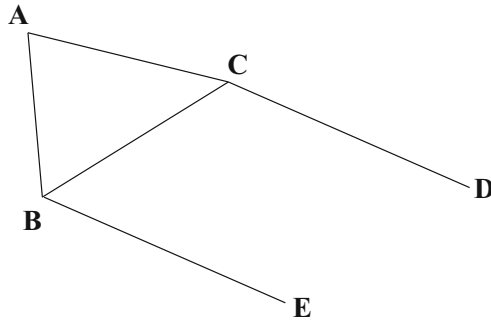
where  $x_{ik}$  indicates a measure of human capital investment. This model imposes a strong assumption of independence. First, the assumptions on the error term ( $\epsilon_i$ ) exclude any type of covariance. Second, the partial derivatives exclude any kind of indirect (interaction or spillover) effect, i.e. an investment in human capital by a researcher  $i$  will affect only his/her own scientific productivity ( $y_i$ ), but not the productivity of any other researcher ( $y_j$ ):

$$\partial E [y_i] / \partial x_{ik} = \widehat{\beta}_k \quad \partial E [y_j] / \partial x_{ik} = \partial E [y_i] / \partial x_{jk} = 0 \quad i, j = 1, \dots, N$$

We can write this model in matrix form as

$$\mathbf{y} = \iota_N \alpha + \mathbf{X}\beta + \boldsymbol{\epsilon} \quad E[\boldsymbol{\epsilon}] = 0 \quad E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] = \sigma^2 \mathbf{I}_N \quad (2)$$

The independence assumption is quite unrealistic, however. In fact, we cannot evaluate the scientific performance of this sample of individuals without taking into account the possibility of knowledge spillovers among them. Suppose that our sample is composed of only five researchers (identified by the letters A, B, C, D, E). Scientific collaborations (co-authorship relations) will determine a *network* or *connectivity scheme* such as the one shown in Fig. 1:



**Fig. 1.** A network scheme of scientific collaborations (co-authorship relations)

Researcher A has a co-authorship (that is a direct link) only with individuals B and C. Researcher B has a co-authorship only with individuals A, C and E; and so on. This network scheme can be translated into a symmetric  $5 \times 5$  binary matrix  $\mathbf{W}^*$ :

$$\mathbf{W}^* = \begin{bmatrix} & A & B & C & D & E \\ A & 0 & 1 & 1 & 0 & 0 \\ B & 1 & 0 & 1 & 0 & 1 \\ C & 1 & 1 & 0 & 1 & 0 \\ D & 0 & 0 & 1 & 0 & 0 \\ E & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

with  $w_{ij}^* = 1$  if  $i$  and  $j$  are classified as co-authors, and  $w_{ij}^* = 0$  otherwise. This binary matrix can be row-standardized so as  $w_{ij} = w_{ij}^* / \sum_j w_{ij}^*$  s.t.  $\sum_j w_{ij}^* = 1$ :

$$\mathbf{W} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Now, we can multiply  $\mathbf{W}$  by the vector  $\mathbf{y}$ :

$$\mathbf{W}\mathbf{y} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1/2y_2 + 1/2y_3 \\ 1/3y_1 + 1/3y_3 + 1/3y_5 \\ 1/3y_1 + 1/3y_2 + 1/3y_4 \\ y_3 \\ y_2 \end{bmatrix}$$

Each element of the vector  $\mathbf{W}\mathbf{y}$  measures the weighted average of the scientific productivity of the co-authors of each individual. We can also compute  $\mathbf{W}\mathbf{X}$  and  $\mathbf{W}\boldsymbol{\varepsilon}$ . These three terms can be used to extend model (2). For example, we can include  $\mathbf{W}\mathbf{y}$  on the r.h.s. of (2):

$$\mathbf{y} = \iota_N \alpha + \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim iid \mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{I}_N) \tag{3}$$

The reduced form of this model is:

$$\mathbf{y} = (\mathbf{I}_N - \rho \mathbf{W})^{-1} (\iota_N \alpha + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \quad (\mathbf{I}_N - \rho \mathbf{W})^{-1} = \mathbf{I}_N + \rho \mathbf{W} + \rho^2 \mathbf{W}^2 + \dots$$

To ensure that  $\mathbf{I}_N - \rho \mathbf{W}$  is invertible, one needs to impose some restrictions on the parameter  $\rho$ , which for a row-normalized interaction matrix  $\mathbf{W}$  correspond to take use of a compact set of  $(1/\omega_{min}, 1)$ , where  $\omega_{min}$  is the minimum eigenvalue of  $\mathbf{W}$  matrix. Once this restriction is satisfied, using the estimated parameters of the model ( $\hat{\rho}$  and  $\hat{\boldsymbol{\beta}}_k$ ), we can compute the impacts of a change in the  $k$ -th explanatory variable, i.e. the partial derivatives of the expected value of the dependent variable  $\mathbf{y}$  with respect to the concerned variable,  $\mathbf{x}_k$ :

$$\Xi_y^{x_k} = \partial E[\mathbf{y}] / \partial \mathbf{x}_k = (\mathbf{I}_N - \hat{\rho} \mathbf{W})^{-1} \hat{\boldsymbol{\beta}}_k \tag{4}$$

Unlike to what we observe for the traditional classical linear regression model, diagonal elements of (4) are different from each other, off diagonal elements differ from zero and the matrix itself is not symmetric. In particular, diagonal elements of (4) represent own-partial derivatives, meaning the impact of a change in the  $k$ -th variable in unit  $i$  on the expected value of the dependent variable in this unit. They are formally written as

$$\partial E[y_i] / \partial x_{ik} = [\Xi_y^{x_k}]_{ii} \quad i = 1, \dots, N \tag{5}$$

These own-partial derivatives are labeled *direct impacts* and include feedback loop effects that arise as a result of impacts passing through interacting units  $j$  and back

to unit  $i$ . As the set of interacting units is different for each unit, the feedback will be heterogeneous by nature, giving birth to the notion of *interactive heterogeneity*. This interactive heterogeneity should not be confused with *parameter heterogeneity*, which refers to instability of parameters (structural breaks, clubs) or heteroskedasticity.

Off-diagonal elements of (4) represent the effects of a change in the  $k$ -th explanatory variable in unit  $j$  on the dependent variable in unit  $i$ . As matrix (4) is asymmetric, this further imply that this impact will not be the same as the one caused by a change in unit  $i$  on unit  $j$ . Formally,

$$\partial E[y_i]/\partial x_{jk} = [\Xi_y^{x_k}]_{ij} \neq \partial E[y_j]/\partial x_{ik} = [\Xi_y^{x_k}]_{ji} \quad (6)$$

These cross-derivative elements are thus labeled *indirect effects*. Using expressions (5) and (6), we can for example say that an investment in human capital by individual A (i.e. an idiosyncratic shock in a  $x_k$  variable) will affect not only the scientific productivity of A (*direct effect*), but also the scientific productivity of his/her own co-authors (individual A will transmit part of the new knowledge to his/her own co-authors B and C), the co-authors of his/her co-authors and so on (*spillover* or *indirect effect*). Thus, we can say that there is a global diffusion of the idiosyncratic shock. Given the stability condition  $|\rho| < 1$ , the intensity of these knowledge spillovers decreases with the increase in the order of co-authorship relations. Since the matrix  $(\mathbf{I}_N - \widehat{\rho}\mathbf{W})^{-1}$  pre-multiplies also the error term, we can also say that there is a global diffusion of shocks in the unobserved term.

Eventually, we may introduce both  $\mathbf{W}\mathbf{y}$  and  $\mathbf{W}\mathbf{X}$  on the r.h.s. of Eq. (1):

$$\mathbf{y} = \iota_N\alpha + \rho\mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{W}\mathbf{X}\delta + \varepsilon \quad \varepsilon \sim iid.\mathcal{N}(0, \sigma_\varepsilon^2\mathbf{I}_N) \quad (7)$$

Again, the reduced form of this model implies a global diffusion of both observed and unobserved shocks. The matrix of partial derivatives of  $\mathbf{y}$  with respect to the  $k$ -th explanatory variable, presented in (8) and computed from the reduced form of model (7), contains the additional term  $\mathbf{W}\delta_k$ .

$$\Xi_y^{x_k} = \partial E[\mathbf{y}]/\partial \mathbf{x}_k = (\mathbf{I}_N - \widehat{\rho}\mathbf{W})^{-1}(\mathbf{I}_N\widehat{\beta}_k + \mathbf{W}\widehat{\delta}_k) \quad (8)$$

Alternatively, we can leave the systematic part of model (2) unchanged and introduce the assumption of spatial autocorrelation in the error term:

$$\begin{aligned} \mathbf{y} &= \iota_N\alpha + \mathbf{X}\beta + \varepsilon & \varepsilon &= \lambda\mathbf{W}\varepsilon + \mathbf{u} \\ |\lambda| < 1 & & \mathbf{u} &\sim iid.\mathcal{N}(0, \sigma_u^2\mathbf{I}_N) \end{aligned} \quad (9)$$

The reduced form of this model

$$\mathbf{y} = \iota_N\alpha + \mathbf{X}\beta + (\mathbf{I}_N - \lambda\mathbf{W})^{-1}\mathbf{u}$$

implies a global diffusion of random shocks, but not spillovers of idiosyncratic shocks in an observed variable. Thus, using model (9), in our example, we would exclude knowledge spillovers from observed changes in human capital investments of researcher A; only spillovers from unobserved factors would take place.

