Modeling regional economic dynamics: spatial dependence, spatial heterogeneity and nonlinearities

Roberto Basile, Saime S. Kayam, Román Mínguez, Jose María Montero and Jesús Mur

1 Introduction

Modeling regional economic dynamics requires the adoption of complex econometric tools, which allow us to deal with some important methodological issues, such as spatial dependence, spatial heterogeneity and nonlinearities. Regional and urban growth theories provide good examples to illustrate these issues. Recent developments in economic growth theory have proposed extensions of multi-region growth models (see, i.a., [18]) that include technological interdependence across regions, in order to consider neighborhood effects (*spatial dependence*) in growth and convergence processes. Relaxing the strong homogeneity assumptions on the cross-region growth process and starting from the more realistic hypothesis that different economies should be described by distinct production functions, other contributions to the economic growth theory have cited the issue of *nonlinearities* (parameter heterogeneity) to explain club convergence (see, i.a., [14]). Nonlinearities and spatial dependence have also been addressed in the literature on the effect of agglomeration

Roberto Basile
Second University of Naples, Department of Economics, Corso Gran Priorato di Malta, 1 - 81043, Capua (CE), Italy, e-mail: roberto.basile@unina2.it

Saime S. Kayam
Istanbul Technical University, Faculty of Management, Macka, Istanbul, Turkey e-mail: kayams@itu.edu.tr

Román Mínguez
University of Castilla-La Mancha, Statistics Department. Cuenca, Spain e-mail: roman.minguez@uclm.es

Jose María Montero
University of Castilla-La Mancha, Statistics Department. Toledo, Spain e-mail: jose.mlorenzo@uclm.es

Jesús Mur
University of Zaragoza, Department of Economic Analysis, Spain e-mail: jmur@unizar.es

economies on local (urban) economic growth. For example, a hump-shaped relationship between economic density and local productivity (or employment) growth has been highlighted by [8]: the positive effect of agglomeration externalities fades as the density of economic activities reaches some threshold value, after which negative effects due to congestion costs prevail. In modeling the effects of agglomeration economies on local economic growth, it is important to recognize that externalities may overcome the administrative boundaries of the regions, thus generating spatial dependence. Finally, as pointed out by [37], the marked unevenness of local economic development can be partly justified on the basis that space is not uniform (*spatial heterogeneity*): "first nature" characteristics of local areas (i.e. unobserved spatial heterogeneity) must be carefully controlled for when specifying a local economic growth model, especially when these unobservables are potential sources of endogeneity.

Many other examples in regional science and economic geography may be used to discuss the issues of nonlinearities, spatial dependence and spatial heterogeneity. Here, we want to point out that applied econometric studies rarely tackle all these issues simultaneously. Most studies focus on one of these matters, disregarding the interdependence between them. For example, some scholars focus on detecting the presence of spatial dependence while assuming a linear functional form for the data generating process ([57]; [51], others only try to assess the existence of nonlinear effects ([20]), some others only control for spatial heterogeneity in a panel framework ([30]). However, nonlinearities, spatial dependence and spatial heterogeneity are not orthogonal issues and disregarding one of them may generate some biases. For example, [46] shows that specification tests may indicate spatial autocorrelation when functional form misspecification is actually the only problem with the model. Thus, incorrect functional forms and omitted variables that are correlated over space produce spurious spatial autocorrelation. [7] have also provided evidence of a trade-off between spatial autocorrelation and nonlinearities: the value of the spatial autocorrelation parameter is lower when possible nonlinearities are taken into account.

Recent developments in the spatial econometrics literature have provided some instruments (such as *Spatial Autoregressive Semiparametric Geoadditive Models*), which address the three issues simultaneously and, therefore, are of great use for practitioners ([26]; [5]; [62]; [61]; [47]; [48]. In this paper we describe some of these methodological contributions and present some applications in the field of regional science and economic geography. We start by briefly reviewing the broad literature on parametric spatial econometric models and raising some critical issues concerning these models (Sect. 2). Then, we introduce semiparametric geoadditive models and describe their potential (Sect. 3). Specifically, we describe a control function approach to estimate a spatial lag semiparametric geoadditive model. In Sect. 4 we discuss an alternative way to specify (spatial lag) semiparametric geoadditive models as mixed models. In Sect. 5 we present selected empirical works using these models in the field of regional science and economic geography in order to show the wide scope for applications of this approach. In particular, we present an application to regional growth. Concluding remarks are reported in Sect. 6.

2 Parametric Spatial Econometric Models

Regional economics works within a spatial realm, and that means heterogeneity (i.e., heteroskedasticity and spatial instability of parameters) and interdependence (i.e., spatial dependence). Spatial econometrics obviously deals with these two topics, predominantly under a parametric approach. This section discusses some of the key aspects of this strategy of building regional models and highlights some of its limits.

Let y_i be the response variable computed for spatial unit i and $\left\{x_{k,i}; k = 1,2,...K\right\}$ a set of K explanatory variables. For example, y_i may measure the productivity growth rate of region i computed for a sufficiently long time span, whereas the x variables may include different indices of human capital and investment effort computed for the same time span. A key element in current spatial econometrics is the Spatial Durbin Model (SDM):

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{k,i} + \rho \sum_{j=1}^{n} w_{ij} y_j + \sum_{k=1}^{K} \theta_k \sum_{j=1}^{n} w_{ij} x_{k,j} + \varepsilon_i \tag{1}$$

where ε_i is a white noise normally distributed error term, $\varepsilon_i \sim iid\mathcal{N}(0;\sigma_\varepsilon^2)$. This equation includes a sequence of (spatial linear) weights for the purpose of measuring the influence received by region i from region j: $\left\{w_{ij}; j = 1,2,...n\right\}$. Using these weights we can obtain spatial lags of the variables of interest. Using matrix notation, we get:

$$\mathbf{y} = \mathbf{X}\beta + \rho \mathbf{W}_n \mathbf{y} + \mathbf{W}_n \mathbf{X}\theta + \varepsilon \tag{2}$$

where $\mathbf{y}$ is the $(n \times 1)$ vector of observations of the explained variable on the n regions, $\mathbf{X}$ is a $(n \times K)$ matrix of observations of the explanatory variables on the same regions and ε is a $(n \times 1)$ vector of error terms. Moreover, $\mathbf{W}_n$ is a $(n \times n)$ weighting matrix, assumed to be the same in the spatial lag of $\mathbf{y}$ and in the lags of the xs variables (the assumption can be relaxed).

Assuming $\theta = \mathbf{0}$ in equation (2), a Spatial Autoregressive Model (SAR) is obtained; a Spatial Lag of $\mathbf{X}$ Model (SLX) results from the restriction that $\rho = 0$; and a Spatial Error Model (SEM) appears under the assumption that $\theta - \rho\beta = \mathbf{0}$; the pure non-spatial model would occur if $\rho = \theta = 0$. The puzzle may also include a SDM or a SAR model with spatially autocorrelated errors, $\varepsilon = \varphi \mathbf{W}_n \varepsilon + \eta$, η being a white noise random vector. Other specifications can be found in [40].

The term $\mathbf{W}_n\mathbf{y}$ that appears in the right–hand side (*rhs* from now on) of (2) is correlated with the error term, $Cov[\mathbf{W}_n\mathbf{y};\varepsilon] \neq \mathbf{0}$, 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), if the assumption of normality cannot be maintained ([39]). Two–Stage Least Squares (2SLS) estimates adapt well to the case of (2) because higher orders of spatial lags of the x variables are natural candidates to be used as instrumental variables ([33]). A more efficient estimator is the method of moments estimator (MM) ([34]). [39] generalized the MM approach into a fully generalized method of

moments (GMM) estimator for the case of the SDM model (2), while [41] 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 GMM estimator allows the researcher to take into account any endogeneity problems in the *r.h.s.*, different from the spatial lag of $\mathbf{y}$.

The above–mentioned spatial econometric models allow for interdependence among regions. All of them correspond to a long-run equilibrium relation between the response variable and its covariates; time dynamics is ruled out. In response to an exogenous variation in a x variable (such as, e.g., an improvement in the human capital stock in Southern Italy's provinces), these spatial models return the expected impact on the dependent variable (e.g., the productivity growth rates) for the whole regional system in the steady state solution.[1] It is customary to distinguish between *local* and *global* spatial spillovers ([1]). The key is the existence of feedback effects in the equation. Model (2) contains feedback effects: a change in a regressor in region i impacts on the outcome of this region, on the outcome of its neighbors, on that of the neighbors of its neighbors and so on. The impact therefore is *global*. On the contrary, the multipliers obtained from a SLX model ($\mathbf{y} = \mathbf{X}\beta + \mathbf{W}_n\mathbf{X}\theta + \varepsilon$) are *local*, since the impact of the change dies just after its effect on the neighbors.

