Mínguez, R., Basile, R., & Durbán, M. (2020). An alternative semiparametric model for spatial panel data. *Statistical Methods & Applications*, 29 (4), pp. 669-708.

DOI: 10.1007/s10260-019-00492-8

An alternative semiparametric model for spatial panel data

Román Mínguez[1] · **Roberto Basile**[2] · **María Durbán**[3]

Abstract
We propose a semiparametric P-Spline model to deal with spatial panel data. This model includes a non-parametric spatio-temporal trend, a spatial lag of the dependent variable, and a time series autoregressive noise. Specifically, we consider a spatio-temporal ANOVA model, disaggregating the trend into spatial and temporal main effects, as well as second- and third-order interactions between them. Algorithms based on spatial anisotropic penalties are used to estimate all the parameters in a closed form without the need for multidimensional optimization. Monte Carlo simulations and an empirical analysis of regional unemployment in Italy show that our model represents a valid alternative to parametric methods aimed at disentangling strong and weak cross-sectional dependence when both spatial and temporal heterogeneity are smoothly distributed.

Keywords Spatial panel · Spatio-temporal trend · Mixed models · P-splines · PS-ANOVA

JEL Classification C33 · C14 · C63

1 Introduction

A 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 (Montero et al. 2012; Basile et al. 2014; Hoshino 2018). This approach combines penalized

✉ Román Mínguez
roman.minguez@uclm.es

[1] Department of Statistics, University of Castilla-La Mancha, Avenida de los Alfares 44, 16002 Cuenca, Spain

[2] Department of Industrial and Information Engineering and Economics, University of L'Aquila, Via Giuseppe Mezzanotte, Localitá Acquasanta 67100, Italy

[3] Department of Statistics, University Carlos III of Madrid , Avda Universidad 30, 28911 Leganés, Madrid, Spain

regression spline (PS) methods (Eilers et al. 2015) with standard spatial autoregressive models (such as SAR, SEM and SDM). 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.

In this paper, we propose a new alternative model for spatial panel data. This model is based on the P-Spline spatial autoregressive model (PS-SAR), extending it to deal with spatio-temporal data when both a large cross-section and a long time series 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 using a smooth interaction between the spatial coordinates and the time trend (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 us to answer a number of questions, including: How do unobserved time-related factors (i.e. common factors), such as economic-wide technological or demand shocks, heterogeneously affect the 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 a spatio-temporal trend may represent an alternative to parametric methods aimed at disentangling *common factor* effects (such as common business cycle effects) and *spatial dependence* effects (i.e. spatial spillover effects generated by local interactions between regions), where the former is sometimes regarded as 'strong' cross-sectional dependence and the latter as 'weak' cross-sectional dependence (Chudik et al. 2011; Ertur and Musolesi 2016).

Bai and Li (2013) and Shi and Lee (2018) propose the quasi-maximum likelihood method (QML) to estimate dynamic spatial panel data models with common shocks, thus accommodating both strong and weak cross-sectional dependence. Spatial correlations and common shocks are also considered by Pesaran and Tosetti (2011), but they specify the spatial autocorrelation on the idiosyncratic errors rather than on the dependent variable.[1] Bailey et al. (2016) and Vega and Elhorst (2016) propose a two-step and one-step approach, respectively, to address both forms of cross-sectional dependence.[2] All these approaches are parametric and do not allow practitioners to properly capture nonlinearities. On the other hand, Su and Jin (2012) examine the problem of estimating semiparametric panel data models with cross-sectional dependence, where the individual-specific regressors enter the model nonparametrically, and the common factors enter the model linearly, thus extending Pesaran (2006)'s common correlated

[1] Actually, the approach proposed by Pesaran and Tosetti (2011) does not explicitly allows for both forms of cross-sectional dependence (strong and weak). Rather they demonstrate that the CCE approach is still valid when in the DGP the errors contain both factors and a spatial-autoregressive terms.

[2] The two-step method proposed by Bailey et al. (2016) consists to model first common factors (e.g. aggregate shocks) using cross-sectional averages of the observations (thus following Pesaran 2006) and, then, to model the de-factored observations using spatial econometric techniques. In the one-step method proposed by Vega and Elhorst (2016) common factors and spatial dependence are modeled simultaneously. Another related article is of Han and Lee (2016), where the authors use a bayesian estimator.

effects (CCE) estimator to a semiparametric framework. Nevertheless, they do not take spatial contagion effects into account. By using on the PS-SAR model with spatio-temporal trends, we simultaneously handle four main econometric issues which are relevant when modeling spatio-temporal data; namely, *functional form bias*, *spatial dependence bias*, *spatial heterogeneity bias*, and *omitted time-related factors bias*.

The econometric model proposed here might appear complicated and computationally demanding, mainly because of the use of a three-dimensional smooth function for the spatio-temporal trend. Nevertheless, we employ an ANOVA decomposition of the spatio-temporal trend into several components (spatial and temporal main effects, and second- and third-order interactions between them), which gives further insights into the dynamics of the data. Thus, we use the acronym PS-ANOVA-SAR for the new data generating process (DGP) proposed here. Furthermore, we use a mixed model representation that allows us to apply the methods already developed in this area for estimation and inference, and to implement the necessary identifiability constraints in a straightforward manner. We also present an extension of the algorithm derived by Rodriguez-Alvarez et al. (2015) (for variance components estimation) to the PS-ANOVA-SAR model, which dramatically reduces the computational time needed to estimate the parameters in the model. Also, the use of nested B-spline bases (Lee et al. 2013) for the interaction components of the spatio-temporal trend contributes to the efficiency of the fitting procedure without compromising the goodness of fit of the model. Finally, we also consider an extension of the PS-ANOVA-SAR including a first-order time series autoregressive process (AR1) in the noise to accommodate residual serial correlation. All these models can be estimated, in a transparent and easy way for the potential user, using a new R package named *sptpsar*. This package is available in GitHub (https://github.com/rominsal/sptpsar) and can be installed in the usual way.[3]

We apply the PS-ANOVA-SAR(AR1) to regional unemployment data in Italy. As is well known, these data are characterized by spatial dependence, unobserved spatial heterogeneity, unobserved common effects, and time persistence. Substantive spatial dependence occurs due to interregional trade, labor migration and commuting, and knowledge spillovers; this spatial dependence can be captured by including spatial interaction effects in the model (Burridge and Gordon 1981; Molho 1995; Overman and Puga 2002; Patacchini and Zenou 2007). Unobserved spatial heterogeneity is mainly due to a strong North-South spatial trend which is largely uncaptured by the explanatory variables: regional unemployment rates register a substantial increase moving from North to South, reflecting the well-known regional development division within the country. A time-invariant smooth spatial trend could be used to filter out unobserved heterogeneity; thus, the spatial trend assumes the same role as the fixed regional effects used in parametric panel data models. Several unobserved common factors (e.g. aggregate demand shocks, aggregate technological shocks, and global labor market policies) may also affect the level of regional unemployment. The spatially heterogeneous effects of these factors may be the result, for instance, of region-specific technological or institutional constraints. These heterogeneous effects can be con-

[3] To install any R package from GitHub you need to have previously installed *devtools* package from CRAN. Then execute the commands *library(devtools)*, to load *devtools*, and *install github("rominsal/sptpsar")* to install *sptpsar* package.

trolled for either by using a model with a multifactor error structure (i.e. by adopting the CCEP approach developed by Pesaran 2006; Chudik et al. 2011; Pesaran and Tosetti 2011) or by including smooth interactions between the time trend and the spatial coordinates (i.e. by adopting a PS-ANOVA specification). Finally, time persistence in regional unemployment is usually (and properly) captured by using a dynamic formulation of the panel data model. In this paper, however, for the sake of simplicity, we adopt a static formulation of the PS-ANOVA-SAR model, and we control for residual serial autocorrelation by including an autoregressive (AR1) noise term. A dynamic development of PS-ANOVA-SAR model is in our future research agenda.

The plan of the paper is as follows. Section 2 provides a brief discussion of the main parametric spatial panel approaches used in the literature to capture strong and weak cross-sectional dependence. Section 3 sets out the PS-ANOVA-SAR(AR1) and discusses various technical aspects related to its identification and estimation. Section 4 reports the results of Monte Carlo experiments, while Sect. 5 discusses the results of the application of the model to regional unemployment data. Section 6 concludes by identifying important areas for extensions and further developments.

2 Parametric spatial panel approaches

Spatial spillover effects and common factors are of increasing empirical importance. Spatial spillovers are due to unobserved idiosyncratic shocks which propagate to all other regions with a distance-decay mechanism driven by network relationships. 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. There is a growing literature dedicated to the separate analysis of the two types of effect.

On the one hand, spatial spillover effects (weak cross-section dependence) can be analyzed by using, for example, the spatial autoregressive model with fixed effects (SAR-FE):

$$y_{it} = \alpha_i + \nu_t + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{1}$$

where y_{it} is a continuous response variable observed on the ith cross-sectional unit at time t for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$; $\mathbf{x}_{it}$ is a $k \times 1$ vector of explanatory variables, and $\boldsymbol{\beta}$ the associated set of coefficients; the nuisance parameters α_i capture unobserved time-invariant spatial heterogeneity (spatial fixed effects), while ν_t capture unobserved temporal heterogeneity, that is the effect of the omitted variables that are specific to each time period. α_i and ν_t are allowed to be correlated with $\mathbf{x}_{it}$, while the idiosyncratic errors, ε_{it}, are assumed to be independently distributed over $\mathbf{x}_{it}$. $W_N = (w_{ij,N})_{N\times N}$ is a spatial weights matrix whose diagonal elements $w_{ii,N}$ are 0. Thus, $\sum_{j=1}^{N} w_{ij,N} y_{jt}$ is the spatial lag of the dependent variable, and ρ the associated parameter. This model can be estimated using a quasi-maximum likelihood estimator (QMLE) (Elhorst 2014; Lee and Yu 2010).

Model (1) allows for interdependence among spatial units and corresponds to a long-run equilibrium relation between the response variable and its covariates. The spatial multiplier matrix,[4] $\mathbf{A}_N \equiv (\mathbf{I}_N - \rho\mathbf{W}_N)^{-1} \equiv \mathbf{I}_N + \rho\mathbf{W}_N + \rho^2\mathbf{W}_N^2 + \cdots$, in the reduced form of any SAR model pre-multiplies both observed and unobserved factors: the outcome in a location i will not only be affected by the exogenous characteristics of i, but also by those in any other location j through the inverse spatial transformation. The impact, therefore, is *global*. The powers of ρ matching the powers of $\mathbf{W}_N$ (higher orders of neighbors) ensure that a distance-decay effect is present. Thus, it is customary to distinguish between *direct*, *indirect* and *total* marginal effects. *Direct* effects measure the impact of a change in regressor x_k in region i on the outcome of said region: $\frac{\partial y_i}{\partial x_{ki}}$. Conversely, *indirect* (or spatial spillover) effects measure the impact of a change in regressor x_k in region j on the outcome of region i: $\frac{\partial y_i}{\partial x_{kj}}$. Total marginal effects are simply the sum of direct and indirect effects.

The problem with these effects is that, conditional on the model, both direct and indirect effects are specific to the pair of regions involved (i, j). Thus, average measures are typically used to summarize the results. In the SAR model, the average total marginal effect is computed as $\overline{M}_{tot}^{k} = (1 - \widehat{\rho})^{-1}\,\widehat{\beta}_k$. The average direct impact is $\overline{M}_{dir}^{k} = N^{-1} tr\left[(\mathbf{I}_N - \widehat{\rho}\mathbf{W}_N)^{-1}\,\mathbf{I}_N\widehat{\beta}_k\right]$, while the average indirect impact is $\overline{M}_{ind}^{k} = \overline{M}_{tot}^{k} - \overline{M}_{dir}^{k}$. In order to draw inference regarding the statistical significance of the average direct and indirect effects, LeSage and Pace (2009, p. 39) suggest simulating the distribution of the direct and indirect effects using the variance-covariance matrix implied by the maximum likelihood (ML) estimates. Efficient simulation approaches can be used to produce an empirical distribution of the parameters $\boldsymbol{\beta}, \boldsymbol{\theta}, \rho, \sigma^2$, which 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.