Finally, we may extend model (2) by introducing on the r.h.s. only  $\mathbf{WX}$ :

$$\mathbf{y} = \iota_N \alpha + \mathbf{X}\beta + \mathbf{WX}\delta + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim iid.\mathcal{N}(0, \sigma_\varepsilon^2 \mathbf{I}_N) \tag{10}$$

This model implies only *local spillovers*: an investment in new knowledge by individual A will spill over only to his/her own co-authors, and *vice-versa*.

Now, let's turn to a *spatial* context and imagine that the network structure depicted in Fig. 1 represents a *spatial* network, identifying direct neighborhood links (i.e. direct proximity relationships) between regions or firms in space. In spatial statistics and spatial econometrics,  $\mathbf{W}^*$  and  $\mathbf{W}$  are called the spatial weights matrix and the row-standardized spatial weights matrix, respectively.  $\mathbf{W}\mathbf{y}$  is called the spatial lag operator; it works to produce a weighted average of the neighboring observations. In spatial econometrics, model (3) is called the *Spatial Lag Model* or *Spatial Autoregressive Model* (SAR), model (7) is known as the *Spatial Durbin Model* (SDM), model (9) is known as the *Spatial Error Model* (SEM), and model (10) is known as the *Spatial in X-variable Model* (SLX). Each of them allows us to capture a different spatial spillover effect.

For example, using cross-regional data, one may estimate a SDM version of the so-called *knowledge production function*, according to which the knowledge produced in a region ( $K_i$ ) (approximated by the number of patents per capita or by the total factor productivity) is an increasing function of both internal and external cumulative research and development ( $R\&D_i, \sum_{j \neq i} w_{ij} \ln R\&D_j$ ), and both internal and external human capital stocks:

$$\begin{aligned} \ln K_i = & \alpha + \beta_1 \ln R\&D_i + \beta_2 \sum_{j \neq i} w_{ij} \ln R\&D_j \\ & + \beta_3 \ln H_i + \beta_4 \sum_{j \neq i} w_{ij} \ln H_j + \rho \sum_{j \neq i} w_{ij} \ln K_j + \varepsilon_i \end{aligned} \tag{11}$$

Technological spillovers among regions may be assumed to be driven by interregional trade relations, as suggested by the endogenous growth theory. Thus, if interregional trade data are available for the regional sample used in the analysis, a researcher may use them to build a  $\mathbf{W}$  matrix. Alternatively, spatial proximity measures (such binary contiguity measures or inverse distance) can be used.

It is worth noticing that a more parsimonious version of (11) is often estimated, which imposes zero values to parameters  $\beta_4$  and  $\rho$ , thus assuming only local spillovers from R&D investments carried out by direct neighboring regions and excluding global spillovers captured by a spatial multiplier mechanism. A natural way to proceed is to estimate model (11) and then test these restrictions on parameters parameters  $\beta_4$  and  $\rho$ .

The term  $\mathbf{W}\mathbf{y}$  that appears on the r.h.s. of (3) and (7) is correlated with the error term,  $Cov[\mathbf{W}\mathbf{y}; \boldsymbol{\varepsilon}] \neq \mathbf{0}$ , so that ordinary least squares (OLS) estimates are biased and inconsistent. Consistent and efficient estimates can be obtained by maximum likelihood (ML) or quasi-maximum likelihood estimates (QML) (Lee 2004). Two-Stage Least Squares (2SLS) estimates adapt well to the case of (3) because higher orders of spatial lags of the  $\mathbf{X}$  variables are natural candidates to be used as instrumental variables (Kelejian and Prucha 1997). A more efficient estimator is the method of moments estimator (MM) (Kelejian and Prucha 2001). Lee (2004) generalized the MM approach into a

fully generalized method of moments (GMM) estimator for the case of the SDM model (7), while Liu et al. (2007) proposed a GMM estimator for a SDM with dependent structures in the error term. The GMM estimator may have, under general conditions, the same limiting distribution as the ML or QML estimators. Moreover, the 2SLS and the GMM estimators allow the researcher to take into account any endogeneity problems in the r.h.s., different from the spatial lag of  $\mathbf{y}$ .

As mentioned above, direct, indirect and total marginal effects change across spatial units. Specifically, they depend on the specific position of the region within the spatial proximity network. Thus, in order to summarize the results, it could be easier to compute average measures of direct, indirect and total effects. In the case of Eq. 3, the average total marginal effect is computed as  $N^{-1}\mathbf{i}'_N \left[ (\mathbf{I}_N - \rho \mathbf{W})^{-1} \mathbf{I}_N \beta_k \right] \mathbf{i}_N$  (see Table 1). The average direct impact is  $N^{-1}tr \left[ (\mathbf{I}_N - \rho \mathbf{W})^{-1} \mathbf{I}_N \beta_k \right]$ , while the average indirect (spatial spillover) impact is the difference between average total and average indirect effects. In order to draw inference regarding the statistical significance of average direct and indirect effects, LeSage and Pace (2009, p. 39) suggest simulating the distribution of these effects using the variance-covariance matrix implied by the ML estimates. Efficient simulation approaches can be used to produce an empirical distribution of the parameters  $\alpha, \beta, \theta, \rho, \sigma^2$  that are needed to calculate the scalar summary measures. This distribution can be constructed using a large number of simulated parameters drawn from the multivariate distribution of the parameters implied by the ML estimates.

## 2.2 Modeling Spatial Spillovers and Unobserved Spatial Heterogeneity: Spatial Autoregressive Models for Panel Data

### 2.2.1 Static Spatial Panel Data Models

Recently, spatial econometric models have been extended to deal with spatial panel data, that is data with both a spatial and a temporal dimension (Elhorst 2014b). The two-dimensional structure of the data allows us to control for unobserved spatial and time heterogeneity by including individual (spatial) and time effects on the r.h.s. of the model. Thus, for example, the static panel data SAR model can be written in vector form for a cross-section of observations at time  $t$  ( $t = 1, 2, \dots, T$ ) as:

**Table 1.** Average total (ATE), direct (ADE), and indirect (AIE) marginal effects

Model	ADE	AIE	ATE
Linear	$\beta_k$	0	$\beta_k$
SAR	$\frac{tr \left[ (\mathbf{I}_N - \rho \mathbf{W})^{-1} \mathbf{I}_N \beta_k \right]}{N}$	Total-Direct	$\frac{\mathbf{i}'_N \left[ (\mathbf{I}_N - \rho \mathbf{W})^{-1} \mathbf{I}_N \beta_k \right] \mathbf{i}_N}{N}$
SEM	$\beta_k$	0	$\beta_k$
SDM	$\frac{tr \left[ (\mathbf{I}_N - \rho \mathbf{W})^{-1} (\mathbf{I}_N \beta_k + \mathbf{W} \theta_k) \right]}{N}$	Total-Direct	$\frac{\mathbf{i}'_N \left[ (\mathbf{I}_N - \rho \mathbf{W})^{-1} (\mathbf{I}_N \beta_k + \mathbf{W} \theta_k) \right] \mathbf{i}_N}{N}$
SLX	$\beta_k$	$\theta_k$	$\beta_k + \theta_k$



$$\begin{aligned} \mathbf{y}_t &= \rho \mathbf{W}\mathbf{y}_t + \alpha + \iota_N \tau_t + \mathbf{X}_t \beta + \varepsilon_t \\ E(\varepsilon_t) &= 0 \quad E(\varepsilon_t \varepsilon_t) = \sigma^2 \mathbf{I}_N \end{aligned} \tag{12}$$

where, again,  $\mathbf{W}$  is a row-standardized  $N \times N$  spatial weights matrix whose diagonal elements  $w_{ii}$  are 0;  $\rho$  is the spatial spillover parameter satisfying the usual stability conditions, and  $\rho \sum_{j=1}^N w_{ij} y_{jt}$  captures the spatial spillover effects net of the unobserved heterogeneity effects filtered out by the spatial fixed effects,  $\alpha_i$ , and time fixed effects,  $\tau_t$ .

Similarly, the static panel SEM can be expressed as:

$$\begin{aligned} \mathbf{y}_t &= \alpha + \iota_N \tau_t + \mathbf{X}_t \beta + \varphi_t \quad \varphi_t = \lambda \mathbf{W} \varphi_t + \varepsilon_t \\ E(\varepsilon_t) &= 0 \quad E(\varepsilon_t \varepsilon_t) = \sigma^2 \mathbf{I}_N \end{aligned} \tag{13}$$

And, the static panel data SDM as:

$$\begin{aligned} \mathbf{y}_t &= \rho \mathbf{W}\mathbf{y}_t + \alpha + \iota_N \tau_t + \mathbf{X}_t \beta + \mathbf{W}\mathbf{X}_t \theta + \varepsilon_t \\ E(\varepsilon_t) &= 0 \quad E(\varepsilon_t \varepsilon_t) = \sigma^2 \mathbf{I}_N \end{aligned} \tag{14}$$

For example, the static panel version of the spatial Durbin knowledge production function (11) reads as:

$$\begin{aligned} \ln K_{it} &= \beta_1 \ln R\&D_{it} + \beta_2 \sum_{j \neq i} w_{ij} \ln R\&D_{jt} \\ &+ \beta_3 \ln H_{it} + \beta_4 \sum_{j \neq i} w_{ij} \ln H_{jt} + \rho \sum_{j \neq i} w_{ij} \ln K_{jt} + \alpha_i + \tau_t + \varepsilon_{it} \end{aligned} \tag{15}$$

Depending on the assumptions about individual and time effects, these models will be estimated using *fixed effects* (FE) or *random effects* (RE). The latter, more efficient, is adequate when the effects (individual and temporal) are independent from all regressors included in the specification and are traditionally assumed normally distributed. When this hypothesis of independence is rejected, either on the basis of a test statistic (Hausman, Lagrange multiplier (LM) or likelihood ratio (LR)) or from economic insights, the fixed effects specification should be preferred. Even though these two estimation procedures are different, they both consist in first transforming the data (either applying the within operator for the fixed effects or a quasi-within transformation when the random effects estimation is used) and then applying standard spatial econometrics techniques (for example, the QML estimator; Lee and Yu 2010a) on these transformed data to obtain the estimated parameters.