When spatial spillovers are global, we may distinguish between *direct* and *indirect* spatial effects. *Direct* effects measure the impact of a change in regressor k in region i on the outcome of the same region: $\frac{\partial y_i}{\partial x_{ki}}$, while *indirect* effects measure the impact of a change in regressor k in region j on the outcome of region i: $\frac{\partial y_i}{\partial x_{kj}}$. The problem with these effects is that, conditional on the model, they are specific to the pair of regions involved (i, j).[2] For this reason [40] propose the use of average indicators. For example, in the SDM model of (2), the *average direct* effect of variable k is:

$$\overline{de_k} = \sum_{i=1}^{n} \frac{(\mathbf{I}_n - \rho \mathbf{W}_n)^{-1} [\mathbf{I}_n(\beta_k + \theta_k \mathbf{W}_n)]_{ii}}{n};\ k = 1,2,....,K \tag{3}$$

This is the mean value of the n *direct* effects (the sub-index ii means the i-th element of the main diagonal of the square matrix $(\mathbf{I}_n - \rho \mathbf{W}_n)^{-1} [\mathbf{I}_n(\beta_k + \theta_k \mathbf{W}_n)]$. In order to obtain the average *indirect* effect we must first define the average *total* effect:

$$\overline{te_k} = \sum\nolimits_{j=1}^{n} \sum\nolimits_{i=1}^{n} \frac{(\mathbf{I}_n - \rho \mathbf{W}_n)^{-1} [\mathbf{I}_n(\beta_k + \theta_k \mathbf{W}_n)]_{ij}}{n^2};\ k = 1,2,....,K \tag{4}$$

[1] If the interest lies in the short-term adjustments, a spatial panel data model would be required instead ([17]).

[2] Although classical spatial econometric models are based on the assumption of linearity and parameter homogeneity, they allow us to assess a form of heterogeneity, called "interdependence heterogeneity" ([19]): the magnitude of spatial direct and indirect partial effects is different among regions, since it depends upon the position of the regions in space, the degree of connectivity among regions, which is governed by the $\mathbf{W}_n$ matrix and the estimated model parameters.

The *average total* effect measures the accumulated impact of a change in the x_k variable on the dependent variable, $\mathbf{y}$, located in any region. The difference between the two corresponds to the *average indirect* effect:

$$\overline{ie_k} = \overline{te_k} - \overline{de_k};\ k = 1, 2, ..., K \tag{5}$$

and measures the importance of the spatial spillovers in the model.

A very important issue in this context concerns the choice of the weighting matrix whose purpose is to reflect the arrangement of the space. Uncertainty is a big problem here because the researcher must provide this matrix and, normally, he/she has a very limited knowledge about it. Of course, the problem of selecting the right $\mathbf{W}_n$ is reflected on the computation of spatial multipliers: a bad choice of the matrix invalidates subsequent analysis. A few rules can be provided to tackle this issue: $\mathbf{W}_n$ must be a square $(n \times n)$ matrix whose elements are, usually, non-negative and its main diagonal is comprised of zeroes (both restrictions can be relaxed). According to [32], this matrix should be uniformly bounded in absolute value in order to assure convergence and, in order to avoid cases of isolation, the row sums should be uniformly bounded away from zero. It is typical to row-standardize the matrix before its use in estimation algorithms.

Thus far we have sketched some of the key elements of spatial econometrics. Now, an important question remains: Why should econometric models take spatial effects into account? [23] observe that if we are using spatial data it is reasonable to find spatial interaction:

> Put very simply, if we assume that firms are heterogeneous and always interacting with each other, then the fact that they are often located in different regions will cause regions to be heterogeneous and interdependent (p. 178).

This is what intuition tells us and the long list of mis-specification tests usually corroborates ([16]). However, from a purely econometric perspective, we are forced to go a bit further because the indiscriminate use of the spatial lag of the endogenous variable can muffle the impact of other specification errors ([52]). In this sense, [40] point to misspecification problems due to (*i*) mismatches in the time structure of the equation; (*ii*) omission of relevant explanatory variables from the *r.h.s.*; (*iii*) the impact of omitted heterogeneity in the model. [46] adds a fourth factor (*iv*), the consequences of a wrong selection in the functional form of the equation. [35] also observe that heteroskedasticity and spatial dependence produce similar signs and are very often confused. Similarly, spatial dependence tests react strongly under heterogeneity in the parameters ([49]).

These observations advocate for a more theory–driven approach. Examples in this direction are the regional production function with externalities in the rate of technological progress ([42]; [18], [19]), or the Verdoorn Law with knowledge spillovers proposed ([21]). Spatial prices and spatial market competition models have also produced many papers in the same spirit (i.a., [55]; [31]). This practice is not typical, however. In fact, a kind of empiricism controls the process of building a model: first, a simple provisional equation is specified, then a large battery of misspecification tests is applied and, finally, spatial interaction mechanisms are

introduced ad hoc as reaction to the tests. The consequence may be that, at the end, we lose perspective.

The observation that spatial lag terms may actually capture the effect of other specification errors may also suggest the need for a more flexible (semiparametric) approach, which relaxes the restrictive assumptions (linearity and parameter homogeneity) of the parametric approach. One step in this direction is the so-called Geographically Weighted Regression (*GWR*) model ([24]; [43]), which is a non-parametric (local linear) method to capture spatial parameter heterogeneity.[3]

3 Semiparametric Geoadditive Models

In this section, we present a semiparametric framework, which allows us to relax the linearity assumption and simultaneously model spatial dependence and spatial heterogeneity. We start by introducing a general specification of the semiparametric geoadditive model without a spatial lag term (Subsect. 3.1) and discussing technical issues concerning its estimation (Subsect. 3.2 and 3.3). In Subsect. 3.4 we extend this model by introducing the spatial lag of the dependent variable on the *rhs* so as to get a spatial autoregressive semiparametric geoadditive model (*SAR-Geo-AM*).

3.1 Model Specification

In Sect. 2 we pointed out that most of the spatial econometric literature focuses on the issue of spatial dependence. A strand of this literature also stresses the problem of spatial heterogeneity, that is, of spatial instability of the parameters (using the *GWR* method), while the issue of nonlinearity (i.e., the choice of the functional form) is strongly neglected. Obviously, nonlinearities might also be captured within a pure parametric framework by using a polynomial expansion, but this may lead to strong collinearity. Semiparametric methods represent a more satisfactory solution since they are more flexible than any parametric specification. By using a particular version of the semiparametric model that allows for additive components ([28]), we are able to obtain graphical representation of the relationship between the response variable and the covariates. Additivity ensures that the effect of each predictor can be interpreted net of the effects of the other regressors, as in multiple linear regressions.[4]

---

[3] The *GWR* has been extended to cross-sectional models with spatial interaction terms by [50] and [49].

[4] Usually, a fully nonparametric model (i.e., a model where all terms are smoothly interacted with each other) cannot be applied to regional data since it would require a very large number of observations to overcome the curse of dimensionality. Additivity is therefore a valid compromise between flexibility and tractability.

The starting point may be a general form of the semiparametric additive model suitable for large cross-regional data:

$$y_i = \mathbf{X}_i^{*'}\beta^* + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + f_4(x_{1i})\, l_i + ... + \varepsilon_i \tag{6}$$

$$\varepsilon_i \sim iid\mathcal{N}\left(0, \sigma_\varepsilon^2\right) \qquad i = 1, ..., n$$

where y_i is a continuous univariate response variable measuring, for example, the average annual productivity growth rate of region i. $\mathbf{X}_i^{*'}\beta^*$ is the linear predictor for any strictly parametric component (including the intercept, all categorical covariates and eventually some continuous covariates), with β^* being a vector of fixed parameters. $f_k(.)$ 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 (see, e.g., [36]). $f_4(x_{1i})\, l_i$ is a varying coefficient term, where l_i is either a continuous or a binary covariate. For example, we may want to test whether the smooth effect of x_1 (e.g., population density) is different in the North and in the South. In this case l_i is a binary variable taking value one if region i belongs to the North and zero if it belongs to the South. Thus, if $l_i = 0$, the effect of x_1 is given by $f_1(x_{1i})$, whereas for $l_i = 1$, the effect is composed as the sum $f_1(x_{1i}) + f_4(x_{1i})$, and $f_4(x_{1i})$ can be interpreted as the deviation of x_1 for the North. Finally, ε_i are *iid* normally distributed random shocks.