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:

$$\begin{aligned} y_{it} &= \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \gamma_i'\mathbf{f}_t + \varepsilon_{it} \\ \mathbf{x}_{it} &= \Gamma_i'\mathbf{f}_t + v_{it} \end{aligned} \tag{2}$$

where $\mathbf{f}_t$ is a $m \times 1$ vector of unobserved common factors (introduced to allow for cross-sectional dependence), and γ_i the corresponding heterogeneous response. $\mathbf{f}_t$ are allowed to be correlated with $\mathbf{x}_{it}$ (Γ_i is a $m \times k$ factor loading matrix), while the idiosyncratic errors, ε_{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 $\boldsymbol{\beta}$ parameters can be estimated using the so-called CCEP estimator, which can be viewed as a generalized fixed effects estimator.[5] The CCEP approach has been proved to be valid in the presence of both strong and weak (or semi-strong and semi-weak) cross dependence (Chudik et al.

[4] It is assumed than for $\mathbf{W}_N$ row-standardized, $|\rho| < 1$ so that this serial expansion holds.

[5] The assumption of fixed $\boldsymbol{\beta}$ parameters can be relaxed, and a random coefficient specification can be assumed: $\boldsymbol{\beta}_i = \boldsymbol{\beta} + u_i$, 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. We do not employ this extension in our analysis.

2011; Pesaran and Tosetti 2011). Thus, it can easily collect even the pure spatial spillover effects. However, economic analyses often require the assessment of the different forms of cross dependence, or better still, 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 (2013) and Shi and Lee (2018) consider a joint modeling of spatial interaction effects and common-shocks effects:

$$y_{it} = \alpha_i + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \mathbf{x}_{it}' \boldsymbol{\beta} + \gamma_i' \mathbf{f}_t + \varepsilon_{it} \tag{3}$$

This model allows one to test which type of effects (common shocks and/or spatial spillovers) is responsible for the cross-sectional dependence. Bai and Li (2013) and Shi and Lee (2018) use principle components to estimate common factors, while here we follow Pesaran (2006) in using cross-sectional averages of y_{it} and $\mathbf{x}_{it}$ as observable proxies for $\mathbf{f}_t$. Thus, we call it the SAR-CCEP model and use the QMLE to estimate it.

A drawback of this approach is worth noting. Specifically, there is a large number of incidental parameters involved in the estimation of the CCEP model, and thus also in the estimation of the SAR-CCEP, especially when the number of regressors is high and/or T is short. This is also documented in many empirical papers adopting the CCE approach, where estimates lack precision. In the following section, we propose an alternative semiparametric approach to disentangle common factor effects and spatial dependence effects which adequately addresses this problem.

3 Spatio-temporal semiparametric autoregressive models

In this section we propose a class of spatio-temporal models for large spatial panel data. They are a flexible alternative to the parametric models presented in Sect. 2 for modeling spatial economic data, as long as the spatio-temporal heterogeneity is smoothly distributed (as is commonly the case, we would argue, in empirical economic analyses), so that we can approximate it with smooth nonparametric functions. The models represent a generalization of the *Spatial Autoregressive Semiparametric Geoadditive Models* introduced by Basile et al. (2014) and Montero et al. (2012). We focus in this paper in the case of Gaussian response variables, however, the results here can be extended to variables with the exponential family of distributions in the context of Generalized Linear Mixed Models.

Let y_{it} be a sample of spatial panel data, where i is an index for the cross-sectional dimension (spatial units), with $i = 1, \ldots, N$, and t is an index for the time dimension (time periods), with $t = 1, \ldots, T$. The general model proposed is written as:

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \widetilde{f}(s_{1i}, s_{2i}, \tau_t) + \sum_{\delta=1}^{k} g_\delta(x_{\delta_{it}}) + \epsilon_{it}$$

where (s_{1i}, s_{2i}) are the spatial coordinates (latitude and longitude) of individual i (when i refers to areal units: municipality, provinces, etc., the standard convention here is to identify representative points for areal units, the most typical being areal centroids), τ_t is the time period, and $x_{\delta_{it}}$ are independent variables; W_{ij} are the spatial weights, and ρ the spatial autoregressive parameter. The functions $g_\delta(.)$ are parametric or nonparametric smooth functions of the covariates $x_{\delta_{it}}$ (they can be linear, or can accommodate varying coefficient terms, smooth interaction between covariates, by-factor smooth curves, and so on), and $\widetilde{f}(s_{1i}, s_{2i}, \tau_t)$ is an unknown nonparametric spatio-temporal trend. The idiosyncratic error term is assumed to follow an AR(1) process, i.e. $\epsilon_{it} = \phi\epsilon_{it-1} + u_{it}$ with $u_{it} \sim N(0, \sigma^2)$. Then, the model can be expressed in vector form as:

$$\mathbf{y} = \rho(\mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + \widetilde{f}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\tau}) + \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) + \boldsymbol{\epsilon}. \tag{4}$$

In many situations the spatio-temporal trend to be estimated by $\widetilde{f}$ can be complex, and the use of a single 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 $\widetilde{f}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\tau})$ where spatial and temporal main effects, and second- and third-order interactions between them can be identified:

$$\begin{aligned}\widetilde{f}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\tau}) = {} & f_1(\mathbf{s}_1) + f_2(\mathbf{s}_2) + f_t(\boldsymbol{\tau}) + f_{1,2}(\mathbf{s}_1, \mathbf{s}_2) \\ & + f_{1,t}(\mathbf{s}_1, \boldsymbol{\tau}) + f_{2,t}(\mathbf{s}_2, \boldsymbol{\tau}) + f_{1,2,t}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\tau}).\end{aligned}$$

Thus, model (4) can be written as:

$$\begin{aligned}\mathbf{y} = {} & \rho(\mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + f_1(\mathbf{s}_1) + f_2(\mathbf{s}_2) + f_t(\boldsymbol{\tau}) + f_{1,2}(\mathbf{s}_1, \mathbf{s}_2) \\ & + f_{1,t}(\mathbf{s}_1, \boldsymbol{\tau}) + f_{2,t}(\mathbf{s}_2, \boldsymbol{\tau}) + f_{1,2,t}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\tau}) + \sum_{\delta=1}^{k} g_\delta(\mathbf{x}_\delta) + \boldsymbol{\epsilon}.\end{aligned} \tag{5}$$

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 geoadditive terms given by $f_1(\mathbf{s}_1)$, $f_2(\mathbf{s}_2)$ and $f_{1,2}(\mathbf{s}_1, \mathbf{s}_2)$ (two 1d smooth functions of latitude and longitude and a 2d smooth function to account for any spatial effect not accounted for by the main effects) 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 make it possible 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 the use of individual regional dummies to capture unobserved heterogeneity, as long as such heterogeneity is smoothly distributed over space. Regional dummies peak at 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, $f_t(\boldsymbol{\tau})$, and the smooth interactions between space and time—$f_{1,t}(\mathbf{s}_1, \boldsymbol{\tau})$, $f_{2,t}(\mathbf{s}_2, \boldsymbol{\tau})$, and $f_{1,2,t}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\tau})$—work as control functions to capture the heterogeneous effect of common shocks. Thus, conditional on a smooth distribution of the spatio-temporal heterogeneity, the PS-ANOVA-SAR model works as an alternative to the models proposed by Bai and Li (2013), Shi and Lee (2018), Pesaran and Tosetti (2011), Bailey et al. (2016) and 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 ρ parameter). The advantage of the PS-ANOVA-SAR model lies in the fact that, contrary to the CCEP approach, its estimation does not involve a large number of incidental parameter.

Furthermore, our framework is also flexible enough to control for the linear and nonlinear functional relationships between the dependent variable and the covariates as well as the heterogeneous effects of these regressors across space (extending, in some sense, the work by Hoshino 2018, to the spatial panel case). The model inherits all the good properties of penalized regression splines, such as, coping with missing observations by appropriately weighting them, and straightforward interpolation of the smooth functions.

The smooth functions in the model above can be estimated using a number of different methods: splines, kernels, wavelets, and so on. However, most methods do not provide a unified approach whereby all terms and parameters in the model can be estimated simultaneously. We use the penalized spline approach introduced by Eilers and Marx (1996), which provides a general framework and can deal with the identifiability problems that appear in ANOVA decompositions.

3.1 Penalized splines methodology

The penalized regression approach is based on a basis representation of the unknown functions, which is combined with a penalty on the likelihood to control the wiggliness of the curve/surface. In particular, we use the approach introduced by Eilers and Marx (1996) which uses cubic B-splines (De Boor 1977) as basis functions, and second-order differences of adjacent coefficients as penalties. However, the methodology presented here can be applied to any basis and quadratic penalty. For the purpose of illustration, we will assume a simple case here, in which the mean of the response variable is an unknown function of a single covariate and the errors are independent:

$$y_i = f(x_i) + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2).$$

Then, the smooth function is represented by:

$$f(\mathbf{x}) = \sum_{j=1}^{c} B_j(\mathbf{x})\theta_j, \quad j = 1, \ldots, c, \tag{6}$$

with B_j a B-spline basis function, and θ_j is a component of a vector of regression coefficients of length c (the number of knots used to construct the basis). The size of the basis should be large enough (in general, between 4 and 40) to identify nonlinearities. The smoothness of the curve is controlled by a quadratic penalty term. Typically, this term is equivalent to an integral of squared second derivatives of the function, but sometimes (especially in the case of interactions) its calculation is not straightforward. Thus, following Eilers and Marx (1996), we use second-order differences among adjacent coefficients, yielding the following penalized regression problem:

$$\|\mathbf{y} - \mathbf{B}\boldsymbol{\theta}\|^2 + \lambda \sum_j (\Delta^2 \theta_j)^2 \tag{7}$$

where $\mathbf{B}$ is a matrix including B_j B-spline basis by columns and Δ^2 is the *difference* operator of order 2. In matrix form the penalty term becomes $\boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}$, with $\mathbf{P} = \lambda\mathbf{D}'\mathbf{D}$, $\mathbf{D}$ is the matrix of second-order differences and λ is the *smoothing parameter* which controls the amount of smoothing. The beauty of the P-spline methodology is that the penalty relaxes the dependence of the fit on the size of the basis or the degree of the polynomial. From (7) it is immediate to see that the estimated smooth function is given by:

$$\hat{f}(\mathbf{x}) = \mathbf{B}(\mathbf{B}'\mathbf{B} + \lambda\mathbf{D}'\mathbf{D})^{-1}\mathbf{B}\mathbf{y}.$$

Asymptotic properties of penalized spline estimators can be found in Claeskens et al. (2007) and Wood (2006).

When the response is an unknown function of two (or more) covariates, the representation of an interaction term is given by:

$$f_{1,2}(\mathbf{x}_1, \mathbf{x}_2) = \sum_{j=1}^{c_i} \sum_{l=1}^{c_k} B_j(\mathbf{x}_i) B_l(\mathbf{x}_k) \theta_{jl}, \text{ with } j = 1, \ldots, c_1 \text{ and } l = 1, \ldots, c_2, \tag{8}$$

where $B_j(\mathbf{x}_i)B_l(\mathbf{x}_k)$ is the tensor product of two marginal B-spline bases, and θ_{jl} is a vector of coefficients of length $c_1c_2 \times 1$. In this case the penalty is given by:

$$\mathbf{P} = \lambda_1 \mathbf{D}_1'\mathbf{D}_1 \otimes \mathbf{I} + \lambda_2 \mathbf{I} \otimes \mathbf{D}_2'\mathbf{D}_2.$$

See Currie et al. (2006) and Eilers et al. (2006) for further details on multidimensional smoothing with penalized splines. A property of this penalty is the fact that it allows for a separate amount of smoothing per covariate (*anisotropy*), which is an important issue when smoothing over space and time.

It is worth noting that the dimension of the matrices involved in interaction terms can increase very quickly if the size of the marginal bases is large, and so the estimation of the model can become very slow or even intractable. In order to reduce the computational burden without compromising the fit of the model, we follow Lee et al. (2013) in using *nested* B-spline bases for the interactions terms. The idea is to use a matrix $\breve{\mathbf{B}}$ in the interaction, such that the space spanned by this matrix is a subset of the space spanned by $\mathbf{B}$. The use of this *simpler* matrix for the construction of the interaction terms will not be a problem since the information about the interaction is generally sparse. In the ANOVA context, the main effects are more important than the interactions, so this would be reasonable in most situations. In order to ensure that the new basis is nested relative to the original basis, we assume that the number of knots ($\texttt{ndx}^*$) in $\breve{\mathbf{B}}$ is a divisor of the number of knots used to construct the original basis ($\texttt{ndx}$):

$$\texttt{ndx}^* \text{ of } \breve{\mathbf{B}} = \frac{\texttt{ndx} \text{ of } \mathbf{B}}{\texttt{div}} \Rightarrow \mathsf{span}(\breve{\mathbf{B}}) \subset \mathsf{span}(\mathbf{B}).$$

Then, the number of parameters is dramatically reduced but the model is still flexible enough to capture the complex space-time structure in the data. Lee et al. (2013) and Rodrìguez-Álvarez et al. (2018), among others, have carried out several simulation studies under different scenarios to test the impact of the use of reduced-rank basis to represent interaction terms. They showed that a moderate reduction of the number of knots (half for a two-way interaction, etc.) has no significant loss in terms of the fit to the data, and produces remarkable results in computing time that would be unbearable otherwise.