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  $\hat{\alpha}_i$  is  $T$ . Importantly, sampling more observations in the cross-sectional domain is not a 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  $\hat{\alpha}_i$  is not transmitted to the estimator of the slope coefficients  $\hat{\beta}$  in the demeaned equation, since this estimator is not a function of the estimated  $\hat{\alpha}_i$ . Consequently, the incidental parameters problem does

not matter when  $\hat{\beta}$  are the coefficients of interest and the spatial fixed effects  $\hat{\alpha}_i$  are not, which is the case in many empirical studies.

Finally, it is important to recognize that, apart from the control for unobserved heterogeneity, the economic interpretation of static spatial autoregressive models is the same as the one for cross-sectional data. Impacts measures implied by a spatial static panel data model are indeed the same as those in a spatial autoregressive model for cross-sectional data, as soon as the interaction matrix and the parameters of interest of the former are assumed constant across time. Different is the case of spatial dynamic panel data models, which give rise to the possibility of evaluating the effects of transitory and permanent shocks both in the short-run and in the long-run equilibrium.

### 2.2.2 Dynamic Spatial Panel Data Models

In order to simultaneously deal with time persistence and spatial interdependence along with spatial and temporal heterogeneity, a dynamic spatial panel data model with fixed spatial and time effects is needed. The spatial econometric literature provides several alternative specifications of spatial dynamic models. A very general one includes time lags of both the dependent and independent variables, contemporaneous spatial lags of both, and lagged spatial lags of both. However, as Elhorst (2014b) have pointed out, this generalized model suffers from identification problems, and is thus not useful for empirical research. A more parsimonious model (written in vector form for a cross-section of observations at time  $t$ ) can be expressed as:

$$\mathbf{y}_t = \tau \mathbf{y}_{t-1} + \rho \mathbf{W} \mathbf{y}_t + \eta \mathbf{W} \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{W} \mathbf{X}_t \boldsymbol{\theta} + \alpha + \lambda_t \mathbf{1}_N + \boldsymbol{\varepsilon}_t \quad (16)$$

$$\boldsymbol{\varepsilon}_t \sim iid \mathcal{N}(0, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}_N)$$

Lee and Yu (2010b); Yu et al. (2008) have proposed bias corrected QML estimators for a dynamic model with spatial and time fixed effects. However, these estimators are based on the assumption of only exogenous covariates except for the time and spatial lag terms. Kukenova and Monteiro (2008) have suggested to use System-GMM estimator Blundell and Bond (1998) for dynamic spatial panel model with several endogenous variables. More specifically, they have investigated the finite sample properties of different estimators for spatial dynamic panel models (namely, spatial ML, spatial dynamic ML, least-square-dummy-variable, Diff-GMM and System-GMM) and concluded that, in order to account for the endogeneity of several covariates, spatial dynamic panel models should be estimated using System-GMM.

The stationarity conditions on the spatial and temporal parameters in a dynamic spatial panel data model like (16) go beyond the standard condition  $|\tau| < 1$  in serial models, and the standard condition  $1/\omega_{min} < \rho < 1$  in spatial models. Indeed, to achieve stationarity in the dynamic spatial panel data model (16), the characteristic roots of the matrix  $(\mathbf{I}_N - \rho \mathbf{W})^{-1}(\tau \mathbf{I}_N + \eta \mathbf{W})$  should lie within the unit circle (Debarsy et al. 2012) which is the case when

$$\begin{aligned} \tau + (\rho + \eta)\omega_{max} &< 1 && \text{if } \rho + \eta \geq 0 \\ \tau + (\rho + \eta)\omega_{min} &< 1 && \text{if } \rho + \eta < 0 \\ \tau - (\rho - \eta)\omega_{max} &> -1 && \text{if } \rho - \eta \geq 0 \\ \tau - (\rho - \eta)\omega_{min} &> -1 && \text{if } \rho - \eta < 0 \end{aligned}$$

Assuming that the matrix  $(\mathbf{I}_N - \rho \mathbf{W})^{-1}$  is invertible, the reduced form of model (16) can be re-written as

$$\mathbf{y}_t = (\mathbf{I}_N - \rho \mathbf{W} \mathbf{y}_t)^{-1} (\tau \iota_N + \eta \mathbf{W}) \mathbf{y}_{t-1} + (\mathbf{I}_N - \rho \mathbf{W} \mathbf{y}_t)^{-1} (\mathbf{X}_t \beta + \mathbf{W} \mathbf{X}_t \theta + \alpha + \lambda_t \iota_N + \varepsilon_t)$$

Taking the partial derivatives of the expected value of  $\mathbf{y}$  with respect to each  $k$ -th variable in  $\mathbf{X}$  in each unit  $i$  at each time  $t$ , we then obtain the so-called impacts matrices in the *short run*:

$$\left[ \frac{\partial E(\mathbf{y})}{\partial x_{k1}} \dots \frac{\partial E(\mathbf{y})}{\partial x_{kN}} \right]_t = (\mathbf{I}_N - \hat{\rho} \mathbf{W})^{-1} (\hat{\beta}_k \mathbf{I}_N + \mathbf{W}_t \hat{\theta}_k)$$

and in the *long run*:

$$\left[ \frac{\partial E(\mathbf{y})}{\partial x_{k1}} \dots \frac{\partial E(\mathbf{y})}{\partial x_{kN}} \right] = [(1 - \hat{\tau}) \mathbf{I}_N - (\hat{\rho} + \hat{\eta}) \mathbf{W}]^{-1} (\hat{\beta}_k \mathbf{I}_N + \mathbf{W}_t \hat{\theta}_k)$$

The diagonal elements of both matrices give a measure of the so-called *direct effect*. The off-diagonal elements of the matrices give a measure of the so-called *indirect or spillover effect* (Table 2).

**Table 2.** Average total, direct, and indirect short-term and long marginal effects in dynamic spatial panels.  $\bar{\cdot}$ : operator that calculates the mean diagonal element of a matrix.  $\overline{rsum}$ : operator that calculates the mean row sum of the non-diagonal elements.

Time horizon	ADE
Short-term	$\left[ (\mathbf{I}_N - \hat{\rho} \mathbf{W})^{-1} (\hat{\beta}_k \mathbf{I}_N + \hat{\theta}_k \mathbf{W}) \right]^{\bar{\cdot}}$
Long-term	$\left[ [(1 - \hat{\tau}) \mathbf{I}_N - (\hat{\rho} + \hat{\eta}) \mathbf{W}]^{-1} (\hat{\beta}_k \mathbf{I}_N + \hat{\theta}_k \mathbf{W}) \right]^{\bar{\cdot}}$
	AIE
Short-term	$\left[ (\mathbf{I}_N - \hat{\rho} \mathbf{W})^{-1} (\hat{\beta}_k \mathbf{I}_N + \hat{\theta}_k \mathbf{W}) \right]^{\overline{rsum}}$
Long-term	$\left[ [(1 - \hat{\tau}) \mathbf{I}_N - (\hat{\rho} + \hat{\eta}) \mathbf{W}]^{-1} (\hat{\beta}_k \mathbf{I}_N + \hat{\theta}_k \mathbf{W}) \right]^{\overline{rsum}}$

Moreover, Debarsy et al. (2012) derive the algorithms to calculate partial derivatives that can quantify the magnitude and timing of dependent variable responses in each region at various time horizons  $t + T$  to changes in the explanatory variables at time  $t$ . They also distinguish between two different interpretative scenarios, one where the change in explanatory variables represents a *permanent* or sustained change in the level and the other where we have a *transitory* (or one-period) change.

In particular, the  $T$ -period-ahead (cumulative) impact arising from a *permanent change* at time  $t$  in the  $k$ -th variable is<sup>1</sup>:

$$\partial \mathbf{Y}_{t+T} / \partial \mathbf{X}^k = \sum_{s=0}^T \mathbf{D}_s [\mathbf{I}_N \beta_k + \mathbf{W} \theta_k] \tag{17}$$

<sup>1</sup> By a permanent change at time  $t$  they mean that:  $\partial \mathbf{X}^k = (\mathbf{x}_t + \delta, \mathbf{x}_{t+1} + \delta, \dots, \mathbf{x}_T + \delta)$ , so the values increase to a new level and remain there in future time periods.

where  $D^s = (-1)^s (\mathbf{B}^{-1} + \mathbf{C})^s \mathbf{B}^{-1}$ , with  $s = 0, \dots, T-1$ ,  $\mathbf{B} = (\mathbf{I}_N - \rho \mathbf{W})$ , and  $\mathbf{C} = -(\tau \mathbf{I}_N + \eta \mathbf{W})$ .