Model (6) captures nonlinearities in the relationship between the response variable y_i and its covariates, but it does not take into account any spatial structure of the data. Removing *unobserved spatial patterns* is a primary task, especially when the researcher considers spatial unobservables as potential sources of endogeneity, that is, when there is a suspected correlation between unobserved and observed variables. This issue can be addressed within the semiparametric framework by incorporating the spatial location as an additional covariate in (6), thus generating what is known in the literature as the geoadditive model:[5]

$$\begin{aligned} y_i = {} & \mathbf{X}_i^{*'}\beta^* + f_1(x_{1i}) + f_2(x_{1i}) + f_3(x_{3i}, x_{4i}) + f_4(x_{1i})\, l_i + ... \\ & + h(no_i, e_i) + \varepsilon_i \end{aligned} \tag{7}$$

The term $h(no_i, e_i)$ in equation (7) is a smooth spatial trend surface, i.e., a smooth interaction between latitude (*northing*) and longitude (*easting*). It allows us to control for unobserved spatial heterogeneity.

When the term $h(no_i, e_i)$ is interacted with one of the explanatory variables, (e.g., $h(no_i, e_i)\, x_{1i}$), it allows us to estimate spatially varying coefficients (like in the *GWR* model). For example, by using this interaction term, we can test the assumption that

[5] Although this model is widely used in environmental studies and in epidemiology (see, i.a., [2]), it is rarely considered for modeling economic data.

the effect of urbanization economies on local productivity in Italy varies moving from the South to the North, or from North–Western to North–Eastern regions.

Finally, equation (7) can be augmented by relaxing the *iid* assumption for the error term, that is assuming an error vector $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 \Lambda)$ with a covariance matrix Λ reflecting spatial error correlation as, for example, in [54].

3.2 Parameter Estimation and Inference

Let us now discuss the issues concerning the estimation of model parameters in equation (7) and the related inference starting from the assumptions of an independent error structure and strict exogeneneity of all explanatory variables. Omitting the subscript i, each *k-th* univariate smooth term in equation (7) can be approximated by a linear combination of known basis functions $b_{q_k}(x_k)$:

$$f_k(x_k) = \sum_{q_k} \beta_{q_k} b_{q_k}(x_k)$$

where β_{q_k} are unknown parameters to be estimated. To reduce mis-specification bias, $q_k^{\prime}s$ must be made fairly large. But this may generate a danger of over-fitting. As we shall clarify further on, by penalizing 'wiggly' functions when fitting the model, the smoothness of the functions can be controlled. Thus, a measure of 'wiggliness' $J \equiv \beta^{\prime} \mathbf{S} \beta$ is associated with each k smooth function, with $\mathbf{S}$ a positive semidefinite matrix. Typically, the quadratic penalty term is equivalent to an integral of squared second derivatives of the function, for example $\int f^{\prime\prime}(x)^2 dx$, but there are other possibilities such as the discrete penalties suggested by [15].

The penalized spline base–learners can be extended to two or more dimensions to handle interactions by using thin–plate regression splines or tensor products ([68], Section 4.1.5). In the case of a tensor product, smooth bases are built up from products of 'marginal' bases functions. For example,

$$f_3(x_3, x_4) = \sum_{q_3} \sum_{q_4} \beta_{q_3,q_4} b_{q_3}(x_3) b_{q_4}(x_4)$$

A similar representation can be given for the smooth spatial trend surface, $h(no, e)$. Corresponding wiggliness measures are derived from marginal penalties ([68]). Moreover, it is worth mentioning that, when $f(x_3, x_4)$ - or $h(no, e)$ - is represented using a tensor product, the basis for $f(x_3) + f(x_4)$ is strictly nested within the basis for $f(x_3, x_4)$. Thus, in order to test for smooth interaction effects, we do not need to include in the model the two further terms $f(x_3)$ and $f(x_4)$.

In the case of a varying coefficient term like $f_4(x_1)\,l$, the basis functions $b_{q_4}(x_1)$ are premultiplied by a diagonal matrix containing the values of the interaction variable (l). Similarly, in the case of a spatially varying coefficient term like $h(no, e)\,x_1$, the basis functions $b_{q_{no}}(no) b_{q_e}(e)$ are premultiplied by a diagonal matrix containing the values of the interaction variable x_1.

Given the bases for each smooth term, equation (7) can be rewritten in matrix form as a large linear model,

$$\begin{aligned}\mathbf{y} &= \mathbf{X}^*\beta^* + \Sigma_{q_1}\beta_{1q_1}b_{1q_1}(x_1) + \Sigma_{q_2}\beta_{2q_2}b_{2q_2}(x_2) + \ldots + \varepsilon \\ &= \mathbf{X}\beta + \varepsilon \end{aligned} \tag{8}$$

where matrix $\mathbf{X}$ includes $\mathbf{X}^*$ and all the basis functions evaluated at the covariate values, while β contains β^* and all the smooth coefficient vectors, β_q.

As mentioned previously, the number of parameters for each smooth term in a semiparametric model must be large enough to reduce misspecification bias, but not too large to escape over-fitting. To solve this trade–off, we need to penalize lack of smoothness. Thus, starting from the assumption of exogeneity of all the *r.h.s.* variables, model (8) can be estimated by solving the following optimization problem

$$\min \|\mathbf{y} - \mathbf{X}\beta\|^2 + \sum_k \lambda_k \beta' \mathbf{S}_k \beta \qquad w.r.t. \qquad \beta \tag{9}$$

subject to any constraints associated with the bases plus any constraints needed to ensure that the model is identifiable. $\|.\|^2$ is the Euclidean norm and $\lambda_k \geq 0$ are the smoothing parameters that control the fit vs. smoothness trade–off. Employing a large number of basis functions yields a flexible representation of the nonparametric effect $f_k(.)$ where the actual degree of smoothness can be adaptively chosen by varying λ_k.[6]

Given smoothing parameters, λ_k, the solution to (9) is the following penalized least square estimator:

$$\widehat{\beta} = \left(\mathbf{X}'\mathbf{X} + \sum_k \lambda_k \mathbf{S}_k\right)^{-1} \mathbf{X}'\mathbf{y}$$

The covariance matrix of $\widehat{\beta}$ can be derived from that of $\mathbf{y}$

$$V_{\widehat{\beta}} = \sigma_\varepsilon^2 \left(\mathbf{X}'\mathbf{X} + \sum_k \lambda_k \mathbf{S}_k\right)^{-1} \mathbf{X}'\mathbf{X} \left(\mathbf{X}'\mathbf{X} + \sum_k \lambda_k \mathbf{S}_k\right)^{-1}$$

If we also assume normality, that is $\varepsilon \sim \mathcal{N}\left(0, \mathbf{I}_n \sigma_\varepsilon^2\right)$, then

$$\widehat{\beta} \sim \mathcal{N}\left(E\left(\widehat{\beta}\right), V_{\widehat{\beta}}\right)$$

It has been observed, however, that frequentist confidence intervals based on the naive use of $\widehat{\beta}$ and the corresponding covariance matrix perform quite poorly in

[6] It is worth noticing that in expression (9), for interactive terms, the penalty matrix $\mathbf{S}_k$ usually depends on both interacting variables, and the associated λ_k will have two components allowing for different degrees of smoothing.

terms of realized coverage probability ([69]). Thus, in practice, in additive models based on penalized regression splines, frequentist inference yields us to reject the null hypothesis too often. To overcome this problem, and following [63] and [60], [68] [69] has implemented a Bayesian approach to coefficient uncertainty estimation. This strategy recognizes that, by imposing a particular penalty, we are effectively including some prior beliefs about the likely characteristics of the correct model. This can be translated into a Bayesian framework by specifying a prior distribution for the parameters β. Specifically, [69] shows that using a Bayesian approach to uncertainty estimation results in a Bayesian posterior distribution of the parameters

$$\beta|\mathbf{y} \sim \mathscr{N}\left(E\left(\widehat{\beta}\right), \sigma_{\varepsilon}^{2}\left(\mathbf{X}'\mathbf{X} + \sum_{k}\lambda_{k}\mathbf{S}_{k}\right)^{-1}\right)$$

This latter result can be used directly to calculate credible intervals for any parameter. Moreover, the credibility intervals derived via Bayesian theory are well behaved also from a frequentist point of view, i.e., their average coverage probability is very close to the nominal level $1-\alpha$, where α is the significance level.

3.3 Smoothing Parameter Selection: GCV Score Minimization

A crucial issue in the use of penalized regression splines within an additive semi-parametric model is the selection of the smoothing parameters, λ_k, controlling the trade–off between fidelity to the data and smoothness of the fitted spline. How should these values be selected? There are two main approaches to identify the optimum smoothing parameters. First, we can use prediction error criteria, such as generalized cross validation (GCV), Akaike information criterion (AIC), Bayesian information criterion (BIC) and so on. Alternatively we can rewrite the penalized additive model as a mixed model by decomposing each smooth term into fixed effect and random effect components and estimate the model by ML or restricted maximum likelihood (REML), treating λ_k as variance parameters (see Sect. 4).