Earlier in this section we could see that the estimate of the unknown function depends on the smoothing parameter λ, thus optimal selection of this parameter is needed. Leave-one-out cross-validation (CV or GCV) provides a mechanism to choose this parameter. However, this approach has a tendency to under-smooth the data, specially if errors are correlated. An alternative way to select the optimal smoothing parameter is through reparameterization of the penalized spline model as a mixed model (Currie and Durbán 2002). This idea comes from the fact that the minimization problem in (7) is the minimization criterion in a random effects model of the form:

$$\mathbf{y} = \mathbf{B}\boldsymbol{\theta}, \quad \boldsymbol{\theta} \sim N(0, \sigma_\theta^2 \mathbf{P}^{-1}) \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I}),$$

but the penalty matrix $\mathbf{P}$ is singular (2 eigenvalues are zero), and so it has a degenerate distribution. One possible solution is to rewrite the model as $\mathbf{B}\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$, such that the 2 columns of $\mathbf{X}$ span the polynomial null space of $\mathbf{P}$ and the columns of Z span its complement:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2), \quad \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}),$$

where matrices $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{G}$ are computed using an appropriate decomposition on the penalty term (see "Appendix A" for details). This approach will allow us the simultaneous estimation of all parameters in the model, and the use of existing algorithms in the mixed model framework.

Model (5) is non-identifiable since, for example, $f_1(\mathbf{s}_1)$ and $f_2(\mathbf{s}_2)$ are confounded with $f_{1,2}(\mathbf{s}_1, \mathbf{s}_2)$ (as in a three-way ANOVA, lower order effect are confounded with higher order interactions). An elegant way to construct an identifiable model is to impose the constraints used in a factorial design to the coefficients of the B-splines basis, i.e. $\sum_j \theta_j = 0$ for 1d smooth functions (see (7)) and $\sum_j \theta_{jl} = \sum_l \theta_{jl} = 0$ in the case of 2d smooth functions (see 8). It is easy to show that this is equivalent to removing from the matrices of fixed effects (given in the mixed model reparametrization above) the columns that are repeated. Furthermore, the mixed model reparametrization will also provide an automatic selection of the smoothing parameter, since the latter becomes a ratio of variances, meaning that it is no longer necessary to estimate λ via a cross-validation method or an information criterion.

3.2 PS-ANOVA-SAR(AR1) model as a mixed model

We propose the estimation of the PS-SAR-ANOVA model (with temporal autoregressive dependence in the noise) by means of its representation as a mixed model. This allows us to use the standard mixed model methodology, which is well known and tested, and is implemented in most statistical software. For the sake of simplicity we assume that there are no covariates in the model (the inclusion of covariates is an immediate step); then model (5), including time autoregressive errors, becomes:

$$\begin{aligned}(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y} = & f_1(\mathbf{s}_1) + f_2(\mathbf{s}_2) + f_t(\boldsymbol{\tau}) + f_{1,2}(\mathbf{s}_1, \mathbf{s}_2) \\ & + f_{1,t}(\mathbf{s}_1, \boldsymbol{\tau}) + f_{2,t}(\mathbf{s}_2, \boldsymbol{\tau}) + f_{1,2,t}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\tau}) + \boldsymbol{\epsilon}\end{aligned} \tag{9}$$

where

$$\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R}), \quad \mathbf{A}_N = \mathbf{I}_N - \rho \mathbf{W}_N, \quad \mathbf{R} = \frac{\sigma^2}{1-\phi^2}(\mathbf{I}_N \otimes \boldsymbol{\Omega})$$

and

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{T-1} \\ \phi & 1 & \phi & \cdots & \phi^{T-2} \\ \phi^2 & \phi & 1 & \cdots & \phi^{T-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \cdots & 1 \end{pmatrix}$$

Again, we can use the reparameterization of this type of model into a mixed model (see "Appendix B" for details) to obtain:

$$(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R}), \quad \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{G})$$

where $\mathbf{G}$ is a diagonal matrix which depends on several variance components $\sigma^2_{\nu_j}$ (one per univariate smooth function, two per bivariate function, and so on). This implies that each function is estimated with a separate amount of smoothing.

Conditional on the correlation parameters and variance components, the estimates of the coefficients $\boldsymbol{\beta}$ and of the random effects $\boldsymbol{\alpha}$ follow from the standard mixed model theory (Searle et al. 1992), and are the solution of the system of equations:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y} \end{bmatrix}. \tag{10}$$

Variance components (and, therefore, smoothing parameters), and correlation parameters may be estimated by maximizing the residual log-likelihood (REML), as in Patterson and Thompson (1971):

$$\begin{aligned}\ell(\sigma^2_{\tau_i}, \sigma^2, \rho, \phi) = &-\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \\ &- \frac{1}{2}\left[(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y}\right]'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}) \\ &\times \left[(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y}\right] \\ &+ \log|\mathbf{A}_N \otimes \mathbf{I}_T| \end{aligned} \tag{11}$$

where $\mathbf{V} = \mathbf{R} + \mathbf{ZGZ}'$.

The numerous variance components (3 for the main effects of latitude, longitude and time, 6 for the 3 two-way interactions, and 3 for the three-way interaction) and correlation parameters (ρ and ϕ in our case) make the maximization of this function using numerical methods very time consuming. Rodriguez-Alvarez et al. (2015) recently developed an algorithm named SAP (Separation of Anisotropic Penalties) that maximizes REML much faster than numerical optimization procedures can. This algorithm can be applied to any mixed model (or models in which mixed model based inference is used) in which the precision matrix of random effects can be expressed as a linear combination over the variance components. SAP is computationally very efficient and the estimates produced match those obtained using REML. In "Appendix B", we show how the SAP algorithm can also be adapted to estimate the correlation parameters in the PS-ANOVA-SAR(AR1) model.

Estimated curves or interaction surfaces in the model can be easily obtained from Eq. (10). For example, the estimated effect of the spatial term $f_{1,2}(\mathbf{s}_1, \mathbf{s}_2)$ is:

$$\hat{f}_{1,2}(\mathbf{s}_1, \mathbf{s}_2) = \mathbf{E}_{1,2}\mathbf{C}^{-1}[\mathbf{X} : \mathbf{Z}]'(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y},$$

where $\mathbf{C}$ is the matrix on the left-hand side of (10) and $\mathbf{E}_{1,2}$ is a diagonal matrix with ones in the diagonal positions corresponding to the interaction smooth term.

The mixed model approach to penalized splines smoothing can also be interpreted from an empirical Bayes viewpoint in which fixed effects have (constant) non-informative prior, and we use this approach for variance estimation. In this framework it is easy to show (see Lin and Zhang (1999)) that the covariance matrix of the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ is given by $Cov(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{C}^{-1}$, and so, again in the case of the space interaction term:

$$\text{Var}(\hat{f}_{1,2}) = \mathbf{H}_{1,2} = \mathbf{E}_{1,2}\,[\mathbf{X} : \mathbf{Z}]\,\mathbf{C}^{-1}\,[\mathbf{X} : \mathbf{Z}]'\,\mathbf{E}'_{1,2}$$

Also, the trace of $\mathbf{H}_{1,2}$ can be used as a measure of the complexity of the smooth surface (as in the linear model) by defining the effective dimension of the smooth term as:

$$e.d.(\hat{f}_{1,2}) = trace(\mathbf{H}_{1,2}).$$

4 Monte Carlo experiments

We propose a Monte Carlo experiment in order to compare the small sample properties of different parametric and semiparametric estimators (namely CCEP, SAR-CCEP, PS-ANOVA, PS-ANOVA-SAR, and PS-ANOVA-SAR-AR1) in the presence of weak and strong cross-sectional dependence. The aim of this study is to assess whether the PS-ANOVA-SAR-AR1 represents a valid alternative to the SAR-CCEP estimator in the presence of spatial dependence, spatial heterogeneity and unobserved common factors. To fulfill this aim, we set two different data generating processes (DGPs).

4.1 DGP 1: SAR-CCEP

The first DGP is a modified version of the one proposed by Pesaran (2006) and Pesaran and Tosetti (2011), where the dependent variable and the individual-specific regressors are assumed to depend on a linear combination of unobserved common factors. We extend this DGP by including a spatial lag of the dependent variable to combine the two sources of cross-sectional dependence, i.e. common factors and spatial dependence as in Bailey et al. (2016) and Vega and Elhorst (2016):

$$
\begin{aligned}
[DGP1] \quad y_{it} &= \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \gamma_{i1} f_{1t} + \gamma_{i2} f_{2t} + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + u_{it} \\
u_{it} &= \phi u_{i,t-1} + \varepsilon_{it} \\
x_{jit} &= a_{ji} + \gamma_{ji1} f_{1t} + \gamma_{ji3} f_{3t} + v_{jit} \quad j = 1, 2
\end{aligned}
$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.[6]

We always assume homogeneous slopes, i.e. $\beta = (1, 1)'$ for x_{1it} and x_{2it}, and serially correlated errors, i.e. u_{it} are generated as stationary AR(1) processes with $\phi = 0.75$, $\varepsilon_{it} \sim \text{iid}N(0, \sigma^2)$ and $\sigma^2 \sim \text{iid}U[0.5, 1.5]$.

In [DGP 1], f_{1t}, f_{2t} and f_{3t} are unobserved common effects following first-order temporal autoregressive processes,

$$
\begin{aligned}
f_{jt} &= 0.5 f_{j,t-1} + v_{f_{jt}} \quad j = 1, 2, 3 \\
v_{f_{jt}} &\sim \text{iid}N(0, 0.75) \quad f_{j0} = 0
\end{aligned}
$$

and the idiosyncratic noises for the individual-specific regressors are generated as

$$
\begin{aligned}
v_{jit} &= \phi_{ji} v_{ji,t-1} + \varsigma_{jit} \\
\varsigma_{jit} &\sim N(0, 1 - \phi_{ji}^2) \\
\phi_{ji} &\sim \text{iid}U[0.05, 0.95]
\end{aligned}
$$

For all serially correlated errors, the first 50 observations are discarded.

[6] All variables involved, both observable and latent, are stationary in the simulations. The analysis of the statistical properties of the proposed estimator under the assumption of nonstationarity is a subject of current research.

The heterogeneous intercepts are generated as

$$\alpha_i \sim \text{iid}N(1, 1)$$
$$(a_{1i}, a_{2i}) \sim \text{iid}N(0.5\boldsymbol{\tau}_2, 0.5\mathbf{I}_2)$$

where $\boldsymbol{\tau}_2 = (1, 1)'$ and $\mathbf{I}_2$ is a 2×2 identity matrix.

The factor loadings associated with the unobserved common effects in the x_{jit} and y_{it} equations are simulated from

$$\begin{pmatrix} \gamma_{1i1} & 0 & \gamma_{1i3} \\ \gamma_{2i1} & 0 & \gamma_{2i3} \end{pmatrix} \sim \text{iid} \begin{pmatrix} N(0.5, 0.5) & 0 & N(0, 0.5) \\ N(0, 0.5) & 0 & N(0.5, 0.5) \end{pmatrix}$$

and

$$\gamma_{i1} \sim \text{iid}N(1, 0.2)$$
$$\gamma_{i2} \sim \text{iid}N(1, 0.2)$$

Finally, ρ is the spatial autoregressive coefficient associated to the spatial lag of the dependent variable, and w_{ij}, for $i, j = 1, \ldots, N$, are elements of a spatial weight matrix $\mathbf{W}$, assumed to be time-invariant. To construct this matrix, we have simulated random spatial coordinates, s_{1i}, s_{2i}, following iid$U[0, 1]$ distributions. The neighborhood criterion corresponds to the five closest observations and the matrix $\mathbf{W}$ is row-normalized. Prior to this, we set up two values of the spatial coefficient, $\rho = (0, 0.5)$, in order to consider both independence and sizeable spatial dependence.