The main diagonal elements of the  $N \times N$  matrix sums in (17) for time horizon  $T$  represent (cumulative) own-region impacts that arise from both time and spatial dependence. The sum of off-diagonal elements of this matrix reflect both *spillovers* measuring contemporaneous cross-partial derivatives, and *diffusion* measuring cross-partial derivatives that involve different time periods.<sup>2</sup>

The  $T$ -horizon impulse response to a *transitory* change in the  $k$ -th explanatory variable at time  $t$  would be given by the main and off-diagonal elements of:

$$\partial \mathbf{Y}_{t+T} / \partial \mathbf{X}^k = \mathbf{D}_T [\mathbf{I}_N \beta_k + \mathbf{W} \theta_k] \quad (18)$$

where  $\mathbf{D}_T = (-1)^T (\mathbf{B}^{-1} \mathbf{C})^T \mathbf{B}^{-1}$ .

Getting back to the example of the knowledge production function, the spatial dynamic version of (15) would be:

$$\begin{aligned} \ln K_{it} = & \beta_1 \ln R \& D_{it} + \beta_2 \sum_{j \neq i} w_{ij} \ln R \& D_{jt} + \beta_3 \ln H_{it} + \beta_4 \sum_{j \neq i} w_{ij} \ln H_{jt} \\ & + \tau \ln K_{i,t-1} + \rho \sum_{j \neq i} w_{ij} \ln K_{jt} + \eta \sum_{j \neq i} w_{ij} \ln K_{j,t-1} + \alpha_i + \tau_i + \varepsilon_{it} \end{aligned} \quad (19)$$

The estimation of this model would allow us to compute not only spatial (contemporaneous) R&D spillovers, but also spatio-temporal diffusion processes of R&D shocks originating in a region (or a country).

### 2.3 Modeling Spatial Dependence, Spatial Heterogeneity and Common Factors: Spatial Autoregressive Models for Large Panel Data

When spatial panel data have both a large cross-sectional and a large time series dimension, it becomes important to distinguish between spatial spillover effects and common factors. As discussed above, spatial spillovers are due to unobserved idiosyncratic shocks which propagate to all other regions with a distance-decay mechanism driven by network relationships. Instead, common factors are unobserved time-related factors which influence all regions (probably heterogeneously). Both determine cross-sectional correlation in the residuals and make it difficult to get unbiased and efficient estimates.

On the one hand, spatial spillover effects can be analyzed by using, for example, the spatial autoregressive model with fixed effects, described above. On the other hand, *strong* cross-sectional dependence can be accommodated by the Common Correlated Effects Pooled (CCEP) estimator proposed by Pesaran (2006). Suppose that  $y_{it}$  is generated by the following DGP with a multifactor error structure:

$$\begin{aligned} y_{it} &= \alpha_i + \mathbf{x}'_{it} \beta + \varepsilon_{it} \\ \mathbf{x}_{it} &= \gamma'_i \mathbf{f}_t + v_{it} \end{aligned} \quad (20)$$

<sup>2</sup> The term *spillover* is referred to contemporaneous cross-partial derivatives, those that involve the same time period. These cross-partial derivatives involving different time periods are referred to as diffusion effects, since diffusion takes time.

where  $\mathbf{f}_t$  is a  $m \times 1$  vector of common factors (introduced to allow for unobserved cross-sectional dependence), and  $\gamma_i$  the corresponding heterogeneous response.  $\mathbf{f}_t$  are allowed to be correlated with  $\mathbf{x}_{it}$ , while the idiosyncratic errors,  $\varepsilon_{it}$ , are assumed to be independently distributed over  $\mathbf{x}_{it}$ . Pesaran (2006) shows that, for sufficiently large  $N$ , it is valid to use cross-sectional averages of  $y_{it}$  and  $\mathbf{x}_{it}$  as observable proxies for  $\mathbf{f}_t$ . Thus, consistent  $\beta$  parameters can be estimated using the so-called CCEP estimator, which can be viewed as a generalized fixed effects estimator<sup>3</sup>:

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\beta + \delta_i\bar{\mathbf{x}}_t + \eta_i\bar{y}_t + \varepsilon_{it} \tag{21}$$

where  $\bar{\mathbf{x}}_t = N^{-1} \sum_{i=1}^N \mathbf{x}_{it}$  and  $\bar{y}_t = N^{-1} \sum_{i=1}^N y_{it}$ .

The CCEP approach has been proved to be valid in presence of both strong and weak (or semi-strong and semi-weak) cross dependence (Chudik et al. 2011; Pesaran and Tosetti 2011). Thus, it can easily collect even the pure spatial spillover effects. However, economic analyses often requires the assessment of the different forms of cross dependence, or better still, they require the assessment of spatial network effects, net of the effects of common factors. A natural way to deal with this problem is to combine the two approaches.

Using slightly different frameworks, Bai and Li (2015); Bailey et al. (2016); Shi and Lee (2016); Vega and Elhorst (2016) consider a joint modeling of spatial interaction effects and common-shocks effects:

$$y_{it} = \alpha_i + \rho \sum_{j=1}^N w_{ij,N} y_{jt} + \mathbf{x}'_{it}\beta + \gamma'_i \mathbf{f}_t + \varepsilon_{it} \tag{22}$$

This model (we may call it SAR-CCEP model) allows one to test which type of effects (common shocks,  $\gamma'_i \mathbf{f}_t$ , and/or spatial spillovers,  $\rho \sum_{j=1}^N w_{ij,N} y_{jt}$ ) is responsible for the cross-sectional dependence. Bai and Li (2015); Shi and Lee (2016) use principle components to estimate common factors, while Bailey et al. (2016); Vega and Elhorst (2016) follow Pesaran (2006) in using cross-sectional averages of  $y_{it}$  and  $\mathbf{x}_{it}$  as observable proxies for  $\mathbf{f}_t$ . Bailey et al. (2016) propose a two-stage estimation and inference strategy, whereby in the first step strong cross-sectional dependence is modeled by means of a factor model. Residuals from such factor models, referred to as de-factored observations, are then used to model the remaining weak cross dependencies, making use of spatial econometrics techniques. Vega and Elhorst (2016), instead, suggest to model common factors and spatial dependence simultaneously in a single-step procedure. All these authors show that the QMLE is an effective way of estimating this model.

Getting back to the example of the knowledge production function, the SAR-CCEP version of (15) would be:

$$\begin{aligned} \ln K_{it} = & \beta_1 \ln R\&D_{it} + \beta_2 \sum_{j \neq i} w_{ij} \ln R\&D_{jt} + \beta_3 \ln H_{it} + \beta_4 \sum_{j \neq i} w_{ij} \ln H_{jt} \\ & + \rho \sum_{j \neq i} w_{ij} \ln K_{jt} + \alpha_i + \gamma'_i \mathbf{f}_t + \varepsilon_{it} \end{aligned} \tag{23}$$

<sup>3</sup> The assumption of fixed  $\beta$  parameters can be relaxed, and a random coefficient specification can be assumed:  $\beta_i = \beta + u_i$ , with  $u_i \sim i.i.d.(\mathbf{0}, \Omega_u)$ . In this case the estimator proposed by Pesaran (2006) is the common correlated effects mean group (CCEMG) estimator.

Strong cross-sectional dependence in the errors of a knowledge production function may arise as a result of unobserved common factors, including, for instance, aggregate technological shocks, national policies intended to raise the level of technology or oil price shocks that may influence TFP through their effects on product costs. The heterogeneous effects of these factors may be the result, for instance, of country-specific technological constraints (Ertur and Musolesi 2016). Cross-sectional dependence in the errors of a knowledge production function can also be regarded as a result of spatial effects. Thus, a SAR-CCEP version of the knowledge production function seems to be a natural choice when the panel data is large enough.

Some drawbacks of this approach are worth noticing. First, there is a large number of incidental parameters under the joint modeling. Admittedly, this is not a serious problem as long as the model is linear, since inconsistency in the estimation of the incidental parameters is not transmitted to the estimation of the slope parameters of interest ( $\beta$ ); but, it may create a problem when nonlinear terms are considered. Second, the ability of the SAR-CCEP method to capture strong cross-sectional dependence and to disentangle spatial spillover effects and common factor effects is crucially affected by the set of covariates included in the model. On the one hand, if the estimated model contains one or only a few regressors, the CCEP estimator may not fully control for cross-sectional correlation (few regressors implies few cross-sectional averages as proxies for unobserved common factors); on the other hand, if the model includes many regressors, the resulting large number of cross-sectional averages hardly leave space for residual spatial spillovers. In Sect. 3.3, we review an alternative semiparametric approach to filter common-factor (or time-related) effects and, thus, to assess the presence of “residual” spatial dependence effects which adequately addresses these problems.

### 3 Semiparametric Spatial Autoregressive Models

Parametric spatial econometric frameworks described above are unfeasible in the simultaneous presence of different sources of model misspecification, such as substantial spatial dependence, nonlinear relationship of spatially correlated independent variables, unobserved spatial heterogeneity, spatially varying relationships, and common factors. Nonlinearities, spatial heterogeneity and time-related factors can cause spatial (or, more generally, cross-sectional) dependence and the reverse is also true. Studies that consider simultaneously spatial dependence, spatial heterogeneity, nonlinearities and common factors are still scarce in spatial econometrics literature. The recent contributions of Basile et al. (2014); Geniaux and Martinetti (2017); Mínguez et al. (2017) represent some attempts to promote more flexible estimation frameworks to address this problem.