As for the first method, we may select the values of $\widehat{\lambda}_k$ that minimize the GCV score:

$$GCV(\lambda_k) = \frac{n \parallel \mathbf{y} - \mathbf{X}\widehat{\beta} \parallel^2}{[n - tr(\mathbf{H})]^2}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X} + \sum \lambda_k \mathbf{S}_k)^{-1}\mathbf{X}'$ is the hat matrix for the model being fitted and its trace, $tr(\mathbf{H})$, gives the effective degrees of freedom edf (i.e., the number of identifiable parameters in the model). The edf are a general measure for the complexity of a function estimates, which allows us to compare the smoothness, even for different types of effects (e.g. nonparametric versus parametric effects). If λ_k=0, then edf is

equal to the size of the β vector minus the number of constraints (i.e., $edf = K$). Positive values of λ_k lead to an effective reduction of the number of parameters (i.e., $edf < K$). If λ_k is high, we have very few edf.

Actually, multiple smoothing parameter selection based on the minimization of the GCV score (10) is often too computationally demanding. To overcome this problem, [65] extended the 'performance iteration' method proposed by [27] for automatically select multiple smoothing parameters to the case of computationally efficient low rank additive models based on penalized regression splines. First, the multiple smoothing parameter model fitting problem is re-written with an extra *overall* smoothing parameter (δ) controlling the trade-off between model fit and overall smoothness, while retaining smoothing parameters multiplying each individual penalty, which now control only the relative weights given to the different penalties. The following steps are then iterated: (1) given the current estimates of the relative smoothing parameters (λ_k/δ), estimate the overall smoothing parameter; and (2) given the overall smoothing parameter, update $log(\lambda_k)$ by Newton's method. In this way, the smoothing parameters for each smooth term in the model are chosen simultaneously and automatically as part of the model fitting. A drawback of this method is that it does not allow users to fix some smoothing parameters and estimate others or to bound smoothing parameters from below. Moreover, the method is not optimally stable numerically.

More recently, [67] proposed an improved (optimally stable) version of the the 'performance iteration' method which is more robust to collinearity or concurvity problems and which can deal with fixed penalties. The second issue is very important when fully automatic smoothing parameter selection result in one or more model terms clearly over-fitting and thus it is necessary to fix or bound smoothing parameters. The issue is also particularly relevant in geoadditive models, since the smooth function of spatial location ($h(no_i, e_i)$), which enter the model as a nuisance term – i.e. only to explain variability that cannot be explained by the covariates that are really of interest – is often estimated with bounded smoothing parameters, while the 'interesting' term are left with free smoothing parameters: in this way the 'interesting' covariates can be forced to do as much of the explanatory work as possible.

3.4 Semiparametric Spatial Autoregressive Geoadditive Models

Matrix $\mathbf{X}$ in model (8) may include spatial lags of the covariates, thus providing a Semiparametric Geoadditive Spatial Lag of $\mathbf{X}$ Model (SLX) (or SLX-Geo-AM). In other words, if $\widetilde{\mathbf{X}}$ is the matrix of regional characteristics, $\mathbf{X}$ includes both $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{W}_n\mathbf{X}}$, where $\mathbf{W}_n$ is a row-standardized spatial weights matrix. It is important to remark again, however, that the $\widetilde{\mathbf{W}_n X}$ only captures *local spatial externalities*. By replacing $\mathbf{X}$ and ε with $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}\mathbf{X}$ and $(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}\varepsilon$, respectively, it is possible to model *global spillovers*.

The introduction of the spatial multiplier effect in the model yields a reduced form as $\mathbf{y} = (\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}\mathbf{X} + (\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}\varepsilon$ and the structural form becomes a

Semiparametric Spatial Autoregressive Geoadditive Model (SAR-Geo-AM):

$$y_i = \mathbf{X}_i^{*'}\beta^* + \rho \sum\nolimits_{j=1}^{n} w_{ij} y_j + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + f_4(x_{1i})\, l_i + ... + h(no_i, e_i) + \varepsilon_i \tag{10}$$

$$\varepsilon_i \sim iid\mathcal{N}(0, \sigma_\varepsilon^2) \qquad i = 1, ..., n$$

where, again, w_{ij} are the elements of $\mathbf{W}_n$, $\sum_{j=1}^{n} w_{ij} y_j$, which captures the spatial lag of the dependent variable (which always enters the model linearly), and ρ is the spatial spillover parameter. This model was first proposed by [26] and [7] and then reformulated by [4], [5], [6], [48], [47], [62] and [61]. It reflects the notion of a spatial correlation comprised of two parts: (*i*) a spatial trend due to unobserved regional characteristics, which is modeled by the smooth function of the coordinates, and (*ii*) local and/or global spatial spillover effects, which can be modeled by including spatial lag terms of the independent and dependent variables. [61] extends this model to allow for both heteroskedasticity and spatial dependence in the error term.

As mentioned in Sect. 2, the spatial lag term $\sum_{j=1}^{n} w_{ij} y_i$ and the error term ε_i are correlated. In order to deal with this endogeneity problem, the "control function" approach ([11]) can be used ([5]). This is a simple two–step procedure. In the first step, an auxiliary semiparametric regression

$$\sum\nolimits_{j=1}^{n} w_{ij} y_i = \mathbf{X}_i^{*'}\beta^* + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + f_4(x_{1i})\, l_i + ... + h(no_i, e_i) + \sum\nolimits_m g_m(Q_{mi}) + \nu_i \tag{11}$$

is considered, with $\mathbf{Q}_i$ a set of m conformable instruments,[7] and ν_i a sequence of random variables satisfying conditional mean restrictions $E(\nu_i|\mathbf{Q}_i) = 0$.

The second step consists of estimating an additive model of the form:[8]

$$y_i = \mathbf{X}_i^{*'}\beta^* + \rho \sum\nolimits_{j=1}^{n} w_{ij} y_j + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + f_4(x_{1i})\, l_i + ... + h(no_i, e_i) + c(\widehat{\nu}_i) + \varepsilon_i \tag{12}$$

Obviously, the endogeneity of any other continuously distributed regressor in model (10) can also addressed via the control function approach if valid instruments are available.[9] Finally, it is important to note that endogeneity problems arising from omitted variables (i.e., missing permanent characteristics that drive both the response variable and the covariates) can be ruled out since in model (10) we directly

[7] For example, in line with [33], $\mathbf{Q}_i$ may contain an intercept, all exogenous terms included in the model and several orders of their spatial lags.

[8] Both first and second step equations can be estimated by using, for example, penalized least squares estimators.

[9] The requirement that the endogenous regressor be continuously distributed is the most important limitation of the applicability of the control function approach to estimation of nonparametric and semiparametric models with endogenous regressors.

control for the effect of "first nature" characteristics by including the smooth interaction between latitude and longitude of the regional units of analysis.

In Sect. 4 we present a different possible solution to this simultaneity problem based on the REML estimation approach.[10] However, before leaving this discussion, an important issue remains to be introduced. Specifically, as in the parametric SAR model, in the semiparametric SAR also the estimated coefficients of parametric terms, as well as the estimated smooth effects of nonparametric terms, cannot be interpreted as marginal impacts of the explanatory variables on the dependent variable, due to the presence of a significant spatial autoregressive parameter (ρ). Taking advantage of the results obtained for parametric SAR models (see Sect. 2), we can define similar algorithms for the semiparametric SAR model. Specifically, we can compute the total effect of variable x_k as

$$\widehat{f}_k^{te_k}(x_k) = \Sigma_q \left[\mathbf{I}_n - \widehat{\rho}\mathbf{W}_n\right]_{ij}^{-1} b_{kq}(x_k)\widehat{\beta}_{kq} \tag{13}$$

Finally, we can compute direct and indirect (or spillover) effects of smooth terms in semiparametric SAR as follows:

$$\widehat{f}_k^{de_k}(x_k) = \Sigma_q \left[\mathbf{I}_n - \widehat{\rho}\mathbf{W}_n\right]_{ii}^{-1} b_{kq}(x_k)\widehat{\beta}_{kq} \tag{14}$$

$$\widehat{f}_k^{ie_k}(x_k) = \widehat{f}_k^{te_k}(x_k) - \widehat{f}_k^{de_k}(x_k) \tag{15}$$

4 Semiparametric Geoadditive Models as Mixed Models

Semiparametric models presented in the previous section can also be expressed as mixed models. Consequently, it is possible to estimate all the parameters of these models using restricted maximum likelihood methods (REML). In Subsect. 4.1 we present some ways to deal with general semiparametric models using mixed models. Moreover, it is possible to estimate the whole set of parameters (including those for smoothing and interaction between variables) using REML. Subsect. 4.2 shows how this methodology can be applied to estimate the parameters of *SAR-Geo-AM* model in a single step.