The reader may note that, in line with Pesaran (2006) and Pesaran and Tosetti (2011), we always consider an autoregressive, but stationary, process for the generation of common factors ($f_{jt} = 0.5 f_{j,t-1} + v_{f_{jt}}$). Thus, we exclude both the case of a pure random distribution of common factors (e.g. $f_{jt} \sim \text{iid}N(1, 1)$) and the case of a nonstationary process (i.e. $f_{jt} = f_{j,t-1} + v_{f_{jt}}$). Under the assumption of totally random common shocks (e.g. a random perturbation of world demand), the CCEP estimator would certainly remain strongly consistent. The PS-ANOVA-SAR approach (Eq. 4) can also be adapted to this situation by introducing random time effects (as in Perperoglou and Eilers 2009), or interactions between a smooth spatial trend and time fixed effects (i.e. $f(s_{1i}, s_{2i}) \times \tau_t$). Nevertheless, since in many situations the common shocks are persistent over time, we prefer to allow for an autoregressive process. Moreover, for the sake of simplicity, we do not extend the analysis to the nonstationarity case. Indeed, all variables involved, both observable and latent (including common factors), are stationary in the simulations. The Pesaran's multifactor error structure approach has already been extended to the cases where the unobservable common factors follow unit root processes. In particular, the Monte Carlo study conducted by Kapetanios et al. (2011) show that the CCE estimator is robust to the case where unobserved common factors are integrated of order 1. The analysis of the statistical properties of the proposed PS-ANOVA-SAR estimator under the assumption of nonstationarity is a subject of current research.

4.2 DGP 2: PS-ANOVA-SAR(AR1)

The second DGP includes smooth spatio-temporal trends in place of heterogeneous intercepts and unobserved common factors. In other words, time-invariant spatial unobserved heterogeneity and time-varying unobserved heterogeneity (due to common factors) are assumed to be smoothly distributed (over time and space) and generated by a specific spatio-temporal trend for each variable (in this DGP, we only consider an individual-specific regressor), that is,

$$
\begin{aligned}
[DGP2] \quad y_{it} &= \alpha + \beta x_{it} + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + f_{spt} + u_{it} \\
u_{it} &= \phi u_{i,t-1} + \varepsilon_{it} \\
x_{it} &= h_{spt} + \epsilon_{it}
\end{aligned}
$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

α is a constant intercept and both $w_{ij,N}$ and u_{it} are generated as in [DGP 1].

The terms f_{spt} and h_{spt} represent spatio-temporal trends including main effects and interaction effects of second- and third-order between spatial and temporal coordinates. They have been generated by the following nonlinear terms:

$$
\begin{aligned}
f_{1spt} &= f_{s_{1i}} + f_{s_{2i}} + f_{\tau_t} + f_{s_{1i},s_{2i}} + f_{s_{1i},\tau_t} + f_{s_{2i},\tau_t} + f_{s_{1i},s_{2i},\tau_t} \\
f_{s_{1i}} &= \sin(2\pi s_{1i}) \\
f_{s_{2i}} &= \cos(3\pi s_{2i}) \\
f_{\tau_t} &= \sin(4\pi \tau_t) \\
f_{s_{1i},s_{2i}} &= \sin(2\pi s_{1i})(s_{2i}^2 - 1) \\
f_{s_{1i},\tau_t} &= 2\sin(2\pi s_{1i}^2) cos(2\tau_t) \\
f_{s_{2i},\tau_t} &= cos(3\pi s_{2i}) sin(3\tau_t^2) \\
f_{s_{1i},s_{2i},\tau_t} &= sin(2\pi s_{1i}) s_{2i}^2 cos(3\tau_t) \\
h_{spt} &= h_{s_{1i},s_{2i}} + h_{\tau_t} \\
h_{s_{1i},s_{2i}} &= 2\cos(3\pi s_{2i})(2 s_{1i} s_{2i}) \\
h_{\tau_t} &= 2\sin(3\pi \tau_t^3)
\end{aligned}
$$

As in [DGP 1], the spatial coordinates, s_{1i}, s_{2i}, follow iid$U[0, 1]$ distributions, while the temporal coordinate is generated as $\tau_t = t/T$, for $t = 1, \ldots, T$. Finally, the noise term of x_{jit} is generated by the following stochastic process:

$$
\begin{aligned}
\epsilon_{it} &= \nu_i \epsilon_{i,t-1} + \varsigma_{it} \\
\varsigma_{it} &\sim N(0, 1 - \phi_i^2) \\
\phi_i &\sim iidU[0.05, 0.95]
\end{aligned}
$$

Table 1 Spatio-temporal competing models

Model I	Model with unobserved common effects (CCEP)
Model II	SAR model with unobserved common effects (SAR-CCEP)
Model III	Spatio-temporal model (PS-ANOVA)
Model IV	Spatio-temporal SAR model (PS-ANOVA-SAR)
Model V	Spatio-temporal SAR-AR(1) model (PS-ANOVA-SAR-AR1)

4.3 Monte Carlo results

For both DGPs, the generated sample panels have $N = 100$ and $T = 30$, and 300 simulations have been used for each value of ρ. For each simulated DGP, we assess the small sample properties of five parametric and semiparametric estimators (Table 1). In particular, we computed the parametric pooled common correlated estimator (CCEP) described in Pesaran (2006) (Eq. 2 in Sect. 2), and its extension (SAR-CCEP) which includes a spatial lag of the dependent variable (Vega and Elhorst 2016) (Eq. 3 in Sect. 2); plus three versions of the semiparametric model described in Sect. 3 (Eq. 4) with the ANOVA decomposition for the spatio-temporal trend (imposing linearity for the effect of the covariates), namely the PS-ANOVA (which excludes the spatial lag term $\mathbf{Wy}$), the PS-ANOVA-SAR (which excludes serial correlation in the error term), and the PS-ANOVA-SAR-AR(1). For all PS-ANOVA specifications, B-spline bases have been included and we have chosen 16 knots for both time and spatial coordinates. Furthermore, we have nested the bases for second- and third-order interaction terms dividing the number of knots by 2 and 4, respectively (as explained in Sect. 3.1). Finally, all the parameters of PS-ANOVA specifications have been estimated using REML applied to the PS-ANOVA models previously reparameterized as mixed models (details in Sect. 3.2).

Table 2 provides estimates of the bias and of the root-mean-square error (RMSE) for parameters β_1, β_2, and ρ over the two DGPs.

The main conclusions can be synthesized as follows:

1. With respect to the estimates of the parameter β_1, the five estimators perform quite differently, depending on the DGP and on the value of the ρ parameter. More specifically, for [DGP 1] and $\rho = 0$, the estimators of the parametric specifications (I and II) are substantially unbiased. Conversely, the estimators of the semiparametric specifications (III to V) are slightly positively biased, and their RMSE is approximately double that of specifications I and II.
   With $\rho = 0.5$, the relative performance of the different estimators changes a great deal. The estimators of the spatial lag models (II, IV and V) are substantially unbiased, with comparatively similar values of RMSE. As expected, the estimators of the models without a spatial lag (I and III) show a positive and non-negligible bias in the estimation of the slope parameter.
   Semiparametric models (III, IV and V) prove their superiority in estimating the parameter β_1 when $\rho = 0.5$ in [DGP 2], which includes smooth spatio-temporal trends in place of unobserved common factors and spatial fixed effects. The esti-

Table 2 Bias and root-mean-square-error (RMSE)

	Model	I	II	III	IV	V
β_1						
DGP 1						
$\rho = 0$	Bias	0.0018	0.0015	0.0440	0.0387	0.0515
	RMSE	0.0285	0.0284	0.0557	0.0500	0.0608
$\rho = 0.5$	Bias	0.0512	0.0026	0.1047	0.0144	0.0088
	RMSE	0.0613	0.0287	0.1198	0.0333	0.0386
DGP 2						
$\rho = 0$	Bias	0.0005	0.0062	0.0027	−0.0112	0.0029
	RMSE	0.0395	0.0346	0.0274	0.0319	0.0160
$\rho = 0.5$	Bias	0.0485	−0.0256	0.0024	0.0063	0.0024
	RMSE	0.0840	0.0435	0.0363	0.0280	0.0162
β_2						
DGP 1						
$\rho = 0$	Bias	−0.0006	−0.0010	−0.0034	−0.0011	−0.0200
	RMSE	0.0262	0.0264	0.0327	0.0306	0.0396
$\rho = 0.5$	Bias	0.0497	0.0008	0.0310	−0.0120	−0.0505
	RMSE	0.0632	0.0305	0.0630	0.0322	0.0627
ρ						
DGP 1						
$\rho = 0$	Bias		−0.0125		0.1903	0.3666
	RMSE		0.0374		0.2053	0.3750
$\rho = 0.5$	Bias		−0.0137		0.1318	0.2220
	RMSE		0.0286		0.1404	0.2320
DGP 2						
$\rho = 0$	Bias		0.4311		−0.2074	0.0063
	RMSE		0.4353		0.2150	0.0224
$\rho = 0.5$	Bias		0.2589		−0.1454	−0.0006
	RMSE		0.2606		0.1512	0.0132

Fixed values of β_1 and β_2 are (1,1) and true values of ρ and ϕ are (0,0.5) and 0.75, respectively

mator of model III is also unbiased when $\rho = 0.5$ although this model does not include a spatial lag (nevertheless, its RMSE is bigger than that computed for estimators of models IV and V). Not surprisingly, the best results are achieved by the estimator of model V, since it is the specification closest to [DGP 2]. Moreover, considering that in [DGP 2] the spatio-temporal trend is different for the dependent variable y_{it} and the regressor x_{it}, semiparametric models (III, IV and V) produce unbiased estimates for the parameters despite the existence of complex unobserved spatio-temporal trends with the only requirement of smoothness.[7]

[7] Following the suggestion of an anonymous referee, we have also simulated (only for DGP2) the CCEP and SAR-CCEP specifications including individual time trends. The results are similar to those reported in Table 2 and are available upon request.

When examining the estimates for parameter β_2, the conclusions for [DGP 1] are similar to those reported for β_1 except for the unexpectedly poor performance of the estimator of model V when $\rho = 0.5$. On the other hand, the assessment of model III is better for parameter β_2 than for β_1. We believe that this behaviour could be explained by the difficulties of PS-ANOVA models to identify which part of the variability corresponds to spatio-temporal trends and which part to other regressors when the DGP is far from the assumptions made for these specifications. This lack of identifiability is aggravated when the specification includes more terms (like V) and tends to be more unstable and likely overparameterized for [DGP 1].

2. As for the estimates of the ρ parameter, the results clearly show that, depending on the DGP considered, either the parametric or the semiparametric estimators are unbiased. Specifically, for [DGP 1] only model III (SAR-CCEP) produces unbiased estimates, while the estimators of models IV and V are positively biased. The size of the bias for the estimators of models IV and V decreases as the true value of ρ increases. This evidence attests to the fact that the smooth-based estimator ”prefers” diffused spatial dependence, so that cross-sectional smoothness increases.

   This situation is reversed when the data are generated using [DGP 2]. In this case, the estimator of model III is positively biased with greater bias and RMSE than in models IV and V in [DGP 1]. When assessing semiparametric models IV and V for [DGP 2], the negative bias of the estimator of model IV is noteworthy. In fact, model V is the only unbiased estimator which is not surprising given that [DGP 2] includes a first-order temporal autoregressive term in the noise. Again, the bias for the estimators of models III and IV decreases as the value of ρ increases.

   Summing up, as also pointed out by one of the anonymous referees, the results of the simulations suggest that SAR-CCEP and PS-ANOVA-SAR remain different tools for different jobs. Nevertheless, they also suggest that, in the presence of spatial dependence, spatial unobserved heterogeneity, and time-related unobserved factors, the semiparametric estimator proposed in this paper represents a valid alternative to parametric estimators proposed in the literature (CCEP and SAR-CCEP) to estimate the slope parameters as long as both types of unobserved effects are smoothly distributed over time and space. In practical terms, we may say that the CCEP estimator can be a better choice in cross-country or cross-industry analyses to deal with strong cross-sectional dependence problems generated by any kind of common (international or inter-industry) shocks. The PS-ANOVA estimator can offer a good alternative in cross-regional analyses, especially when the spatial unit is relatively small (e.g. provinces, cities, metropolitan areas, and so on), to deal with strong cross-sectional dependence generated by any kind of common shocks (technological shock or demand shocks) propagating smoothly over space and time.

5 Empirical case

Ever since Partridge and Rickman (1997) and Taylor and Bradley (1997), regional unemployment differentials have been the subject of intensive research in the literature.

Recent contributions apply spatial econometric models both in a cross-sectional setting (Molho 1995; Aragon et al. 2003; Cracolici et al. 2007) and in a spatial panel framework (Lottmann 2012; Basile et al. 2012; Ríos 2014). Here, we analyze the performance of the PS-ANOVA-SAR(AR1) model compared to different competing parametric and semiparametric models using panel data on regional unemployment in Italy. We first describe these data and their features in terms of spatial and temporal trends (Sect. 5.1). Then, we briefly discuss the theoretical background and the set of variables used to explain regional unemployment differentials (Sect. 5.2). Finally, we report the results of the econometric analysis (Sect. 5.3).