#### 3.1 Modeling Spatial Heterogeneity and Spatial Dependence: MGWR-SAR

What are the economic motivations underlying the specification of a spatially-varying coefficient model? First, one can argue that models which only consider spatial autocorrelation are not capable of correcting all the problems related to non-observable spatial heterogeneity. This has pushed several authors to consider a non-stationary intercept term amongst the regression variables, for example by means of a smooth interaction of

the spatial coordinates, known as spatial trend (Wood 2006).<sup>4</sup> Nevertheless, this argument can be extended to consider a model with spatially-varying slope coefficients. It is also possible to consider a non-stationary spatial autocorrelation parameter. Indeed, when the spatial weight matrix  $\mathbf{W}$  is unknown and spatial locations are irregularly distributed over space, the choice of a neighboring scheme based only on distance or first nearest neighbors can be tricky. Choosing one weighting scheme instead of the other can lead to a spatial interaction matrix that is too dense or too dispersed in the heterogeneous parts of the space, resulting in under or overestimation of the parameters. Hence, the use of a non-stationary spatial autocorrelation parameter could mitigate the effect of the spatial weight matrix misspecification.

Very recently, Geniaux and Martinetti (2017) have introduced a new class of models, called MGWR-SAR (Mixed Geographically Weighted Regression Simultaneous AutoRegressive models), where the regression parameters and the spatial dependence coefficient can vary over space. In its most general form, the MGWR-SAR is specified as:

$$\mathbf{y} = \rho(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}; h)\mathbf{W}\mathbf{y} + \mathbf{X}^*\beta^* + \beta(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}; h)\mathbf{X} + \varepsilon \tag{24}$$

where  $\mathbf{y}$  is the  $N$ -vector of the continuous dependent variable,  $\mathbf{X}^*$  is a matrix of  $k_1$  exogenous explanatory variables entering the model linearly (i.e. with spatially stationary coefficients  $\beta^*$ ), while  $\mathbf{X}$  is a matrix of  $k_2$  exogenous explanatory variables with non-stationary coefficients  $\beta(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}; h)$ ,  $\mathbf{x}_{s_1}, \mathbf{x}_{s_2}$  are spatial coordinates,  $\mathbf{W}$  is the spatial weights matrix,  $\rho$  the spatial spillover parameter,  $\varepsilon$  is an i.i.d. error vector.

Thus, Geniaux and Martinetti (2017) relax one of the main hypothesis generally adopted by existing estimators of SAR models, i.e. the spatial parameter  $\rho$  and the regression parameters  $\beta$  are constant over the coordinates space. In fact, in equation (24) the value of  $\rho$  and  $\beta$  depends on the coordinates. The parameters  $\rho(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$  and  $\beta(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$  are only required to be spatially smoothed. The degree of smoothness depends on the bandwidth parameter  $h$  which allows to define the local sub-sample around the coordinates of each point  $(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$  using a given kernel function.

Because of the presence of the endogenous spatial lag term ( $\mathbf{W}\mathbf{y}$ ) on the r.h.s. of Eq. (24), the marginal effects of a change in  $\mathbf{X}^*$  or in  $\mathbf{X}$  must be computed starting from the reduced form of the model. Specifically, the marginal effect of a change in  $\mathbf{X}^*$  is:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{X}^*} = [\mathbf{I}_N - \rho(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}; h)\mathbf{W}]^{-1} \beta^* \tag{25}$$

while the marginal effect of a change in  $\mathbf{X}$  is:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{X}} = [\mathbf{I}_N - \rho(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}; h)\mathbf{W}]^{-1} \beta(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}; h) \tag{26}$$

For the estimation of these new models, Geniaux and Martinetti (2017) resort to the Spatial Two-Stage Least Squares (S2SLS) technique. In particular, they use a 5-step approach, a local linear estimator (a variant of the GWR) and Cross Validation for the selection of the bandwidth parameter.

<sup>4</sup> It is worth noting that the spatial trend term can be included even in a model for cross-sectional data, while spatial fixed or random effects can be included in a model to control for spatial unobserved heterogeneity only when spatial panel data are available.

Using cross-regional data, one may for example estimate a *knowledge production function* with heterogeneous parameters:

$$\begin{aligned} \ln K_i = & \alpha(\mathbf{x}_{s_1,i}, \mathbf{x}_{s_2,i}) + \beta_1(\mathbf{x}_{s_1,i}, \mathbf{x}_{s_2,i}) \ln R\&D_i \\ & + \beta_2(\mathbf{x}_{s_1,i}, \mathbf{x}_{s_2,i}) \ln H_i + \rho(\mathbf{x}_{s_1,i}, \mathbf{x}_{s_2,i}) \sum_{j \neq i} w_{ij} \ln K_j + \varepsilon_i \end{aligned} \quad (27)$$

The regional learning process of generating and transferring knowledge may be affected by local social capital, i.e. the institutional and cultural context of local networks, trust and conventions. Therefore, heterogeneous region-specific conditions are a source of spatial heterogeneity in intra-regional knowledge creation. In addition, heterogeneous region-specific conditions are related with the regional capacity of exploiting external knowledge sources. Thus, model 27 would allow a researcher to assess the spatial stationarity (homogeneity) of the parameters associated to R&D investments and to human capital investments, as well as the spatial stationarity of the spatial knowledge spillover parameter ( $\rho$ ). Nonstationarity may be evident by inspection of basic maps, and can be formally tested. For example, Kang and Dallerba (2016) have investigated the spatial heterogeneity in the marginal effects of a regional knowledge production function by using nonparametric local modeling approaches such as GWR and mixed GWR with two distinct samples of the US Metropolitan Statistical Area (MSA) and non-MSA counties. The results indicate a high degree of spatial heterogeneity in the marginal effects of the knowledge input variables, more specifically for the local and distant spillovers of private knowledge measured across MSA counties. On the other hand, local academic knowledge spillovers are found to display spatially homogeneous elasticities in both MSA and non-MSA counties.

A characteristic of this approach is that it only considers spatial parameter heterogeneity (i.e. parameter heterogeneity over the coordinates space), while neglecting the possibility of pure nonlinearities (i.e. parameter heterogeneity over the domain of the explanatory variable). Nevertheless, it remains very important to assess the existence of pure nonlinearities in the relationship between the response variable and the covariates. In fact, regional and urban economic development literature often predicts threshold effects (for example in growth theory) or monotonic relationships (for example in urban economics). Moreover, keeping the spatial autocorrelation parameter ( $\rho$ ) constant over space is a valid option: in that case, the feedback effects of spatial autocorrelation have a clearer definition and the interpretation of direct and indirect effects is easier.

### 3.2 Modeling Spatial Dependence, Spatial Heterogeneity and Nonlinearities: P-Spline Models for Cross-Sectional Data and Short Panels

Another recent strand of the spatial econometric literature has proposed *Spatial Autoregressive Semiparametric Geoadditive Models* as a means of simultaneously dealing with different critical issues typically encountered when using spatial economic data; namely, spatial dependence, spatial heterogeneity and unknown functional form (Basile et al. 2014; Montero et al. 2012). This approach combines penalized regression spline (PS) methods (Eilers et al. 2015) with standard spatial autoregressive models (such as SAR, SEM, SDM and SLX). An important feature of these models is that they make it



possible to include within the same specification: (i) spatial autoregressive terms to capture spatial interaction or network effects; (ii) parametric and nonparametric (smooth) terms to identify nonlinear relationships between the response variable and the covariates; and (iii) a geoadditive term, i.e. a smooth function of the spatial coordinates, to capture a spatial trend effect, that is, to capture spatially autocorrelated unobserved heterogeneity.

The structural form of the Penalized-Spline Spatial Lag model (PS-SAR) is:

$$\begin{aligned}
 \mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}^* \boldsymbol{\beta}^* + f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + f_3(\mathbf{x}_3, \mathbf{x}_4) \\
 + f_4(\mathbf{x}_1) \mathbf{z} + \dots + h(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + \varepsilon
 \end{aligned}
 \tag{28}$$

where  $\mathbf{y}$  is a continuous univariate output variable,  $\mathbf{W}\mathbf{y}$  its spatial lag,  $\mathbf{X}^* \boldsymbol{\beta}^*$  is the linear predictor for any strictly parametric component (including the intercept, all categorical covariates and eventually a set of continuous covariates).  $f_k(\cdot)$  are unknown smooth functions of univariate continuous covariates or bivariate interaction surfaces of continuous covariates, capturing nonlinear effects of exogenous variables. Which of the explanatory variables enter the model parametrically or non-parametrically may depend on theoretical priors or can be suggested by the results of model specification tests (Kneib et al. 2009).  $f_4(\mathbf{x}_1) \mathbf{z}$  is a varying coefficient term, where  $\mathbf{z}$  is either a continuous or a binary covariate. The term  $h(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$  is a smooth spatial trend surface, i.e. a smooth interaction between latitude and longitude. It allows us to control for unobserved spatial heterogeneity, which is a primary task when dealing with spatial data. When the term  $h(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$  is interacted with one of the explanatory variables (e.g.,  $h(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) \mathbf{x}_1$ ), it allows us to estimate spatially varying coefficients (like in the *GWR* model). Finally,  $\varepsilon$  are *iid* normally distributed random shocks.<sup>5</sup>

This model reflects the notion of spatial dependence made of two parts: (i) a spatial trend due to unobserved regional characteristics, which is modeled by the smooth function of the coordinates, and (ii) global spatial spillover effects, which are modeled by including the spatial lag of the dependent variable. The introduction of the spatial lags of the exogenous ( $\mathbf{X}$ ) variables results in what can be called the Penalized-Spline Geoadditive Spatial Durbin Model (PS-SDM).