[10] It is important to mention that a semiparametric spatial lag model has also been proposed within a partial linear framework. For example, [62] develop a profile quasi–maximum likelihood estimator for the partially linear spatial autoregressive model which combines the spatial autoregressive model and the nonparametric (local polynomial) regression model. Furthermore, [61] proposes a semiparametric GMM estimator of the SAR model under weak moment conditions which allows for both heteroskedasticity and spatial dependence in the error terms.

4.1 Model Specification and REML Estimation

The estimation of model (8) can be based on the reparameterization of such a model in the form of a mixed model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{U} + \varepsilon \quad \mathbf{U} \sim \text{ i.i.d. } N(\mathbf{0}, \mathbf{G}) \quad \varepsilon \sim \text{ i.i.d. } N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}) \tag{16}$$

where $\mathbf{G}$ is a block–diagonal matrix, which depends on both $\sigma^2_{u_k}$ and σ^2_ε variances. The smoothing parameters are defined by the ratios $\lambda_k = \frac{\sigma^2_\varepsilon}{\sigma^2_{u_k}}$. Again, matrix $\mathbf{X}$ may include parametric components such as the intercept, continuous covariates and categorical covariates.

This reparameterization consists in postmultiplying $\mathbf{X}$ and premultiplying β in model (8) by an orthogonal matrix resulting from the singular value decomposition of the penalty matrices $\mathbf{S}_k$ ([64]; [38]; [72]). Therefore, the type of penalizations determines the transformation matrix and, thus, the fixed and random effects obtained in the mixed model. The resulting coefficients associated with the fixed effects are not penalized, while those associated with the random effects are penalized. The penalization of random effects is given by the variance–covariance matrix of these coefficients.

It is worth pointing out that when the model is a pure additive model $\mathbf{y} = \sum_{k=1}^K f(x_k) + \varepsilon$ (i.e. there are no interaction terms), $\mathbf{G}$ is block–diagonal, each block matrix $\mathbf{G}_k$ depending only on λ_k, the smoothing coefficient associated to each variable x_k. Thus, model (16) becomes a variance components model that can be estimated by using standard software on the topic. When the model contains interaction terms, it is not longer a pure additive model. Therefore, each block $\mathbf{G}_k$ depends on more than one smoothing coefficient λ_k, except in the isotropic case,[11], where coefficients λ_k are the same for all variables ([72]; [38]). As a consequence, the resulting mixed model is not an orthogonal variance component model and standard software cannot be used to estimate it.

A recent reparameterization, proposed by [72], allows us to express a semiparametric model including additive and interaction effects as a mixed model with orthogonal variance components, allowing for different degrees of smoothing for the variables that interact using only one smoothing coefficient for each term. An alternative reparameterization from a P–Spline approach with a B–Spline basis and penalization matrices for the basis coefficients based on discrete differences is considered in [15] and [38]. Two other interesting reparameterizations are based on (i) a truncated polynomial basis and ridge penalizations ([59]), and (ii) on a thin plate regression splines basis and penalizations based on the integral of the second derivatives of the spline functions ([66]). These last three alternatives cannot be

[11] For the sake of clarity, isotropy means that the degree of smoothness is the same for all the covariates, that is, the degree of flexibility in all of them is the same. Nevertheless, the usual situation in real cases is anisotropy, since the covariates are usually measured in different units of measure or, in the case of equal measurement units (e.g. spatial location variables), the variability of such covariates differing greatly.

estimated with standard software on mixed models when the interactions between the variables are considered (except in the isotropic case).

Once the mixed model is defined, the parameters associated to fixed (β) and random effects (λ_k and σ_ε^2) can be estimated by using a ML algorithm. If the noise term follows a Gaussian distribution, the log–likelihood function is given by:

$$\log L(\beta, \lambda_1, \cdots, \lambda_K, \sigma_\varepsilon^2) = \text{constant} - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$$

where $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \sigma_\varepsilon^2\mathbf{I}$ and the smoothing parameters λ_k are included in $\mathbf{V}$.

However, the ML estimates are biased since this method does not take into account the reduction in the degrees of freedom due to the estimation of the fixed effects. The restricted maximum likelihood (REML) method can be used to solve the problem. The REML method looks for the linear combinations of the dependent variable that eliminates the fixed effects in the model ([45]). In this case the objective function to maximize is given by:

$$\begin{aligned} \log L_{\mathrm{R}}(\lambda_1, \cdots, \lambda_K, \sigma_\varepsilon^2) &= \text{constant} - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \\ &\quad - \frac{1}{2}\mathbf{y}'\left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right)\mathbf{y} \end{aligned}$$

An estimation of the variance components parameters can be obtained after maximizing $\log L_{\mathrm{R}}(.)$. In a second step, the estimates of β and $\mathbf{U}$ are given by ([45]):

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y} \\ \hat{\mathbf{U}} &= \hat{\mathbf{G}}\mathbf{X}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}) \end{aligned}$$

Finally, the estimated values of the observed variable can be obtained as:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} + \mathbf{Z}\hat{\mathbf{U}}$$

To build confidence intervals for the estimated values, an approximation of the variance–covariance matrix of the estimation error is given by $V(\mathbf{y} - \hat{\mathbf{y}}) = \sigma_\varepsilon^2\mathbf{H}$ where, as shown previously in the GCV method, $\mathbf{H}$ is the hat matrix of the model ([59]). For the mixed model, it can be proved that:

$$\mathbf{H} = \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} \end{pmatrix}$$

Recently [71] has proposed a Laplace approximation to obtain an approximated REML or ML for any generalized linear model, which is suitable for efficient direct optimization. Simulation results indicate that these novel REML and ML procedures offer, in most cases, significant gains (in terms of mean–square error) with respect to GCV or AIC methods.

4.2 Semiparametric Spatial Autoregressive Geoadditive Models as Mixed Models

In a mixed–model form the SAR-Geo-AM can be expressed as:

$$\mathbf{y} = \rho \mathbf{W}_n \mathbf{y} + \mathbf{X}\beta + \mathbf{Z}\mathbf{U} + \varepsilon \quad \mathbf{U} \sim \text{ i.i.d. } N(\mathbf{0}, \mathbf{G}) \quad \varepsilon \sim \text{ i.i.d. } N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I})$$

In reduced form we have:

$$\mathbf{y} = \mathbf{A}\mathbf{X}\beta + \mathbf{A}\mathbf{Z}\mathbf{U} + \mathbf{A}\varepsilon \tag{17}$$

where $\mathbf{A} = (\mathbf{I} - \rho \mathbf{W}_n)^{-1}$.

As pointed out in [48] and [47], the log–REML function for model (17) is:

$$\begin{aligned} \log L_{\mathrm{R}}(\rho, \lambda_1, \cdots, \lambda_K, \sigma_\varepsilon^2) = {} & \text{constant} - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}| + \log |\mathbf{A}| \\ & - \frac{1}{2} \mathbf{y}' \mathbf{A}' \left(\mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \right) \mathbf{A}\mathbf{y} \end{aligned}$$

As usual, $\log L_{\mathrm{R}}(.)$ is maximized with respect to the parameter vector $(\lambda_1, \cdots, \lambda_K, \sigma_\varepsilon^2)'$. Note that the maximization process requires the computation of the log-determinant of matrix $\mathbf{A}$, a dense $n \times n$ inverse matrix depending on ρ. As a consequence, the maximization of such a function constitutes a challenging task. Nevertheless, to evaluate $\mathbf{A}$ for different values of ρ when n is large, it is possible to use Monte Carlo procedures ([40]).

Finally, fixed and random effects can be estimated as:

$$\begin{aligned} \hat{\beta} &= (\mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X}) \mathbf{X}' \hat{\mathbf{V}}^{-1} \hat{\mathbf{A}} \mathbf{y} \\ \hat{\mathbf{U}} &= \hat{\mathbf{G}} \mathbf{X}' \hat{\mathbf{V}}^{-1} (\hat{\mathbf{A}} \mathbf{y} - \mathbf{X} \hat{\beta}) \end{aligned}$$

Unlike model (16), the spatial lag model (17) cannot be estimated by using standard software regardless of the type of reparameterization used to express it as a mixed model.[12]

5 Empirical Applications

The scientific contributions to the empirical literature on regional economic dynamics and economic geography may be roughly classified into four main categories depending on whether they apply *i*) parametric non-spatial models, *ii*) parametric spatial models, *iii*) semiparametric non-spatial models and, *iv*) semiparametric spatial geoadditive models. In this Section we mention examples for each category, focusing on empirical studies on regional and urban growth and on economic geog-

[12] Nevertheless, there are some *R* codes using *spdep* package available from [48] and [47].

raphy. In particular, in Sect. 5.2 we present an application of the SAR-Geo-AM to regional growth data.