5.1 Regional unemployment data

The data on regional unemployment rates ($unrate_{i,t}$) for each Italian province $i = 1, \ldots, N$ (N=103) (corresponding to Italian NUTS-3 regions) and for each time period $t = 1996, \ldots, 2014$ (T=19) used in this analysis are provided online by the Italian National Institute of Statistics (ISTAT). They are defined as $unrate_{i,t} = 100 \times \frac{U_{i,t}}{LF_{i,t}}$, where $U_{i,t}$ is the number of unemployed people and $LF_{i,t}$ is the total labor force.

Regional unemployment rates differ widely in Italy, especially between northern and southern provinces. The North-South division can be depicted by mapping the predicted values of a simple regression of provincial unemployment rates on the smooth interaction between longitude and latitude (Fig. 1). A clear spatial trend emerges and is persistent over time. These findings suggest that the nature of regional unemployment disparities in Italy is the result of a long-run equilibrium rather than a short-term disequilibrium caused by temporary shocks. As such, as Marston (1985) points out, "If unemployment is of equilibrium nature, any policy oriented to reduce regional disparities is useless since it cannot reduce unemployment anywhere for long". Nevertheless, we cannot exclude the possibility that the strong persistence of regional unemployment differentials is caused by both structural problems in the economy and the inability of Italian regions to absorb specific shocks (on either the demand or the supply side).

A nonlinear time trend also characterizes unemployment data. The national unemployment rate (red line in Fig. 2) shows a fall from 1996 (11.2%) to 2007 (6.1%); however, with the outbreak of the financial crisis, and its subsequent impact on the productive economy, there was an upturn in the unemployment rate, reaching 12.7% in 2014. Both northern and southern provinces followed a similar time path, suggesting that common business cycle factors affect all the regions. However, there are relevant differences across provinces, indicating that regions may differ in their elasticity to common shocks. This feature is rather usual in regional unemployment studies. Thus, in order to obtain coefficients of the determinants that measure their impact on regional unemployment rates net of aggregate cyclical factors, these studies adopt one of two main approaches. The first is to include time-period fixed effects in the model (Elhorst 1995; Partridge and Rickman 1997). However, this is a homogeneous approach since it assumes that the impact of common factors is the same across regions, whereas it is often the case in applied settings that certain regions are found to be more sensitive

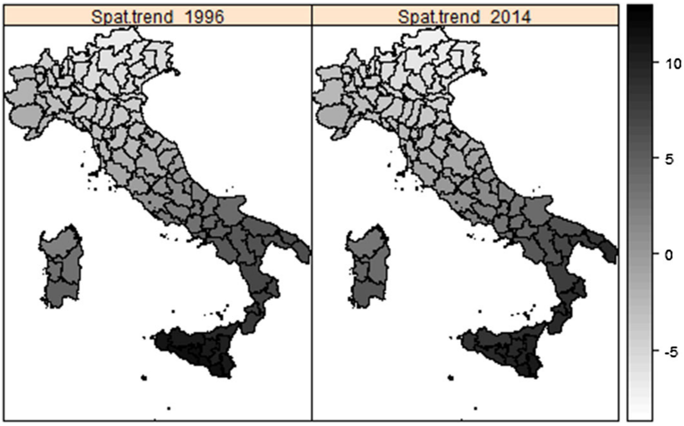
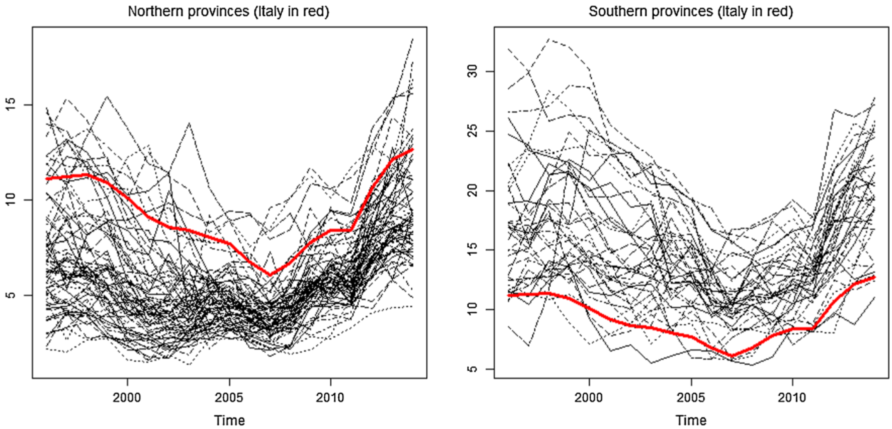


Fig. 1 Spatial trend of provincial unemployment rates. To simulate a continuous trend, shades of grey are assigned to each percentile. Darker regions are those with higher unemployment rates



(a) North **(b)** South

Fig. 2 Time trend of provincial unemployment rates: 1996–2014

than others to aggregate fluctuations. The alternative approach is to take the difference between the regional and national unemployment rates as a way to appraise dispersion and factor out country-specific dynamics (Thirlwall 1966; Blanchard et al. 1992; Decressin and Fatas 1995). This ‘factoring out’ of aggregate cyclical factors also bears a clear resemblance to the common factor approach proposed in Pesaran (2007), where common factors are modeled by cross-sectional averages of the variables at each point in time.

The presence of common cyclical factors is expected to generate significant cross-sectional correlation in the data. This hypothesis can be assessed using the CD test developed by Pesaran (2004, 2015). This test uses the pair-wise correlation coefficients between the time-series for each panel unit. The CD statistics computed on our sample of regional unemployment rates is highly significant, confirming the existence of

Table 3 Cross-sectional dependence test

	Without controlling for serial correlation	Controlling for serial correlation
Unemployment rate (*unrate*)	180.7***	87.4***
Employment growth rate (*enpgrow*)	54.1***	58.7***
Participation rate (*partrate*)	36.8***	25.9***
Agriculture (*agri*)	238.8***	106.9***
Industry (*ind*)	248.9***	262.0***
Construction (*cons*)	69.7***	28.6***
Services (*serv*)	185.8***	22.2***

***; ** and * significance at 1%, 5% and 10%, respectively

cross-sectional dependence (Table 3). Applying the same test on the residuals of an AR(2) model (to accommodate serial correlation), we obtain a CD value of 87.4, still highly significant. Moreover, the result of the estimation of the exponent of cross-section dependence (introduced by Bailey et al. 2016) provides clear evidence of strong cross-sectional dependence (the parameter is equal to 1). This result has relevant implications in terms of econometric modeling, underlining the importance of using a proper approach to control for strong cross-sectional dependence (either the CCEP or the PS-ANOVA approach).

Nevertheless, net of the effect of common factors (-'strong'-cross-sectional dependence), significant cross-sectional correlation in the data could also be generated by spatial autocorrelation (-'weak'-cross-sectional dependence). From a theoretical point of view, spatial autocorrelation in regional unemployment rates can be justified on the basis of a framework which builds on Blanchard et al. (1992) regional labor market model, including neighboring effects due to interregional trade, migration, and knowledge spillovers (Zeilstra and Elhorst 2014). Starting from a steady-state pattern of regional unemployment, a region-specific shock not only affects the respective labor market, but also spills over to neighboring regions. Given this interdependence, the resulted changes in unemployment in neighboring areas may spill over again to adjacent labor markets, including the location where the shock originated. This implies that the unemployment rate of a particular region is affected not only by its own labor market characteristics, but also by the labor market performance of all other regions. Thus, in principle, we cannot exclude the possibility that potential sources of interaction between regions are both weak -due to, for example commuting flows-, and strong -due to common factors. A joint modeling of weak and strong cross-sectional dependence is, therefore, needed.

Finally, another important issue is the assessment of the stationarity of regional unemployment data. To this end, we use a panel unit root test proposed by Pesaran (2007) which is robust against cross-sectional dependence. This test clearly rejects the hypothesis of a unit root at all reasonable significance levels (Table 4). Hence, these results give a strong indication regarding the stationarity of the data once cross-sectional dependence is taken into account.

Table 4 IPS panel unit root tests robust against cross-sectional dependence

	None	Drift	Drift and trend
Unemployment rate (*unrate*)	– 1.462*	– 2.387***	– 10.109***
Employment growth rate (*enpgrow*)	– 1.871***	– 2.581***	– 2.700**
Participation rate (*partrate*)	– 1.309	– 1.905	– 8.974***
Agriculture (*agri*)	– 1.713***	– 2.670***	– 10.550***
Industry (*ind*)	– 1.830***	– 2.47***	– 8.955***
Construction (*cons*)	– 1.786***	– 1.899	– 5.274***
Services (*serv*)	– 1.379	– 1.408	– 2.454

Deterministic components: none, drift, drift and trend
***; ** and * significance at 1%, 5% and 10%, respectively

Table 5 Summary statistics

Variable	Min	1st qu.	Median	Mean	3rd qu.	Max	SD
unrate	1.33	4.78	7.71	9.23	12.50	32.72	5.65
empgrowth	– 14.68	– 1.54	0.40	0.37	2.27	13.91	3.21
agri	0.05	3.52	6.80	7.77	10.99	30.57	5.30
ind	5.54	13.57	20.53	21.32	27.96	46.33	9.22
cons	3.59	6.71	7.75	7.83	8.77	14.68	1.64
serv	45.09	58.75	63.96	64.27	69.51	86.09	7.88
partrate	27.04	37.73	42.85	41.52	45.16	53.19	4.70

5.2 Observable and unobservable determinants

The unemployment rate can be considered as a reduced form function of a variety of factors affecting labor demand, labor supply and wages.[8] According to the pioneering work of Partridge and Rickman (1997), these factors can be broadly categorized as either disequilibrium factors (e.g. employment growth rates) or market equilibrium factors (e.g. industry and services shares, demographic variables and amenities). For the choice of the actual variables in these categories, we take into account the empirical regional unemployment literature. However, our set of variables is limited by data availability. Table 5 reports simple descriptive statistics of these variables, while Tables 3 and 4 show that all of them are affected by significant cross-sectional dependence and, except for $serv$, they can be considered stationary variables.

In order to account for regional disequilibrium labor market dynamics, the employment growth rate ($empgrowth_{i,t}$) is included in the set of explanatory variables. Obviously, it is expected to have a negative effect on unemployment.

The other covariates are proxies of equilibrium variables. First of all, differences in the industrial mix should impact the geographical distribution of unemployment. Provinces specializing in a declining economic sector, such as agriculture or indus-

[8] If some of these covariates were considered as endogeneous, the methodology outlined in Sect. 3 can be extended using the control function approach, as explained in Basile et al. (2014).

try, might show higher structural unemployment rates than provinces specializing in services and construction. The share of employment in agriculture ($agri_{i,t}$), in industry ($ind_{i,t}$), in services ($serv_{i,t}$), and in construction ($cons_{i,t}$) over total provincial employment are proxies of the provincial economic structure. However, we excluded ind from our model due to multicollinearity problems.

The labor force participation rate ($partrate_{i,t}$), i.e. the ratio between the total labor force and the working population (population aged between 16 and 65 years), is used as an indicator of labor supply. The expected sign of its coefficient is not unambiguous. On the one hand, factors determining low participation rates in a particular region also reflect relatively low investments in human capital and low commitment to working life, which make it more likely that people with these characteristics will become unemployed. On the other hand, a positive effect may occur if the growth of the labor force (i.e. young people) outpaces the growth of new jobs (or vacancies).

As stated above, *spatial amenities* are considered a compensating differential for the higher probability of unemployment. Variables used to proxy for producer and consumer amenities are largely conditioned by the availability of data, which are usually very scarce.[9] Thus, scholars often introduce spatial fixed effects in the model to control for time-invariant unobservable equilibrium effects (e.g. Ríos 2014). The alternative approach used in the present paper consists of including a *spatial trend* (i.e. the smooth interaction between latitude and longitude) in the model as a way to clean up the residuals. Specifically, the spatial trend can be regarded as an alternative to the use of individual regional dummies to capture unobserved spatial heterogeneity, as long as said heterogeneity is smoothly distributed over space. Regional dummies peak at 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 spatial heterogeneity is spatially autocorrelated), the smooth spatial trend is able to capture them. This is the case in our empirical application (and, we would argue, in most spatial economic analyses). Figure 3 shows the maps of the distribution of the spatial fixed effects estimated with different models (the one-way fixed effects model [FE], the two-way fixed effects model [FE/TE], the one-way fixed effects SAR model [SAR-FE], and the two-way fixed effects SAR model [SAR-FE/TE]). In all cases, the maps do indeed clearly show a smooth distribution of the fixed effects, very similar to the North-South spatial trend described above in Fig. 1.