When the  $\rho$  parameter is not statistically different from zero, i.e. in the case of a simpler semiparametric geoadditive model without the spatial lag of the dependent variable (*PS* model), if all regressors are manipulated independently of the errors,  $\hat{f}_k(x_k)$  can be interpreted as the conditional expectation of  $y$  given  $x_k$  (net of the effect of the other regressors). Blundell and Powell (2003) use the term Average Structural Function (ASF) with reference to this function. Instead, when  $\rho$  is different from zero, the estimated smooth functions —  $\hat{f}_k(x_k)$  — cannot be interpreted as ASF. Taking advantage of the results obtained for parametric SAR, we can compute the total smooth effect (total-ASF) of  $x_k$  as

$$\hat{f}_k^T(x_k) = \Sigma_q [\mathbf{I}_n - \hat{\rho} \mathbf{W}_n]_{ij}^{-1} b_{kq}(x_k) \hat{\beta}_{kq}
 \tag{29}$$

<sup>5</sup> This assumption can be relaxed by a more general specification, such as  $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 \Lambda)$  being  $\Lambda$  a covariance matrix reflecting cross-sectional dependence in the errors as, for example, in Pinheiro and Bates (2000).

where  $b_{kq}(x_k)$  are P-spline basis functions, and  $\widehat{\beta}_{kq}$  the corresponding estimated parameters.

We can also compute direct and indirect (or spillover) effects of smooth terms in the PS-SAR case as:

$$\widehat{f}_k^D(x_k) = \Sigma_q [\mathbf{I}_n - \widehat{\rho} \mathbf{W}_n]_{ii}^{-1} b_{kq}(x_k) \widehat{\beta}_{kq} \tag{30}$$

$$\widehat{f}_k^I(x_k) = \widehat{f}_k^T(x_k) - \widehat{f}_k^D(x_k) \tag{31}$$

Similar expressions can be provided for the direct, indirect and total effects of the PS-SDM (Table 3).

**Table 3.** Total, direct, and indirect smooth effects

Model	Direct Smooth Effect: $\widehat{f}_k^{de_k}(x_k)$
PS model	$b_{kq}(x_k) \widehat{\beta}_{kq}$
PS-SLM	$\Sigma_q [\mathbf{I}_N - \widehat{\rho} \mathbf{W}_N]_{ii}^{-1} b_{kq}(x_k) \widehat{\beta}_{kq}$
PS-SEM	$b_{kq}(x_k) \widehat{\beta}_{kq}$
PS-SDM	$\Sigma_q [\mathbf{I}_N - \widehat{\rho} \mathbf{W}_N]_{ii}^{-1} [b_{kq}(x_k) \widehat{\beta}_{kq} + b_{kg}(\mathbf{W}_N x_k) \widehat{\beta}_{kg}]$
PS-SLX	$b_{kq}(x_k) \widehat{\beta}_{kq}$
Model	Indirect Smooth Effect: $\widehat{f}_k^{ie_k}(x_k)$
PS model	0
PS-SLM	$\widehat{f}_k^{te_k}(x_k) - \widehat{f}_k^{de_k}(x_k)$
PS-SEM	0
PS-SDM	$\widehat{f}_k^{te_k}(x_k) - \widehat{f}_k^{de_k}(x_k)$
PS-SLX	$b_{kg}(\mathbf{W}_N x_k) \widehat{\beta}_{kg}$
Model	Total Smooth Effect: $\widehat{f}_k^{te_k}(x_k)$
PS model	$b_{kq}(x_k) \widehat{\beta}_{kq}$
PS-SLM	$\Sigma_q [\mathbf{I}_N - \widehat{\rho} \mathbf{W}_N]_{ij}^{-1} b_{kq}(x_k) \widehat{\beta}_{kq}$
PS-SEM	$b_{kq}(x_k) \widehat{\beta}_{kq}$
PS-SDM	$\Sigma_q [\mathbf{I}_N - \widehat{\rho} \mathbf{W}_N]_{ij}^{-1} [b_{kq}(x_k) \widehat{\beta}_{kq} + b_{kg}(\mathbf{W}_N x_k) \widehat{\beta}_{kg}]$
PS-SLX	$b_{kq}(x_k) \widehat{\beta}_{kq} + b_{kg}(\mathbf{W}_N x_k) \widehat{\beta}_{kg}$

The Spatial Error Geoadditive Model (PS-SEM) proposed by Mínguez et al. (2012) augments the PS model by including a spatial autoregressive error term, while leaving the systematic part unchanged:

$$\begin{aligned} \mathbf{y} &= \mathbf{X}^* \beta^* + f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + f_3(\mathbf{x}_3, \mathbf{x}_4) \\ &\quad + f_4(\mathbf{x}_1) \mathbf{z} + \dots + h(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + \mathbf{u} \\ \mathbf{u} &= \lambda \mathbf{W} \mathbf{u} + \varepsilon \quad \varepsilon \sim iid \mathcal{N}(0, \sigma_\varepsilon^2) \end{aligned} \tag{32}$$

where  $\lambda$  is a spatial autoregressive parameter. As in the case of the pure PS model, if all regressors are exogenous,  $\widehat{f}_k(x_k) = \sum_q b_{kq}(x_k) \widehat{\beta}_{kq}$  can be directly interpreted as the conditional expectation of  $y$  given  $x_k$  (ASF).

Getting back to the example of the knowledge production function, the PS-SAR counterpart of model (15) for a short panel data can be for example specified as:

$$\ln K_{it} = \alpha + f(\ln R\&D_{it}, \ln H_{it}) + \rho \sum_{j \neq i} w_{ij} \ln K_{jt} + h(x_{s_1,i}, x_{s_2,i}) + \varepsilon_{it} \quad (33)$$

The nonparametric part of model 33 relaxes the standard assumptions of linearity and additivity regarding the effect of R&D and human capital. Charlot et al. (2015) use a similar specification to analyze the genesis of innovation in the regions of the European Union. Their results unveil nonlinearities, threshold effects, complex interactions and shadow effects that cannot be uncovered by standard parametric formulations.

### 3.3 Modeling Spatial Spillovers, Spatial Heterogeneity, Nonlinearities and Time-Related Factors: Spatio-Temporal Semiparametric Autoregressive Models for Large Panel Data

In this section we propose a class of spatio-temporal models for large spatial panel data which represent a generalization of the *Spatial Autoregressive Semiparametric Geoaddivitive Models* discussed in Sect. 3.2. They are a flexible alternative to the parametric models presented in Sect. 2.3 for modeling spatial panel data as long as the spatio-temporal heterogeneity is smoothly distributed (a very common case, one may say, in empirical economic analyses), so that we can approximate it with smooth nonparametric functions.

The general model proposed is written as:

$$\mathbf{y} = \widetilde{f}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t) + \rho \mathbf{W}\mathbf{y} + \sum_{\delta=1}^k g_{\delta}(\mathbf{x}_{\delta}) + \varepsilon \quad (34)$$

where  $\widetilde{f}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t)$  is a smooth spatio-temporal trend, i.e. a three-dimensional smooth function of the spatial coordinates  $(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$ , and of the time component  $\mathbf{x}_t$ ;  $g_{\delta}(\cdot)$ ,  $\delta = 1, \dots, k$ , are also smooth functions of the covariates  $x_{\delta,it}$  (they can be linear, or can accommodate varying coefficient terms, smooth interactions between covariates, factor-by-smooth curves, and so on);  $\mathbf{W}$  is the spatial weights matrix,  $\rho$  the spatial spillover parameter, and  $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$  where  $\mathbf{R}$  can be multiple of the identity (if errors are independent), or include a temporal correlation structure.

In many situations the spatio-temporal trend to be estimated by  $\widetilde{f}$  can be complex, and the use of a multidimensional smooth function may not be flexible enough to capture the structure in the data. To solve this problem, Lee and Durbán (2011) proposed an ANOVA-type decomposition of  $\widetilde{f}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t)$  where spatial and temporal main effects, and second- and third-order interactions between them can be identified:

$$\begin{aligned} \widetilde{f}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t) = & f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_t(\mathbf{x}_t) + f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) \\ & + f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_t) + f_{2,3}(\mathbf{x}_{s_2}, \mathbf{x}_t) + f_{1,2,3}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t) \end{aligned}$$

Thus, model (34) can be written as:

$$\begin{aligned} \mathbf{y} = & f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_t(\mathbf{x}_t) + f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_t) \\ & + f_{2,3}(\mathbf{x}_{s_2}, \mathbf{x}_t) + f_{1,2,3}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t) + \rho \mathbf{W}_N \mathbf{y} + \sum_{\delta=1}^k g_{\delta}(\mathbf{x}_{\delta}) + \varepsilon \end{aligned} \quad (35)$$

We will refer to it as the PS-ANOVA-SAR(AR1) model. It is flexible enough to simultaneously control for different sources of bias: *spatial heterogeneity bias*, *spatial dependence bias*, *omitted-time related factors bias*, and *functional form bias*.

First, as already pointed out in Basile et al. (2014), the geoaddivitive terms given by  $f_1(\mathbf{x}_{s_1})$ ,  $f_2(\mathbf{x}_{s_2})$  and  $f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$  work as control functions to filter the spatial trend out of the residuals, and transfer it to the mean response in a model specification. Thus, they allow to capture the shape of the spatial distribution of  $\mathbf{y}$ , eventually conditional on the determinants included in the model. These control functions also isolate stochastic spatial dependence in the residuals, that is spatially autocorrelated unobserved heterogeneity. Thus, they can be regarded as an alternative to individual regional dummies to capture unobserved heterogeneity as long as the latter is smoothly distributed over space. Regional dummies peak significantly higher and lower levels of the mean response variable. If these peaks are smoothly distributed over a two-dimensional surface (i.e., if unobserved heterogeneity is spatially autocorrelated), the smooth spatial trend is able to capture them.