5.1 Studies Based on Parametric Models

The main difference between the first two categories lies in the treatment of spatial effects. Studies belonging to the first category are based on the assumption of independence and, thus, spillover or feedback effects are ruled out. For example, [44] provides a review of empirical studies on regional growth and convergence, which mainly focus on testing the β-convergence hypothesis, by regressing the so-called linear Barro's equation, derived from the neoclassical Solow model and based on the assumption of cross-region independence. Similarly, in the field of urban economics, a large number of studies have investigated the relationship between agglomeration externalities and local economic growth, ignoring regional interdependence ([58]; [9]). Moreover, several studies have put into empirical test selected propositions of New Economic Geography (NEG) models, such as the prediction of a positive relationship between local wages and market potential (home market effect). However, none of the empirical studies on the NEG wage equation take spatial dependence into account ([29]; [13]; [56]).

Recent studies ([42]; [53]; [18], [19]) have shown that spatial technological interdependence can be explicitly modeled in multi-region exogenous and endogenous growth frameworks to account for neighborhood effects in growth and convergence processes. These studies have provided sound theoretical foundations for the specific form taken by spatial autocorrelation in econometric growth models. Thus, they have stimulated the empirical assessment of the existence of neighboring effects in regional growth ([57]).

In the same spirit, [10] stressed the importance of extending the two–region NEG models into multi–region models:

> when there are just two regions, there is only one way in which these regions can interact, namely directly; whereas with three regions, there are two ways in which these regions can interact, namely directly and indirectly. In other words, in multiregional systems the so-called 'three–ness effect' enters the picture and introduces complex feedbacks into the models, which significantly complicates the analysis. Dealing with these spatial interdependencies constitutes one of the main theoretical and empirical challenges NEG and regional economics will surely have to face in the future (p. 461).

Multi–region NEG models have been empirically tested by [12] and [22]. In particular, [12] derive a reduced–form linearized wage equation, which is a simple SAR model of order one in short–run deviations of local wages from their equilibrium values. This model relates these deviations in each region to the weighted sum of the deviations in all regions. The spatial weights are bilateral elasticities of the wage rate in one region with respect to the wage rate in another region. Estimation of this model thus becomes a test of whether or not local wage shocks propagate through the system of observed regional wages in the way predicted by the NEG model.

While the insights of these empirical studies employing sophisticated spatial econometric techniques are valuable to measure spatial spillover and feedback effects, their foundation on a priori functional form specification limits the scope of these methods in uncovering the process dictating regional dynamics.

5.2 Studies Based on Semiparametric Models

In this section, we mention examples of recent empirical works that adopt semiparametric additive models in regional science and economic geography. First, it is important acknowledge the existence of a several studies using semiparametric methods to identify nonlinearities, and of a few studies using spatial lag semiparametric models to tackle both spatial dependence and nonlinearities.

Following the cross–country growth literature ([14]), an emerging issue in regional growth analyses has been the evidence of strong nonlinearities in regional growth models ([25]; [3]). Using semiparametric methods, these studies have uncovered the existence of significant nonlinearities across an array of variables within cross–region growth regressions. Moreover, in the field of urban growth, [8] propose a semiparametric geoadditive model to identify important nonlinearities in the relationship between local industry structure (e.g. population density) and local employment growth and to control for unobserved spatial heterogeneity. Although these studies are able to relax functional form assumptions, their consistency still depends on restrictive assumptions about interregional independence. Finally, the need to consider jointly spatial dependence and nonlinearities has been raised by [26], [7], [4], [5] and [6].

As an example of the application of a semiparametric spatial lag geoadditive model (SAR-Geo-AM), we report the results of the estimation of a regional growth regression model on a sample of 249 NUTS2 regions belonging to the enlarged Europe (EU27). The dependent variable, y, is the per-worker income growth rate — $y = (ln(p_T) - ln(p_0))/T$ —, computed for the 1990–2004 period. The covariates are the rates of investment in physical and human capital — ($ln(s_k)$ and $ln(s_h)$, respectively) —, initial conditions — $ln(p_0)$ — and the effective depreciation rate — ($ln(n+g+\delta)$), with n the working–age population growth rate, g the common exogenous technology growth rate and δ the rate of depreciation of physical capital assumed identical in all economies. Basic data to measure these variables come from the EUROSTAT Regio and Cambridge Econometrics databases, which include information on real gross value added, employment, investment and tertiary education (for further details, see [6]). The estimated model is

$$\begin{aligned} y_i = {} & \beta_0 + \rho \textstyle\sum_{j=1}^{n} w_{ij} y_j + f_1\left(ln(p_0)_i\right) + f_2\left(ln(s_k)_i\right) \\ & + f_3\left(ln(s_h)_i\right) + f_4\left(ln(n+g+\delta)_i\right) + h\left(no_i, e_i\right) + \varepsilon_i \end{aligned} \tag{18}$$

$$\varepsilon_i \sim iid\mathcal{N}(0, \sigma_\varepsilon^2) \qquad i = 1, \ldots, n$$

The matrix $\mathbf{W}_n$ used to estimate this model has been selected among a number of inverse–distance spatial weights matrices. The model has been estimated using spline–based penalized regression smoothers which allow for automatic and integrated smoothing parameters selection via *GCV*. A control function approach has been used in order to deal the endogeneity of the spatial lag term, and the spatial lags of the exogenous variables have been used as valid instruments.

The F tests for the overall significance of the smooth terms have p values lower than 0.05 in all cases, while the number of *edf* suggests that the relationship between regional growth and its determinants is far from being linear. Figures 1a–d show the fitted univariate smooth functions, alongside Bayesian credibility intervals at the 95% level of significance. The value of the spatial autocorrelation parameter ρ is equal to 0.88 and statistically significant at 1%, confirming the role of spatial frictions in the interregional diffusion of technological spillovers.

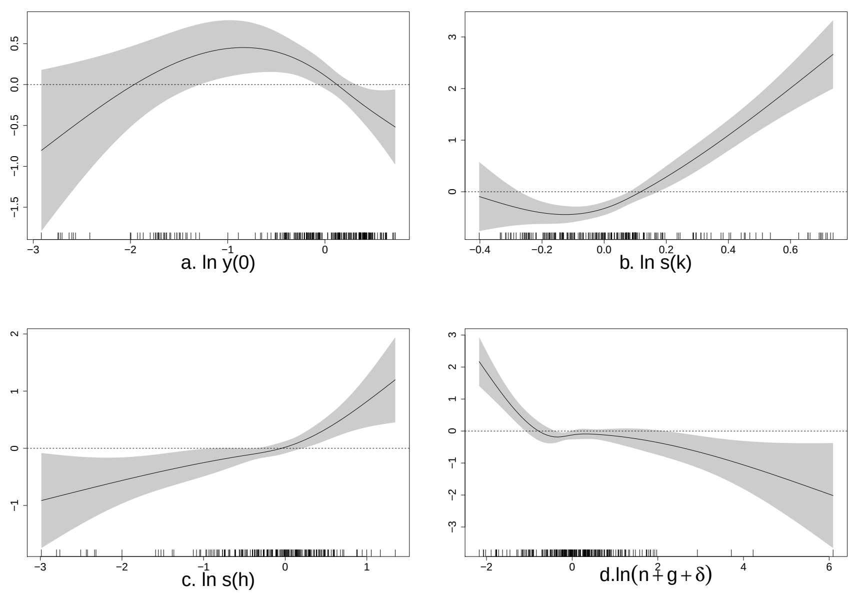

Fig. 1 Smooth effects of growth determinants. **a** Initial conditions. **b** Physical capital accumulation rate. **c** Human capital accumulation rate. **d** Effective rate of depreciation. Solid lines represent smooth functions of each term, alongside Bayesian confidence intervals (shaded grey areas) at the 95% level of significance. In each plot, the vertical axis displays the scale of the estimated smooth function, while the horizontal ones report the scale of each determinant (in deviations to the EU average). Rug plot along the horizontal axis represents observed data.

A hump–shaped relationship between growth and initial conditions emerges. Specifically, a diverging behavior characterizes the group of Eastern regions (45 regions), while Western regions maintain a conditional predicted convergence path.

The assumption of identical speed of convergence is therefore rejected. Nonlinearities in the effects of gross physical investments are also clearly detected. Specifically, an increase in the saving rate is associated with an increase in growth rates only when the saving rate is above the EU average. A similar threshold effect is also evident in the smooth effect of human capital investments. The influence of the employment growth rate on regional growth is negative, although the effect is not homogeneous across the sample. Moreover, Fig. 2 displays the effect of the smooth interaction between latitude and longitude. It can be observed that, ceteris paribus, some North–Eastern and some North–Western regions (mainly the UK and Ireland) have higher predicted growth rates.