Unobserved common business cycle factors may also influence labor market dynamics with a heterogeneous effect across regions. Some authors include time fixed effects or measures of cyclical output fluctuations (e.g. Ríos 2014) to control for these effects, although the CCEP approach (or a combination of the CCE with a spatial model) is surely a better solution (Vega and Elhorst 2016). The PS-ANOVA model with a spatio-

[9] In addition, there are many other equilibrium and disequilibrium variables affecting regional unemployment differentials. These include, for example, demographic factors (workers migration, commuting, age structure of the population and human capital variables), and institutional factors (unemployment benefits, tax wedge, employment protection legislation, collective bargaining labor relations, and so on). Valid measures for all these variables are often difficult to find at the adopted spatial unit level of the analysis. This means that there is a huge amount of spatial *unobserved heterogeneity* when modeling regional unemployment rates.

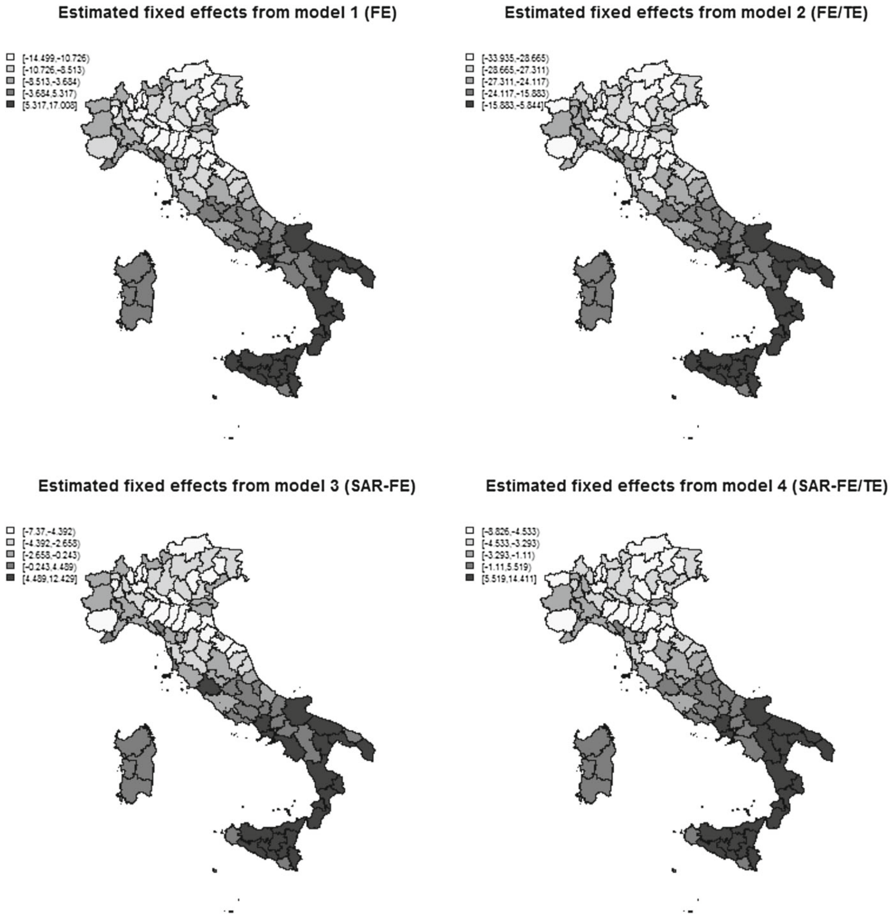


Fig. 3 Maps of spatial fixed effects estimated with models 1, 2, 3 and 4

temporal trend is used here as an alternative model to control for unobserved common factors.

Finally, we include the *spatial lag of the regional unemployment rate* ($\sum_{j \neq i} w_{ij}\, unrate_{jt}$) on the right-hand side of the model. It is important to remark again that this endogenous variable should only capture substantive spatial dependence (i.e. externalities in regional labor markets), which implies that the unemployment rate of a particular region is affected by both its own labor market characteristics and the labor market performance of other regions, rather than spatially correlated unobserved heterogeneity or common time effects.

A distance-based spatial weights matrix (W) is used to estimate spatial lag models. A general element of this matrix, w_{ij}, represents a combination of a binary spatial weight based on the critical cut-off criterion and a decreasing function of pure geographical distance, namely the inverse distance function, d_{ij}^{-1}:

$$
\nu_{ij} = \begin{cases} d_{ij}^{-1} / \sum_{j \neq i} d_{ij}^{-1} & if \ 0 < d_{ij} < d^* \\ 0 & if \ i = j \ or \ if \ d_{ij} > d^* \end{cases}
$$

where d_{ij} is the great-circle distance between the centroids of provinces i and j. The selected cut-off distance (d^*) corresponds to the minimum distance that allows all provinces to have at least one neighbor.

5.3 Econometric results

5.3.1 Model selection and diagnostics

We use the data described above to compare the performance of PS-ANOVA-SAR(AR1) specifications with that of different competing parametric and semiparametric models in terms of: i) goodness of fit, by means of a BIC criterion defined as BIC=Residual sum of squares+ $log(N \times T) \times EDF$, where EDF corresponds to the number of parameters fitted in the case of parametric models, and to the effective dimension of the model defined in "Appendix B" in the case of models with non-parametric smooth terms; and ii) residual diagnostics, focusing on the tests for cross-sectional dependence, and for serial correlation (see Table 6 for the complete list of models).

The most restricted specifications are the *parametric* one-way (FE) and two-way (FE/TE) fixed effects models (**Models 1 and 2**), estimated using the standard within-group estimator. Clearly, they cannot capture the presence of cross-sectionally correlated error terms, either strong or weak, as indicated by the results of the CD test (Table 7). Using the CCEP estimator proposed by Pesaran (2006) (**Model 5**), the

Table 6 List of models

Linear parametric panel data models	
Model 1	Fixed spatial effects model (FE)
Model 2	Fixed spatial and time effects model (FE/TE)
Model 3	SAR model with fixed spatial effects (SAR-FE)
Model 4	SAR model with fixed spatial and time effects (SAR-FE/TE)
Model 5	Model with unobserved common effects (CCEP)
Model 6	SAR model with unobserved common effects (SAR-CCEP)
Spatio-temporal penalized spline (PS) ANOVA models	
Model 7	Spatio-temporal model with linear terms (PS-ANOVA-Linear)
Model 8	PS-ANOVA-Linear-AR1
Model 9	Spatio-temporal SAR model with linear terms (PS-ANOVA-SAR-Linear)
Model 10	PS-ANOVA-SAR-Linear-AR1
Model 11	Spatio-temporal model with nonlinear terms (PS-ANOVA-Nonlinear)
Model 12	PS-ANOVA-Nonlinear-AR1
Model 13	Spatio-temporal SAR model with nonlinear terms (PS-ANOVA-SAR-Nonlinear)
Model 14	PS-ANOVA-SAR-Nonlinear-AR1

Table 7 Model selection and diagnostics

Model	CD test	rho	phi	EDF	σ^2	BIC	LM-AR1 test
Parametric models (p values in parenthesis)							
Model 1	139.57***			108.00	6.01	4225.60	4538.37***
	(0.00)						(0.00)
Model 2	2.67***			126.00	3.29	3161.00	1746.82***
	(0.01)						(0.00)
Model 3	− 0.61	0.55***		109.00	3.53	3188.10	16.92***
	(0.54)	(0.00)					(0.00)
Model 4	− 0.09	0.28***		127.00	2.94	2951.30	12.02***
	(0.92)	(0.00)					(0.00)
Model 5	− 0.67			726.00	1.03	4666.00	18.53***
	(0.50)						(0.00)
Model 6	− 0.41	0.10***		727.00	1.02	4649.45	3.97**
	(0.68)	(0.00)					(0.05)
Spatio-temporal models (p values in parenthesis)							
Model 7	− 0.31			143.15	2.43	2679.19	548.8***
	(0.76)						(0.00)
Model 8	− 1.59		0.901***	46.37	1.44	1031.89	1.09
	(0.11)		(0.00)				(0.30)
Model 9	− 0.32	− 0.021		145.25	2.42	2688.94	541.12***
	(0.75)	(0.51)					(0.00)
Model 10	− 1.64	0.084***	0.897***	46.36	1.44	1026.83	1.196
	(0.10)	(0.00)	(0.00)				(0.27)
Model 11	0.07			171.95	2.10	2578.75	458.02***
	(0.94)						(0.00)
Model 12	− 1.04		0.904***	72.19	1.36	1077.89	0.994
	(0.30)		(0.00)				(0.32)
Model 13	0.07	− 0.003		173.14	2.10	2586.66	457.1***
	(0.94)	(0.92)					(0.00)
Model 14	− 1.11	0.065**	0.903***	71.95	1.35	1071.30	1.01
	(0.27)	(0.00)	(0.00)				(0.31)

The CD test is Pesaran's cross-sectional dependence test. ρ is the spatial spillover parameter, and ϕ is the time series correlation coefficient in the noise. The EDF values include the parametric (fixed part in mixed model) and non-parametric (random part in mixed model) for each additive covariate. Therefore, they correspond to the total value of estimated degrees of freedom for each variable. σ^2 is the model variance. BIC is the value of the Bayesian Information Criterion. LM-AR1 test is a Wooldridge-type Lagrange Multiplier test for AR(1) errors

***; **, and * statistical significance at 1%, 5%, and 10%, respectively

evidence of cross-dependence disappears. These results strongly confirm the existing literature. However, with the CCEP method, we cannot disentangle strong and weak cross-dependence, that is we cannot assess the presence of spatial interaction (network) effects net of the effect of strong cross-sectional dependence.

On the other hand, with the spatial lag fixed effects models (SAR-FE and SAR-FE/TE; **Models 3 and 4**) widely used in the recent applied spatial panel data literature (Elhorst 2014), we are implicitly assuming that only weak cross-dependence (i.e. spatial dependence) exists. The CD test for the residuals of models 3 and 4 reveals that the null cannot be rejected, but the ρ parameter is quite high, suggesting that the spatial lag term has likely captured all cross-dependence (both strong and weak). Combining the SAR-FE specification and the CCEP model (**Model 6**), that is, estimating a linear spatial lag model with interactions between spatial fixed effects and the cross-sectional average of dependent and independent variables, in line with recent contributions (Bai and Li 2013; Shi and Lee 2018; Bailey et al. 2016; Vega and Elhorst 2016), we allow for both strong and weak cross-dependence. Indeed, the value of the ρ parameter (0.10) now appears much lower than before, while still remaining statistically significant. This value seems to be much more plausible than those estimated with SAR models, since we believe that most of the cross-sectional dependence in local labor market performance is due to unobserved time-related factors which influence all regions, rather than to unobserved idiosyncratic shocks which propagate to all regions with a distance decay mechanism driven by network relationships. Finally, **Model 6** also proves to be more robust against residual serial correlation, as shown by the result of the Wooldridge-type Lagrange Multiplier test for AR(1) errors (LM-AR1) reported in the last column of Table 7, while all other parametric models show a strongly significant error correlation in the residuals.[10]

The results of the CD test on the residuals confirm that the smooth spatio-temporal trend (**Models 7–14**) is able to capture the unobserved cross-sectional dependence and thus represents a valid alternative to the inclusion of cross-sectional averages in the model.[11] With respect to fixed effects models and to CCEP models, the PS-ANOVA models are less affected by the incidental (nuisance) parameter problem as a result of the effective penalizing estimation procedure described in Sect. 3. Indeed, the BIC values of PS-ANOVA models are lower than those computed for any parametric model. The PS-ANOVA model has been estimated by including either linear or nonlinear terms for the explanatory variables, and eventually adding a spatial autoregressive term (**Wy**) and a serially correlated (AR1) error term. It clearly emerges that the AR1 extension is needed to control for the strong persistence in the residuals. In absolute terms, **Model 10** (the PS-ANOVA-SAR-Linear-AR1) shows the best performance with a BIC value of 965, and non-significant cross-sectional dependence and serial correlation in the residuals. Moreover, the ρ parameter estimated with **Model 10** (0.08) is statistically significant and very close in magnitude to that estimated with **Model 6**, confirming the existence of some weak dependence net of the effect of common business cycle effects.

---

[10] Testing error persistence in the case of fixed effects models (like **Models 1–6**) is complicated by the 'artificial' serial correlation induced by time-demeaning. In fact, if the original errors are serially uncorrelated, the transformed ones are negatively serially correlated with coefficient $-1/(T-1)$. Thus, following Millo (2015), the null hypothesis for the Wooldridge-type test of serial correlation in the case of **Models 1–6** is $H_0 : \psi = -1/(T-1)$, while in the case of **Models 7–14** (which do not include fixed effects) is simply $H_0 : \psi = 0$.