Second, the smooth time trend,  $f_t(\mathbf{x}_t)$ , and the smooth interactions between space and time -  $f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_t)$ ,  $f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_t)$ , and  $f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t)$  - work as control functions to capture the heterogeneous effect of common shocks. Thus, the PS-ANOVA-SAR model works as an alternative to the models proposed by Bai and Li (2015); Bailey et al. (2016); Pesaran and Tosetti (2011); Shi and Lee (2016); Vega and Elhorst (2016) based on extensions of common factor models to accommodate both strong cross-sectional dependence (through the estimation of the spatio-temporal trend) and weak cross-sectional dependence (through the estimation of the  $\rho$  parameter). The advantage of the PS-ANOVA-SAR model lies in the fact that its ability to fully control for the residual cross-sectional dependence and to assess the presence of network effects net of common factor effects, is not crucially affected by the set of covariates included in the model.

Furthermore, this framework is also flexible enough to control for the linear and nonlinear functional relationships between the dependent variable and the covariates.

Getting back to the example of the knowledge production function, the PS-ANOVA-SAR version of (33) for a panel data with a long time series would be:

$$\begin{aligned} \ln K_{it} = & f(\ln R \& D_{it}, \ln H_{it}) + \rho \sum_{j \neq i} w_{ij} \ln K_{jt} \\ & + f_1(x_{s_1,i}) + f_2(x_{s_2,i}) + f_t(x_t) + f_{1,2}(x_{s_1,i}, x_{s_2,i}) + f_{1,t}(x_{s_1,i}, x_t) \\ & + f_{2,3}(x_{s_2,i}, x_t) + f_{1,2,3}(x_{s_1,i}, x_{s_2,i}, x_t) + \varepsilon_{it} \end{aligned} \quad (36)$$

## 4 Software

Nowadays there is a wide range of software allowing to estimate most of the econometric models exposed in this Chapter. Some of them, like *GeoDa* (Anselin et al. 2006),

use a menu interface which permits the user to perform spatial exploratory analysis, and to estimate parametric spatial econometric models for cross-sectional data without the need to learn new commands. Nevertheless, other well-known software alternatives require some skills in the corresponding programming language to deal with the spatial data. This is the case of some specialized packages in *R* (R Core Team 2016), the library *PySAL* (Rey and Anselin 2007) written in *Python* (Van Rossum 1995), the toolbox for spatial econometric models written by LeSage (2009) in *MATLAB* (MATLAB 2017), some functions, also in *MATLAB*, to estimate static and dynamic spatial panel data models developed by Elhorst (Elhorst et al. 2013), and a suite of commands for spatial data in *SAS* (SAS Institute Inc. 2013) or *Stata* (StataCorp. 2015). Bivand and Piras (2015) compare the results obtained by using different software alternatives and conclude that all of them provide similar results.

In this overview we focus on *R*, for the following reasons<sup>6</sup>:

- it is a well-tested free software with a growing number of packages in all statistical fields (spatial analysis included);
- it has a huge community of users;
- the possibility to combine functional programming with object-oriented programming (Chambers 2016) allows the developers to build new packages making use of the existing ones;
- it allows to estimate most of the spatial econometric models exposed in this chapter including both parametric models (for cross-sectional and static panel data) and semiparametric models.

The *R* packages **spdep** (Bivand 2013) and **sp** (Bivand et al. 2013; Pebesma and Bivand 2005) facilitate the creation, transformation and manipulation of spatial objects, neighborhood matrices and the computation of descriptive measures of spatial autocorrelation. Moreover, the package **spdep** allows researchers to estimate the whole set of cross-sectional spatial autoregressive models exposed in Sect. 2.1 including SAR, SEM, SDM, SLX and SAC models using either ML or GMM estimation in an efficient way. Furthermore, this package also permits us to compute the marginal effects and make inference on their values. To extend the range of standard spatial models considered, Piras (2010) created the **sphet** package for estimating and testing parametric spatial models with heteroskedastic innovations using estimation procedures based on GMM.

To deal with the static spatial panel data models discussed in Sect. 2.2.1, Millo and Piras (2012) have developed the **splm** package. It includes a set of functions able to estimate a full range of static spatial panel data models including fixed or random effects; spatial lags for the error term or dependent variable and, possibly, serial correlation in the noise of the model. Millo (2014) provides an extensive overview of these models including algorithms to estimate them using MLE. These packages can also be used to estimate the SAR-CCEP model discussed in Sect. 2.3. Unfortunately, there is not a freely available *R* package for the estimation and inference of dynamic spatial panel data models, revised in Sect. 2.2.2, while some functions are available in *MATLAB* (Elhorst 2014a; Elhorst et al. 2013).

<sup>6</sup> Two recent references of the use of *R* for spatial statistical and econometric analysis are Arbia (2014); Brunsdon and Comber (2015). A more classical reference is given by Bivand et al. (2013).

Focusing on semiparametric spatial data models (Sect. 3), McMillen (2013) has written the **McSpatial** package which includes routines to estimate nonparametric and conditionally parametric versions of spatial linear regression and spatial models with binary dependent variable. It mainly uses kernel techniques to perform the nonparametric estimations. Moreover, the package **GWmodel** (Gollini et al. 2015; Lu et al. 2014) deals with geographical weighted (GW) models, and includes functions for computation of GW summary statistics and regression, GW principal components analysis, and GW discriminant analysis. The techniques to estimate MGWR-SAR models discussed in Sect. 3.1 are already included in the forthcoming *R* package **gwrsar** (Geniaux and Martinetti 2017).

Finally, considering semiparametric regression models that include spatial or spatio-temporal trends, both packages **mgcv** (Wood 2006) and **R2BayesX** (Belitz et al. 2016; Umlauf et al. 2015) include some functions to estimate models including complex spatial and spatio-temporal trends, parametric and non-parametric covariates and interactions between them. Both packages have the possibility to choose P-spline methodology or the combination of other type of spline bases with penalty matrices for the non-parametric terms. The full class of models are usually estimated either by restricted maximum likelihood (REML) or bayesian methods. The techniques to estimate PS-SAR and PS-ANOVA-SAR models (Mínguez et al. 2017) discussed in Sect. 3.2 will also be included in a forthcoming *R* package.

## 5 Conclusions

Spatial econometrics is commonly conceived as a powerful method for capturing spatial spillover (or spatial interaction) effects. It is based on the assumption that, when an idiosyncratic shock hits a specific spatial unit (a country, a region, a firm, etc.), then its effects propagate to all other spatial units in the sample with a distance-decay mechanism. For example, in estimating a regional knowledge production function using a simple cross-section of regional data, we must be able to assess the impact of the investment in R&D in a region on both its own productivity outcome (*TFP*) and on the outcome of all other regions in the sample. Spatial econometricians have also derived statistical measures of direct and indirect (spillover) marginal impacts to quantify this phenomenon (LeSage and Pace 2009).

Nevertheless, is also important to recognize that the evidence of spatial spillovers might (at least partially) mask other specification errors, such as wrong functional form, unobserved spatial heterogeneity, heteroskedasticity, unobserved common factors, time persistence, and so on. Without a proper control for these sources of bias, the estimated spatial spillover effect often appears very (unrealistically) strong. For example, in estimating a regional knowledge production function using a simple cross-section of regional data without any control for nonlinearities and spatial unobserved heterogeneity, one may find evidence of an average indirect (spillover) impact of R&D on *TFP* similar to the corresponding average direct marginal effect. This is obviously unreasonable.

In this Chapter we have reviewed different parametric and semiparametric approaches recently developed to mitigate this problem. Not surprisingly, parametric spatial panel models received most attention in the literature. In particular, dynamic spatial panel data models and spatial panel autoregressive models with common factors turn

to be very important tools for simultaneously control for spatial spillovers, unobserved spatial heterogeneity, unobserved common factor and time persistence. However, in the Chapter we have also pointed out that spatial autoregressive semiparametric geosadditive models (PS-SAR models; Basile et al. 2014) may play a prominent role in those context in which the theory suggests the existence of spatial interdependence and heterogeneous behavior of the spatial units. These methods represent indeed some flexible approaches which are able to address simultaneously spatial dependence, heterogeneity and nonlinearity. Moreover, we have reviewed more recently developed semiparametric models for longitudinal data including a non-parametric spatio-temporal trend, a spatial lag of the dependent variable, and a time series autoregressive noise (PS-ANOVA-SAR-AR1) which represent a valid alternative to parametric methods aimed at disentangling strong and weak cross-sectional dependence (Mínguez et al. 2017). Natural directions in which these methods can be extended are a specification for a dynamic framework.

**Acknowledgements.** The work of Román Mínguez was supported by the research project MTM2014-52184 and ECO2015-65826-P from the Spanish Ministry of Economy and Competitiveness.