Finally, it is important to remark again that the estimated smooth effects of nonparametric terms cannot be interpreted as marginal impacts of the explanatory variables on the dependent variable. Therefore, using Eqs. (13), (14) and (15), we have computed direct, indirect and total smooth effects. Actually, these effects are not smooth at all over the domain of variable x_k due to the presence of the spatial multiplier matrix in these algorithms. A wiggly profile of direct, indirect and total effects

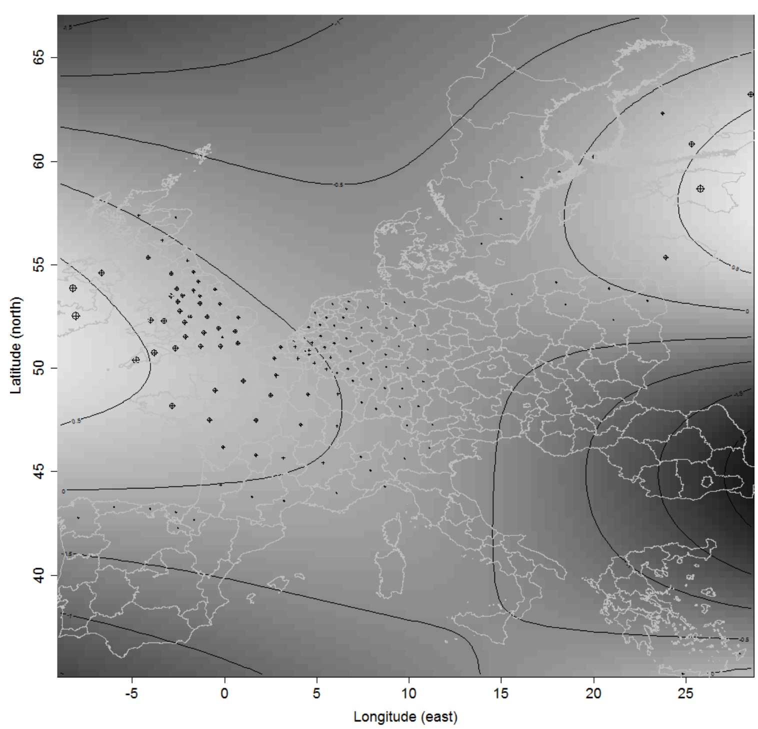

Fig. 2 Spatial trend surface

would appear even if the model were linear.[13] Therefore, in the spirit of this paper, we have applied a spline smoother to obtain smooth curves (see Fig. 3). Briefly commenting these evidences, we first note that the shape of direct effects is very similar to the one displayed in Fig. 1, which means that the feedback effect is rather negligible. Moreover, indirect effects (spillover effects) are always lower that direct effects.

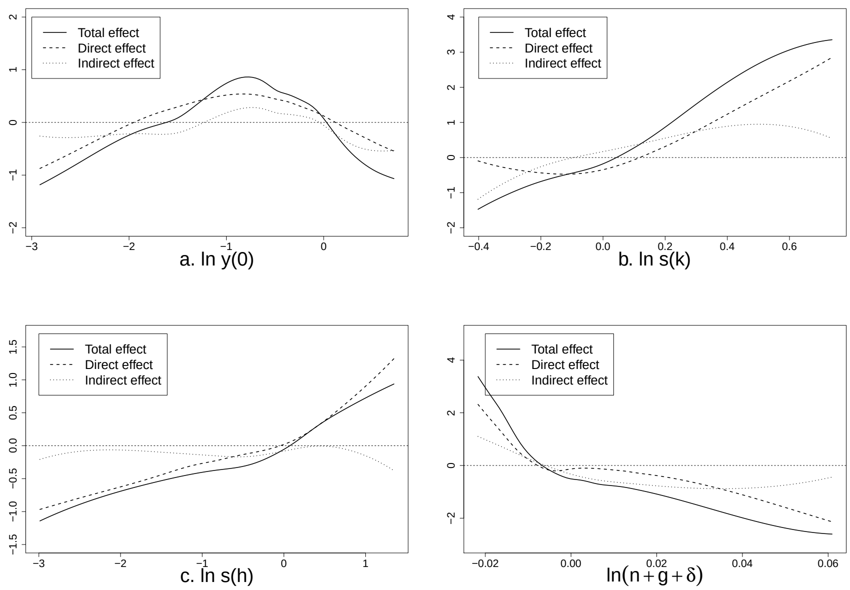

Fig. 3 Direct, Indirect and Total Smooth Effects (see Eqs. 13, 14 and 15). **a** Initial conditions. **b** Physical capital accumulation rate. **c** Human capital accumulation rate. **d** Effective rate of depreciation.

6 Concluding Remarks

In this paper we have reviewed recently developed spatial lag semiparametric geoadditive models and presented some applications of them in the fields of regional science and economic geography. These methods play a prominent role in those context in which the theory suggests the existence of spatial interdependence and heterogeneous behavior of the spatial units. Natural directions in which these methods can be

[13] This kind of heterogeneity is called interactive heterogeneity by [19], and this is why scholars usually compute average marginal effects for parametric SAR and SDM models.

extended are a specification for longitudinal data and, eventually, a dynamic framework. Also testing spatial autocorrelation in the residuals in both non-spatial and spatial lag semiparametric geoadditive models is an important task.

Acknowledgements This paper has been presented at the EU-COST Meeting in Lisboa. We thank all participants for useful comments. We are also grateful to two referees that with their comments helped us to reformulate the analysis. We are responsible for any remaining errors. The work of Román Mínguez was supported by the research project MTM-2011-28285-C02-C2 from the Spanish Governments Ministry of Economy and Competitiveness. Jesús Mur likes to thank the financial support of the research project ECO2012-36032-C03-01 from the Spanish Governments Ministry of Economy and Competitiveness.

References

1. Anselin L (2003) Spatial Externalities, Spatial Multipliers and Spatial Econometrics. Int Reg Science Rev 26: 153–166
2. Augustin N, Musio M, von Wilpert K, Kublin E, Wood SN, Schumacher M (2009) Modeling Spatio–Temporal Forest Health Monitoring Data. J Am Stat Assoc 104: 899–911
3. Azomahou TT, El Ouardighi J, Nguyen-Van P, Pham TKC (2011) Testing Convergence of European Regions: A semiparametric approach. Econ Model 28: 1202–1210
4. Basile R (2008) Regional Economic Growth in Europe: a Semiparametric Spatial Dependence Approach. Pap Reg Science 87: 527–544
5. Basile R (2009) Productivity Polarization across Regions in Europe: The Role of Nonlinearities and Spatial Dependence. Int Reg Science Rev 32: 92–115
6. Basile R, Capello R, Caragliu A (2012) Technological Interdependence and Regional Growth in Europe. Pap Reg Science 91: 697–722
7. Basile R, Gress B (2005) Semi-parametric Spatial Auto-covariance Models of Regional Growth Behavior in Europe. Région et Développement 21: 93–118
8. Basile R, Donati C, Pittiglio R (2013) Industry Structure and Employment Growth: Evidence from Semiparametric Geoadditive Models. Mimeo
9. Beaudry C, Schiffauerova A (2009) Who's Right, Marshall or Jacobs? The Localization versus Urbanization Debate. Res Pol 38: 318–337
10. Behrens K, Thisse JF (2007) Regional Economics: A New Economic Geography Perspective. Reg Science Urb Econ 37: 457-4-65
11. Blundell R, Powell J (2003) Endogeneity in Nonparametric and Semiparametric Regression Models. In: Dewatripont M, Hansen L and Turnsovsky SJ (eds) Advances in Economics and Econometrics Theory and Application. Cambridge University Press, Cambridge
12. Bode E, Mutl J (2010) Testing Nonlinear New Economic Geography Models. Reihe Ökonomie Working Paper No. 253.
http://hdl.handle.net/10419/45569
13. Brakman S, Garretsen H, Schramm M (2006) Putting New Economic Geography to the Test: Free-ness of Trade and Agglomeration in the EU Regions. Reg Science Urb Econ 36: 613-635
14. Durlauf SN, Johnson PA, Temple JRW (2005) Growth Econometrics. In: Aghion P, Durlauf S (eds) Handbook of Economic Growth 555–677. North Holland, New York
15. Eilers PHC, Marx BD (1996) Flexible Smoothing with B-Splines and Penalties. Statist Science 11 89–121
16. Elhorst P (2010) Spatial Panel Data Models. In: Fisher M, Getis A (eds) Handbook of Applied Spatial Analysis 377–407. Springer, Berlin
17. Elhorst P (2012) Dynamic Spatial Panels: Models, Methods, and Inferences. J Geogr Systems 14: 5–28