[11] All the computations for smooth spatio-temporal models have been made with the R package *sptpsar* available in github (https://github.com/rominsal/sptpsar).

5.3.2 Estimation results

Table 8 reports the estimated marginal effects of the linear terms included in models 1–10, along with the associated standard errors. Obviously, for non-SAR models (i.e. models 1, 2, 5, 7, and 8), only direct effects have been reported, with indirect (spatial spillover) effects being equal to zero by construction. For all the SAR specifications direct, indirect and total marginal effects,[12] as well as their standard errors, have been computed using the algorithms discussed in Sect. 2.

We previously pointed out the evidence of a ρ parameter from SAR-FE (0.55) and SAR-FE/TE (0.28) that is rather high with respect to the values obtained with SAR-CCEP (0.10) and PS-ANOVA-SAR-AR1-Linear (0.08). It is also important to discuss here the consequences of these differences in terms of the magnitude of direct and indirect effects. In particular, in the SAR-FE model the indirect (spillover) effect of any variable is very close to the corresponding direct effect. This would imply that, if there is an idiosyncratic shock in a specific province (for example, an increase in the employment growth rate, i.e. an increase in labor demand), this shock would have the same impact on this province (direct effect) as it does in the rest of the country (spillover effect). Of course, this is implausible, since we would expect a spillover effect to be much lower than a direct effect. By including a time fixed effect in the model (SAR-FE/TE), the spillover effect turns out to be about one third of the direct effect, but it is still very high. Much more reasonable magnitudes of spillover effects (one tenth of the direct effect) emerge once we control for the common factor effects (strong cross-sectional correlation) either through the SAR-CCEP or through the PS-ANOVA-SAR-AR1-Linear model.

Focusing on **Model 10** (PS-ANOVA-SAR-AR1-Linear model), that is, the best-performing model, the results suggest that there is a clear explanation of unemployment differentials in terms of spatial equilibrium and disequilibrium factors. First, higher employment growth rates lower provincial unemployment rates, as suggested by the disequilibrium approach. Both average direct and indirect marginal effects of the variable $empgrowth$ have a negative sign and are strongly significant, indicating that an increase in the employment growth rate in one region reduces the unemployment rate not only in that region, but also in other regions, with a distance decay effect. However, as observed above, spatial spillovers (indirect effects) appear much lower than direct effects. Second, regional unemployment rates are also positively influenced by labor force participation rates. Both direct and indirect marginal effects of the

[12] It is worth noticing that the ratio between the indirect effect and the direct effect is the same for every explanatory variables in Table 8. This is the consequence of the SAR specification, where we only consider a spatial lag in the dependent variable, and not in the independent variables. As well known (Elhorst 2014), a Spatial Durbin specification, including also **WX** terms, would allow for different ratios between direct and indirect effects across the different explanatory variables. First of all, we must observe that this generalization (i.e. the inclusion of **WX** terms) does not have any effect on the estimators (either SAR-CCEP or PS-ANOVA-SAR). Indeed, as it is well known, we might define a larger matrix including both **X** and **WX** terms, and transform the Durbin specification into a SAR model. Second, in our empirical case, we have tried to estimate a Spatial Durbin version of the regional unemployment model, but the **WX** terms did not enter significantly the model. However, this is not surprising since the **WX** terms mainly work to capture unobserved heterogeneity in cross-sectional settings. In panel data settings, when unobserved heterogeneity is properly captured through other tools (fixed effects or smooth trends), the spatial lags of the exogenous variables often lose their relevance.

Table 8 Direct, indirect and total marginal effects

Model		*empgr.*	*partr.*	*agri*	*cons*	*serv*
Model 1	Direct	− 0.241***	0.591***	− 0.116***	− 0.906***	− 0.060*
		(0.019)	(0.037)	(0.043)	(0.061)	(0.032)
Model 2	Direct	− 0.168***	0.577***	0.051	− 0.092*	0.137***
		(0.015)	(0.029)	(0.036)	(0.053)	(0.029)
Model 3	Direct	− 0.187***	0.473***	− 0.080***	− 0.646***	− 0.059**
		(0.015)	(0.033)	(0.033)	(0.055)	(0.027)
	Indirect	− 0.177***	0.447***	− 0.075***	− 0.611***	− 0.056**
		(0.017)	(0.044)	(0.032)	(0.067)	(0.026)
	Total	− 0.365***	0.920***	− 0.155***	− 1.256***	− 0.115***
		(0.030)	(0.073)	(0.065)	(0.118)	(0.054)
Model 4	Direct	− 0.161***	0.520***	0.003	− 0.139***	0.104***
		(0.014)	(0.029)	(0.036)	(0.052)	(0.027)
	Indirect	− 0.055***	0.179***	0.001	− 0.048***	0.036***
		(0.008)	(0.021)	(0.013)	(0.018)	(0.010)
	Total	− 0.217***	0.699***	0.004	− 0.187***	0.140***
		(0.020)	(0.045)	(0.049)	(0.070)	(0.036)
Model 5	Direct	− 0.184***	0.570***	0.054	0.048	− 0.013
		(0.011)	(0.036)	(0.057)	(0.071)	(0.043)
Model 6	Direct	− 0.181***	0.570***	0.047	0.038	− 0.016
		(0.008)	(0.028)	(0.042)	(0.057)	(0.036)
	Indirect	− 0.018***	0.058***	0.005	0.004	− 0.002
		(0.005)	(0.017)	(0.004)	(0.006)	(0.004)
	Total	− 0.200***	0.628***	0.052	0.042	− 0.018
		(0.011)	(0.036)	(0.046)	(0.063)	(0.040)
Model 7	Direct	− 0.116***	0.213***	0.044**	− 0.024	0.060***
		(0.013)	(0.027)	(0.018)	(0.042)	(0.012)
Model 8	Direct	− 0.194***	0.636***	− 0.026	0.022	0.051**
		(0.008)	(0.030)	(0.030)	(0.045)	(0.022)
Model 9	Direct	− 0.116***	0.213***	0.044**	− 0.022	0.059***
		(0.013)	(0.028)	(0.018)	(0.040)	(0.012)
	Indirect	0.0022	− 0.004	− 0.0008	0.0004	− 0.001
		(0.004)	(0.007)	(0.0016)	(0.0016)	(0.002)
	Total	− 0.114***	0.209***	0.044**	− 0.022	0.059***
		(0.013)	(0.028)	(0.018)	(0.040)	(0.012)
Model 10	Direct	− 0.194***	0.638***	− 0.025	0.024	0.052**
		(0.008)	(0.030)	(0.032)	(0.045)	(0.022)
	Indirect	− 0.018***	0.058***	− 0.002	0.002	0.0047*
		(0.006)	(0.020)	(0.003)	(0.004)	(0.0026)
	Total	− 0.212***	0.696***	− 0.028	0.026	0.056**
		(0.010)	(0.039)	(0.035)	(0.049)	(0.024)

Standard errors (in parenthesis). For the SAR model, standard errors are computed using Monte Carlo simulations following the method discussed in Sect. 2

***; **, and * statistical significance at 1%, 5%, and 10%, respectively

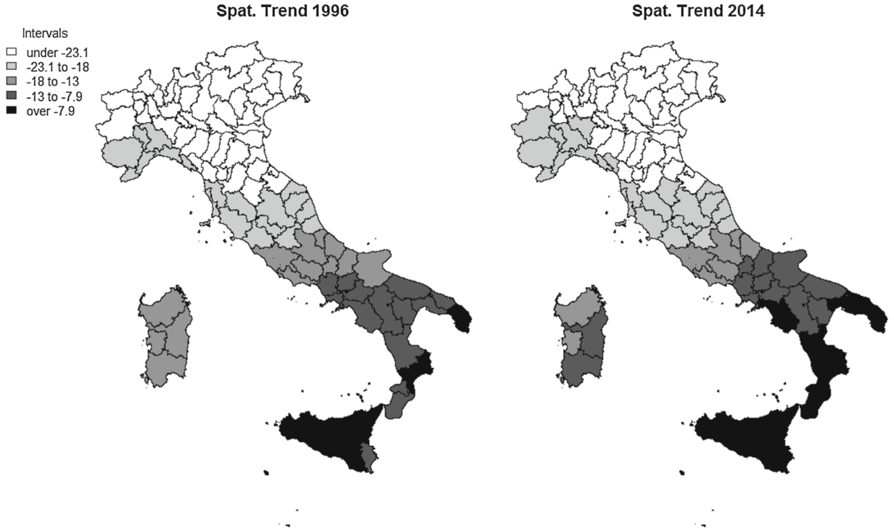


Fig. 4 Spatial trends of unrate in 1996 and 2014 for the spatio-temporal ANOVA SAR Model with linear terms and AR(1) in the noise (**Model 10**)

variable *partr* have a positive sign, but again the indirect effects are much lower than the direct effects. The positive impact of the participation rate along with the negative effect of the employment growth rate suggests, in particular, that labor market conditions in the South have worsened as a result of a growth in the labor force (i.e. more young people) that has outpaced the growth of new jobs (or vacancies). It is worth noting that the coefficients of *empgrowth* and *partrate* in **Model 10** are very close in magnitude to those in **Model 6**, confirming that the SAR-CCEP and the PS-ANOVA-SAR-AR1-Linear model behave very similarly. Third, the coefficients of the regressors related to the structure of the economy are not stable across the various model specifications. In particular, they lose statistical significance once we control for the effect of common factors.

Finally, Figs. 4 and 5 report the yearly estimated spatial trend maps, and the region-specific time trends, respectively, from **Model 10**. The maps clearly show that, even after controlling for the role of equilibrium and disequilibrium factors, the spatial distribution of expected regional unemployment rates remains persistently characterized by a strong North-South spatial trend; the estimated region-specific temporal trends also confirm the presence of common business cycles factors heterogeneously affecting all the regions.

6 Conclusions

Many large spatial panel data sets used in cross-regional and cross-country empirical analyses exhibit cross-sectional dependence which may arise from both spatial spillovers and common factors. Spatial spillovers are the results of local interactions

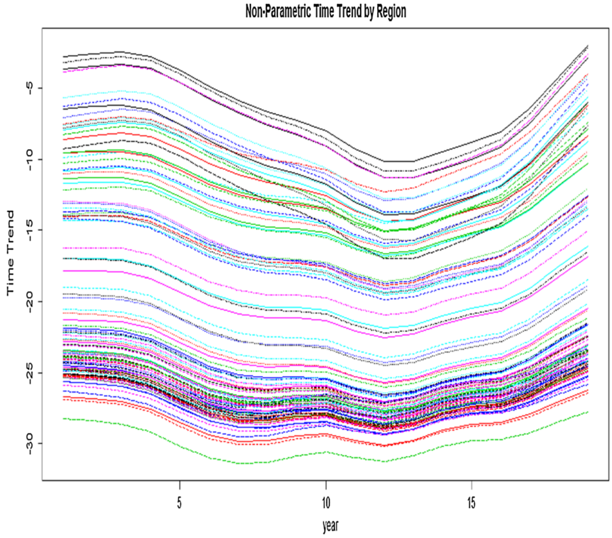


Fig. 5 Regional time trends estimated by the Spatio-Temporal ANOVA SAR with linear terms and AR(1) in the noise (**Model 10**)

and are thus classified as weak dependence effects. Common factors, on the other hand, represent latent economic-wide technological and/or demand shocks, heterogeneously affecting the dynamics in all the different regions and are thus classified as strong dependence effects.

In the present paper, we propose a wide class of models called spatio-temporal autoregressive semiparametric models (PS-SAR-ANOVA-AR1), including a non-parametric spatio-temporal trend, a spatial lag of the dependent variable, and a time series autoregressive noise. Using generated data, we have illustrated the relative small sample properties of the PS-ANOVA-SAR(AR1) model as compared to alternative DGPs such as spatial panel data models. Simulation results indicate that the PS-ANOVA-SAR(AR1) model represents a competitive alternative to parametric methods (such as the SAR-CCEP model) for estimating model parameters in the presence of time-invariant (i.e. spatial fixed effects) and time-varying unobserved heterogeneity (i.e. common factor effects) as long as both types of unobserved effects are smoothly distributed across time and space (which is quite commonly the case in regional economic analyses). More specifically, we show that the spatio-temporal trend can be interpreted as an alternative to the use of cross-sectional averages of the observations to capture the heterogeneous effect of unobserved common factors when the spatio-temporal heterogeneity is smoothly distributed.