## References

- Anselin, L.: *Spatial Econometrics: Methods and Models*, vol. 4. Kluwer Academic Pub., Dordrecht (1988)
- Anselin, L., Syabri, I., Kho, Y.: *GeoDa: an introduction to spatial data analysis*. *Geograph. Anal.* **38**(1), 5–22 (2006)
- Arbia, G.: *A primer for spatial econometrics. With Applications in R*. Palgrave Texts in Econometrics: Palgrave Macmillan UK, UK (2014)
- Bai, J., Li, K.: *Dynamic Spatial Panel Data Models with Common Shocks*. Manuscript: Columbia University (2015)
- Bailey, N., Holly, S., Pesaran, M.H.: A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *J. Appl. Econom.* **31**(1), 249–280 (2016)
- Basile, R., Durbán, M., Mínguez, R., Montero, J.M., Mur, J.: Modeling regional economic dynamics: spatial dependence, spatial heterogeneity and nonlinearities. *J. Econ. Dyn. Control* **48**, 229–245 (2014)
- Belitz, C., Brezger, A., Kneib, T., Lang, S., Umlauf, N.: *BayesX: Software for Bayesian Inference in Structured Additive Regression Models*. Version 1.1 (2016)
- Bivand, R.: *spdep: spatial dependence, weighting schemes, statistics and models*. R package version 0.5-56 (2013). <http://CRAN.R-project.org/package=spdep>
- Bivand, R., Piras, G.: Comparing implementations of estimation methods for spatial econometrics. *J. Stat. Softw.* **63**(1), 1–36 (2015)
- Bivand, R.S., Pebesma, E., Gómez-Rubio, V.: *Applied Spatial Data Analysis with R*. Use R!, 2nd edn. Springer, New York (2013)
- Blundell, R., Bond, S.: Initial conditions and moment restrictions in dynamic panel data models. *J. econ.* **87**(1), 115–143 (1998)
- Blundell, R., Powell, J.: Endogeneity in nonparametric and semiparametric regression models. In: *Advances in Economics and Econometrics Theory and Application*. Cambridge University Press (2003)
- Brunsdon, C., Comber, L.: *An Introduction to R for Spatial Analysis and Mapping*. SAGE Publications Ltd., London (2015)

- Brunsdon, C., Fotheringham, A.S., Charlton, M.E.: Geographically weighted regression: a method for exploring spatial nonstationarity. *Geograph. Anal.* **28**(4), 281–298 (1996)
- Chambers, J.M.: *Extending R. The R Series*. Chapman and Hall/CRC, Boca Raton (2016)
- Charlot, S., Crescenzi, R., Musolesi, A.: Econometric modelling of the regional knowledge production function in Europe. *J. Econ. Geogr.* **15**(6), 1227–1259 (2015)
- Cho, S.-H., Lambert, D.M., Roberts, R.K., Kim, S.G.: Moderating urban sprawl: is there a balance between shared open space and housing parcel size? *J. Econ. Geogr.* **10**(5), 763–783 (2010)
- Chudik, A., Pesaran, M.H., Tosetti, E.: Weak and strong cross-section dependence and estimation of large panels. *Econom. J.* **14**(1), 45–90 (2011)
- Debarsy, N., Ertur, C., LeSage, J.P.: Interpreting dynamic space-time panel data models. *Stat. Methodol.* **9**(1), 158–171 (2012)
- Eilers, P.H., Marx, B.D., Durbán, M.: Twenty years of P-splines. *SORT-Stat. Oper. Res. Trans.* **39**(2), 149–186 (2015)
- Elhorst, J.: Matlab software for spatial panels. *Int. Reg. Sci. Rev.* **37**(3), 389–405 (2014a)
- Elhorst, J.: *Spatial Econometrics. From Cross-Sectional Data to Spatial Panels*. SpringerBriefs in Regional Science. Springer, New York (2014b)
- Elhorst, J., Zandberg, E., de Haan, J.: The impact of interaction effects among neighbouring countries on financial liberalization and reform: a dynamic spatial panel data approach. *Spat. Econ. Anal.* **8**(3), 293–313 (2013)
- Ertur, C., Musolesi, A.: Weak and strong cross-sectional dependence: a panel data analysis of international technology diffusion. *J. Appl. Econom.* **32**, 477–503 (2016)
- Geniaux, G., Martinetti, D.: A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Reg. Sci. Urban Econ.* (2017). Forthcoming
- Gollini, I., Lu, B., Charlton, M., Brunsdon, C., Harris, P.: GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *J. Stat. Softw.* **63**(17), 1–50 (2015)
- Kang, D., Dallerba, S.: Exploring the spatially varying innovation capacity of the US counties in the framework of Griliches knowledge production function: a mixed GWR approach. *J. Geogr. Syst.* **18**(2), 125–157 (2016)
- Kelejian, H., Prucha, I.: Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem. *Int. Reg. Sci. Rev.* **20**, 103–111 (1997)
- Kelejian, H., Prucha, I.: On the asymptotic distribution of moran I test statistic with applications. *J. Econom.* **104**, 219–257 (2001)
- Kneib, T., Hothorn, T., Tutz, G.: Variable selection and model choice in geoadditive regression models. *Biometrics* **65**(2), 626–634 (2009)
- Kukenova, M., Monteiro, J.A.: Spatial dynamic panel model and system GMM: a Monte Carlo investigation (2008). Unpublished article
- Lee, D., Durbán, M.: P-spline ANOVA type interaction models for spatio-temporal smoothing. *Stat. Modell.* **11**, 49–69 (2011)
- Lee, L.: Asymptotic distribution of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica* **72**, 1899–1926 (2004)
- Lee, L.-F., Yu, J.: Estimation of spatial autoregressive panel data models with fixed effects. *J. Econom.* **154**(2), 165–185 (2010a)
- Lee, L.-F., Yu, J.: A spatial dynamic panel data model with both time and individual fixed effects. *Econom. Theory* **26**(02), 564–597 (2010b)
- LeSage, J., Pace, K.: *Introduction to Spatial Econometrics*. CRC Press, Boca Raton (2009)
- LeSage, J.P.: *Econometrics Toolbox for MATLAB* (2009). [www.spatial-econometrics.com](http://www.spatial-econometrics.com)
- Liu, X., Lee, L., Bollinger, C.: An efficient GMM of spatial autoregressive models. *J. Econom.* **159**, 303–319 (2007)



- Lu, B., Harris, P., Charlton, M., Brunsdon, C.: The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-Spat. Inf. Sci.* **17**(2), 85–101 (2014)
- MATLAB: version 9.2 (R2017a), Natick, Massachusetts (2017)
- McMillen, D.: McSpatial: nonparametric spatial data analysis. R package version 2.0 (2013)
- McMillen, D.P.: Perspectives on spatial econometrics: linear smoothing with structured models. *J. Reg. Sci.* **52**(2), 192–209 (2012)
- Millo, G.: Maximum likelihood estimation of spatially and serially correlated panels with random effects. *Comput. Stat. Data Anal.* **71**, 914–933 (2014)
- Millo, G., Piras, G.: splm: spatial panel data models in R. *J. Stat. Softw.* **47**(1), 1–38 (2012)
- Mínguez, R., Basile, R., Durbán, M.: Spatio-temporal autoregressive semiparametric models for the analysis of regional economic data (2017)
- Mínguez, R., Durbán, M., Montero, J., Lee, D.: Competing Spatial Parametric and Non-Parametric Specifications. Mimeo, New York (2012)
- Montero, J., Mínguez, R., Durbán, M.: SAR models with nonparametric spatial trends. A P-spline approach. *Estadística Española* **54**(177), 89–111 (2012)
- Pace, R.K., LeSage, J.P.: Spatial autoregressive local estimation. *Spat. Econom. Spat. Stat.*, 31–51 (2004)
- Páez, A., Uchida, T., Miyamoto, K.: A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environ. Plann. A* **34**(4), 733–754 (2002)
- Pebesma, E.J., Bivand, R.S.: Classes and methods for spatial data in R. *R News* **5**(2), 9–13 (2005)
- Pesaran, M.H.: Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74**(4), 967–1012 (2006)
- Pesaran, M.H., Tosetti, E.: Large panels with common factors and spatial correlation. *J. Econom.* **161**(2), 182–202 (2011)
- Pinheiro, J., Bates, D.: *Mixed-Effects Models in S and S-PLUS*. Springer, New York (2000)
- Piras, G.: sphet: spatial models with heteroskedastic innovations in R. *J. Stat. Softw.* **35**(1), 1–21 (2010)
- R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2016)
- Rey, S.J., Anselin, L.: PySAL: a python library of spatial analytical methods. *Rev. Reg. Stud.* **37**(1), 5–27 (2007)
- SAS Institute Inc.: *SAS/IML(R) 13.1 User's Guide*. SAS Institute Inc., Cary (2013)
- Shi, W., Lee, L.: *A Spatial Panel Data Model with Time Varying Endogenous Weights Matrices and Common Factors* (2016)
- StataCorp.: *Stata Statistical Software: Release 14* College Station. StataCorp LP, TX (2015)
- Umlauf, N., Adler, D., Kneib, T., Lang, S., Zeileis, A.: Structured additive regression models: an R interface to BayesX. *J. Stat. Softw.* **63**(21), 1–46 (2015)
- Van Rossum, G.: *Python Reference Manual*. CWI report CS-R9524 (1995)
- Vega, S.H., Elhorst, J.P.: A regional unemployment model simultaneously accounting for serial dynamics, spatial dependence and common factors. *Reg. Sci. Urban Econ.* **60**, 85–95 (2016)
- Wood, S.: *Generalized Additive Models. An Introduction with R*. Chapman and Hall, London (2006)
- Yu, J., De Jong, R., Lee, L.-F.: Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both  $n$  and  $T$  are large. *J. Econom.* **146**(1), 118–134 (2008)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