18. Ertur C, Koch W (2007) Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence. J Appl Econometrics 22: 1033–1062
19. Ertur C, Koch W (2011) A Contribution to the Schumpeterian Growth Theory and Empirics. J Economic Growth 16: 215–255
20. Ertur C, Le Gallo J (2009) Regional Growth and Convergence: Heterogeneous Reaction versus Interaction Spatial Econometric Approaches. In: Capello R, Nijkamp P (eds) Handbook Of Regional Growth And Development Theories 374–388. Edward Elgar, Cheltenham
21. Fingleton B (2004) Regional Economic Growth and Convergence: Insights from a Spatial Econometric Perspective. In: Anselin L, Florax R, Rey S (eds) Advances in Spatial Econometrics 397–432. Springer, Berlin
22. Fingleton B (2006) The New Economic Geography versus Urban Economics: an Evaluation Using Local Wage Rates in Great Britain. Oxf Econ Pap 58: 501–530
23. Fingleton B, Lopez-Bazo E (2006) Empirical Growth Models with Spatial Effects. Pap Reg Science 85: 177–198
24. Fotheringham AS, Brunsdon C, Charlton ME (2002) Geographically Weighted Regression. Wiley, Chichester
25. Fotopoulos G (2012) Nonlinearities in Regional Economic Growth and Convergence: the Role of Entrepreneurship in the European Union Regions. Ann Reg Science 48: 719–741
26. Gress B (2004) Using Semi-Parametric Spatial Autocorrelation Models to Improve Hedonic Housing Price Prediction. Mimeo. Department of Economics, University of California
27. Gu C, Wahba G (1991) Minimizing GCV/GML scores with multiple smoothing parameters via the Newton method. SIAM J Sci Stat Comp 12: 383–398
28. Hastie TJ, Tibshirani RJ (1990) Generalized Additive Models. Chapman and Hall, London
29. Head K, Mayer T (2004) The Empirics of Agglomeration and Trade. In: Henderson JV, Thisse JF (eds) Handbook of Urban and Regional Economics 4. North Holland, Amsterdam
30. Henderson V (1997) Externalities and Industrial Development. J Urb Econ 42: 449–470
31. Holly S, Pesaran H, Yamagata T (2010) A Spatio-Temporal Model of House Prices in the US. J Econometrics 158: 160–173
32. Kapoor M, Kelejian H, Prucha I (2007) Panel Data Models with Spatially Correlated Error Components. J Econometrics 140: 97–130
33. Kelejian H, Prucha I (1997) Estimation of Spatial Regression Models with Autoregressive Errors by Two-Stage Least Squares Procedures: A Serious Problem. Int Reg Science Rev 20: 103–111
34. Kelejian H, Prucha I (2001) On the Asymptotic Distribution of Moran I Test Statistic with Applications. J Econometrics 104: 291–257
35. Kelejian H, Robinson D (2004) The Influence of Spatially Correlated Heteroskedasticity on Tests of Spatial Correlation. In: Rey S, Florax R, Anselin L (eds): Advances in Spatial Econometrics: Methodology, Tools and Applications. Springer, Berlin
36. Kneib T, Hothorn T, Tutz G (2009) Variable selection and model choice in geoadditive regression models. Biometrics 65: 626–34
37. Krugman P (1993) First Nature, Second Nature, and Metropolitan Location. J Reg Science 33: 129–144
38. Lee DJ, Durbán M (2011) P-Spline ANOVA Type Interaction Models for Spatio-Temporal Smoothing. Statistical Modelling 11: 49–69
39. Lee L (2004) Asymptotic Distribution of Quasi-Maximum Likelihood Estimators for Spatial Econometric Models. Econometrica 72: 1899–1926
40. Le Sage J, Pace K (2009) Introduction to Spatial Econometrics. Boca Ratón: Taylor and Francis
41. Liu X, Lee L, Bollinger C (2007) An Efficient GMM of Spatial Autoregressive Models. J Econometrics 159: 303–319
42. López-Bazo E, Vayá E, Artís M (2004) Regional Externalities and Growth: Evidence from European Regions. J Reg Science 44: 43–73
43. Lloyd C (2011) Local Models for Spatial Analysis 2nd edn. CRC Press, Boca Ratón
44. Magrini S (2004) Regional (di)Convergence. In Henderson JV and Thisse JF (eds): Handbook of Regional and Urban Economics. North Holland, New York

45. McCulloch CE, Searle SR, Neuhaus JM (2008) Generalized, Linear, and Mixed Models 2nd edn. Wiley Series in Probability and Statistics
46. McMillen DP (2003) Spatial Autocorrelation or Model Misspecification? Int Reg Science Rev 26: 208–217
47. Mínguez R, Durbán M, Montero JM, Lee DJ (2012) Competing Spatial Parametric and Non-Parametric Specifications. Mimeo
48. Montero JM, Mínguez R, Durbán M (2012) SAR Models with Nonparametric Spatial Trends. A P-Spline Approach. Estadística Española 54: 89–111
49. Mur J, López E, Angulo A (2009) Testing the Hypothesis of Stability in Spatial Econometric Models. Pap Reg Science 88: 409–444
50. Pace K, Lesage J (2004) Spatial Autoregressive Local Estimation. In Getis A, Mur J and Zoller H (eds.) Spatial Econometrics and Spatial Statistics 31–51. Palgrave, Basingstoke
51. Paci R, Usai S (2008) Agglomeration Economies, Spatial Dependence and Local Industry Growth. Revue D'Economie Industrielle 123: 1–23
52. Partridge M, Boarnet M, Brakman S, Ottaviano G (2012) Introduction: Whither Spatial Econometrics. J Reg Science 52: 167–171
53. Pfaffermayer M (2009) Conditional β and σ-Convergence in Space: A Maximum Likelihood Approach. Reg Science Urb Econ 39: 63–78
54. Pinheiro J, Bates D (2000) Mixed-effects models in S and S-PLUS. Springer-Verlag, New York
55. Pinkse J, Slade M, Brett C (2002) Spatial Price Competition. A Semiparametric Approach. Econometrica 70: 1111–1153
56. Redding SJ (2010) The Empirics of New Economic Geography. J Reg Science 50: 297–311
57. Rey SJ, Le Gallo J (2009) Spatial Analysis of Economic Convergence. In: Mills TC, Patterson K (eds) Handbook of Econometrics Volume II: Applied Econometrics. Palgrave, Basingstoke
58. Rosenthal SS, Strange WC (2004) Evidence on the Nature and Sources of Agglomeration Economies. In: Henderson JV, Thisse JF (eds) Handbook of Urban and Regional Economics. North Holland, New York
59. Ruppert D, Wand MP, Carroll RJ (2003) Semiparametric Regression. Cambridge University Press, Cambridge
60. Silverman BW (1985) Some aspects of the spline smoothing approach to non-parametric regression curve fitting. J Roy Statist Soc Ser B 47: 1–52.
61. Su LJ (2012) Semiparametric GMM Estimation of Spatial Autoregressive Models. J Econometrics 167: 543–560
62. Su LJ, Jin SN (2010) Profile Quasi-Maximum Likelihood Estimation of Partially Linear Spatial Autoregressive Models. J Econometrics 157: 18–33
63. Wahba G (1983) Bayesian Confidence Intervals for the Cross Validated Smoothing Spline. J Roy Statist Soc Ser B 45: 133–150
64. Wand MP (2003) Smoothing and Mixed Models. Computational Statist 18: 223–249
65. Wood SN (2000) Modelling and smoothing parameter estimation with multiple quadratic penalties. J Roy Statist Soc Ser B 62: 413-428
66. Wood SN (2003) Thin Plate Regression Splines. J Roy Statist Soc Ser B 65: 95–114
67. Wood SN (2004) Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models. J Amer Statist Assoc 99: 673–686
68. Wood SN (2006a) Generalized Additive Models. An Introduction with R. London: Chapman and Hall
69. Wood SN (2006b) On Confidence Intervals for Generalized Additive Models Based on Penalized Regression Splines. Aust NZ J Statist 48: 445–464
70. Wood SN (2008) Fast Stable Direct Fitting and Smoothness Selection for Generalized Additive Models. J Roy Statist Soc Ser B 70: 495–518
71. Wood SN (2011) Fast Stable Restricted Maximum Likelihood and Marginal Likelihood Estimation of Semiparametric Generalized Linear Models. J Roy Statist Soc Ser B 73: 3–36
72. Wood SN, F Scheipl, Faraway JJ (2012) Straightforward Intermediate Rank Tensor Product Smoothing in Mixed Models. Statistics and Computing, doi: 10.1007/s11222-012-9314-z