The models proposed do not impose a predetermined structure to capture weak and strong dependence; the ANOVA decomposition of the spatio-temporal trend into a spatial trend, a time trend, and second- and third-order interactions works effectively to control for both unobserved spatial heterogeneity and unobserved common factors. Thus, the inclusion of the ANOVA decomposition of the spatio-temporal trend helps us to interpret the evidence of significant spatial spillovers as weak cross-dependence net of common effects (strong dependence). With respect to the fully parametric approaches (such as the CCEP and the SAR-CCEP models), our framework does not involve the estimation of a large number of incidental parameters. A further advantage of our approach is that it allows for non-linear relationships between the covariates and the response.

Although PS-ANOVA-SAR(AR1) models can be applied to any type of large spatial panels, we have focused on their performance in the analysis of regional economic data. In particular, we have implemented this new framework using real data on unemployment rates in Italy. The econometric results show that the PS-ANOVA-SAR-AR1 performs better than several competing parametric and nonparametric models both in terms of model fitting and diagnostics of the residuals. In particular, the spatio-temporal trend effectively captures the strong cross-sectional dependence (due to common factors), while the parameter associated with the spatial lag term reveals the existence of significant spatial spillovers net of the effect of the observed and unobserved common factors.

As a concluding remark, it is worth noting that regional unemployment rates, like many other regional and national economic variables, are typically characterized by strong persistence over time. Thus, in order to control for serial correlation, we have extended our model by including an auto-regressive (AR1) error term. In line with Chudik and Pesaran (2015), which have developed a CCEP estimator for dynamic panel data models, our future research includes the extension of the PS-ANOVA-SAR model to a dynamic specification. We believe that this dynamic setting would not imply an increase in the number of incidental parameters, and it could be a very promising alternative when dealing with spatio-temporal panels not very large in temporal dimension.

Funding Funding was provided by Ministerio de Economía, Industria y Competitividad, Gobierno de España (Grant Nos. MTM2014-52184 and ECO2015-65826-P) and Grant 2019-GRIN-26913 provided by the University of Castilla- La Mancha (UCLM) and the European Fund for Regional Development (EFRD) to the Research Group “Applied Economics and Quantitative Methods”.

Appendix A: Penalized splines as mixed models

Given the model:

$$y_i = f(x_i) + \varepsilon_i \quad \varepsilon \sim N(0, \sigma^2 \mathbf{I}),$$

using the penalized regression approach we have (in matrix form):

$$\mathbf{y} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}),$$

where $\mathbf{B}$ is a matrix of B-spline bases, and $\boldsymbol{\theta}$ a vector of regression parameters to be estimated via penalized sum of squares:

$$(\mathbf{y} - \mathbf{B}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}.$$

The reformulation of a P-spline into a mixed model can be viewed as a reparameterization of the original non-parametric model; B-spline bases are transformed into a new model basis, i.e. $\mathbf{B} \to [\mathbf{X} : \mathbf{Z}]$, and coefficients $\boldsymbol{\theta} \to (\boldsymbol{\beta}, \boldsymbol{\alpha})'$. Hence, this representation decomposes the fitted values into the sum of a polynomial (unpenalized) part ($\mathbf{X}\boldsymbol{\beta}$) and a nonlinear (penalized) ($\mathbf{Z}\boldsymbol{\alpha}$) smooth term. To carry out this transformation, we need to find an (orthogonal) transformation matrix $\mathbf{T}$, so that $\mathbf{BT} = [\mathbf{X} : \mathbf{Z}]$ and $\mathbf{T}'\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})'$. There are several possibilities for this matrix; we choose one based on the singular value decomposition of the penalty matrix $\mathbf{P} = \lambda \mathbf{D}'\mathbf{D}$, that is:

$$\mathbf{D}'\mathbf{D} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}',$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix that contains the eigenvalues of $\mathbf{D}'\mathbf{D}$, with 2 zero eigenvalues, and $\mathbf{U}$ is the corresponding matrix of eigenvectors that can be decomposed into two parts: $\mathbf{U}_n$ of dimension $c \times 2$ containing the null-part eigenvectors and $\mathbf{U}_s$ of dimension $c \times (c-2)$ (where c is the rank of the basis and 2 the order of the penalty) with non-null-part eigenvectors. Note that we can write $\boldsymbol{\Sigma}$ as $\boldsymbol{\Sigma} = blockdiag\left(\mathbf{0}_2, \tilde{\boldsymbol{\Sigma}}\right)$, where $\tilde{\boldsymbol{\Sigma}}$ is a diagonal matrix that contains the non-zero eigenvalues of $\mathbf{D}'\mathbf{D}$ and $\mathbf{0}_2$ is a 2×2 matrix of zeroes. Therefore, we define the transformation matrix $\mathbf{T}$ as:

$$\mathbf{T} = [\mathbf{U}_n : \mathbf{U}_s \tilde{\boldsymbol{\Sigma}}^{-1/2}],$$

where the fixed and random effect matrices are $\mathbf{X} = \mathbf{B}\mathbf{U}_n$, and $\mathbf{Z} = \mathbf{B}\mathbf{U}_s \tilde{\boldsymbol{\Sigma}}^{-1/2}$, respectively. Also, given this transformation matrix, the new coefficients are $\boldsymbol{\beta} = \mathbf{U}_n'\boldsymbol{\theta}$ and $\boldsymbol{\alpha} = \mathbf{U}_s' \tilde{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{\theta}$. The fixed effect matrix $\mathbf{X}$ may be replaced by any sub-matrix such that $[\mathbf{X} : \mathbf{Z}]$ has full rank and $\mathbf{X}'\mathbf{Z} = \mathbf{0}$ (that is, $\mathbf{X}$ and $\mathbf{Z}$ are orthogonal). So, for example, if we assume a second-order penalty ($d = 2$), the fixed effect matrix can be taken as $\mathbf{X} = [\mathbf{1} : \mathbf{x}]$, where $\mathbf{1}$ is a vector of ones and $\mathbf{x}$ is the explanatory variable. Also, the penalty term $\boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}$ becomes $\boldsymbol{\alpha}'\mathbf{F}\boldsymbol{\alpha}$, where $\mathbf{F} = \lambda\mathbf{I}$. This follows since $\mathbf{T}$ is orthogonal and $(\boldsymbol{\beta}, \boldsymbol{\alpha})' = \mathbf{T}'\boldsymbol{\theta}$. Hence, given the new basis and the new penalty, the penalized sum of squares,

$$(\mathbf{y} - \mathbf{B}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{B}\boldsymbol{\theta}) + \boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta},$$

becomes:

$$(\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\boldsymbol{\alpha})' (\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\boldsymbol{\alpha}) + \lambda\boldsymbol{\alpha}' I_{c-2}\boldsymbol{\alpha},$$

This corresponds to the joint log-likelihood of a linear mixed model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}), \quad \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, \mathbf{G}),$$

with $\mathbf{G} = \sigma_v^2\mathbf{I}_{c-2}$ and $\lambda = \sigma^2/\sigma_v^2$. Therefore, the smoothing parameters is estimated via the estimation of the variance components in the mixed model.

Appendix B: Mixed model representation of the semiparametric spatio-temporal autoregressive model and parameter estimation

For the sake of simplicity, we assume here that there are no covariates. The inclusion of covariates with a linear or non-linear functional relationship with the response is immediate by augmenting the matrices for fixed and random effects accordingly, as well as the corresponding covariance matrices. We therefore focus on the following model:

$$\begin{aligned}\mathbf{y} = & f_1(\mathbf{s}_1) + f_2(\mathbf{s}_2) + f_t(\boldsymbol{\tau}) + f_{1,2}(\mathbf{s}_1, \mathbf{s}_2) + f_{1,t}(\mathbf{s}_1, \boldsymbol{\tau}) \\ & + f_{2,t}(\mathbf{s}_2, \boldsymbol{\tau}) + f_{1,2,t}(\mathbf{s}_1, \mathbf{s}_2, \boldsymbol{\tau}) + \rho(\mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + \boldsymbol{\epsilon}\end{aligned}$$

where the errors are assumed to follow a temporal AR(1) process, see (9). In matrix form:

$$(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \frac{\sigma^2}{1-\phi^2}(\mathbf{I}_N \otimes \boldsymbol{\Omega})\right) \qquad \mathbf{A}_N = \mathbf{I}_N - \rho\mathbf{W}_N$$

The regression matrix of the model above will be the concatenation of B-spline bases for each of the smooth terms in the model:

$$\mathbf{B} = [\mathbf{1}|\mathbf{B}_{s_1}|\mathbf{B}_{s_2}|\mathbf{B}_{s_2}|\mathbf{B}_{s_1}\Box\mathbf{B}_{s_2}|\mathbf{B}_{s_1} \otimes \mathbf{B}_\tau|\mathbf{B}_{s_2} \otimes \mathbf{B}_\tau|(\mathbf{B}_{s_1}\Box\mathbf{B}_{s_2}) \otimes \mathbf{B}_\tau,]$$

where $\mathbf{B}_{s_1}$, $\mathbf{B}_{s_2}$ and $\mathbf{B}_\tau$ correspond to the marginal B-spline basis for the spatial coordinates $(\mathbf{s}_1, \mathbf{s}_2)$ and time $(\boldsymbol{\tau})$, and $\Box$ represents the row-wise tensor product defined as:

$$\mathbf{B}_i \Box \mathbf{B}_j = (\mathbf{B}_i \otimes \mathbf{1}'_{c_i}) * (\mathbf{1}'_{c_j} \otimes \mathbf{B}_j),$$

and $\mathbf{1}$ is a column vector of ones, c_i is the rank of $\mathbf{B}_i$, and $\otimes$ and $*$ are the Kronecker and element-wise matrix products, respectively.

The penalty matrix is now block-diagonal with blocks corresponding to the different terms in the model: $\lambda_i\mathbf{D}'_i\mathbf{D}_i$ for main effects, $\lambda_i\mathbf{D}'_i\mathbf{D}_i \otimes \mathbf{I}_{c_k} + \lambda_k\mathbf{I}_{c_i} \otimes \mathbf{D}'_k\mathbf{D}_k$ for the second-order interactions, and $\lambda_i\mathbf{D}'_i\mathbf{D}_i \otimes \mathbf{I}_{c_k} \otimes \mathbf{I}_{c_l} + \lambda_k\mathbf{I}_{c_i} \otimes \mathbf{D}'_k\mathbf{D}_k \otimes \mathbf{I}_{c_j} + \lambda_l \otimes \mathbf{I}_{c_i} \otimes \mathbf{I}_{c_k} \otimes \mathbf{D}'_l\mathbf{D}_l$ for the three-way interaction.

In this case, several constraints need to be imposed, since the space spanned by any product $\mathbf{B}_i \otimes \mathbf{B}_j$, contains the space spanned by the marginal bases $\mathbf{B}_i$ and $\mathbf{B}_j$. The mixed model reparameterization of this model will automatically provide the necessary constraints. To find that parameterization, a new transformation matrix is needed (again based on the singular value decomposition of the penalty $\mathbf{P}$) (see Lee 2010, for details). Then, the model is written as:

$$\begin{aligned}&(\mathbf{A}_N \otimes \mathbf{I}_T)\,\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \\ &\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{G}) \qquad \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \frac{\sigma^2}{1-\phi^2}(\mathbf{I}_N \otimes \boldsymbol{\Omega})\right)\end{aligned} \tag{12}$$

with

$$\begin{aligned}
\mathbf{X} &= \left[(\mathbf{X}_{s_1} \square \mathbf{X}_{s_2}) \otimes \mathbf{X}_\tau\right] \\
\mathbf{Z} &= \left[(\mathbf{Z}_{s_1} \square \mathbf{X}_{s_2}) \otimes \mathbf{X}_\tau | (\mathbf{X}_{s_1} \square \mathbf{Z}_{s_2}) \otimes \mathbf{X}_\tau | (\mathbf{X}_{s_1} \square \mathbf{X}_{s_2}) \otimes \mathbf{Z}_\tau | (\mathbf{Z}_{s_1} \square \mathbf{Z}_{s_2}) \otimes \mathbf{X}_\tau | \right. \\
&\quad \left. (\mathbf{Z}_{s_1} \square \mathbf{X}_{s_2}) \otimes \mathbf{Z}_\tau | (\mathbf{X}_{s_1} \square \mathbf{Z}_{s_2}) \otimes \mathbf{Z}_\tau | (\mathbf{Z}_{s_1} \square \mathbf{Z}_{s_2}) \otimes \mathbf{Z}_\tau \right]
\end{aligned}$$

where $\mathbf{X}_k$, $\mathbf{Z}_k$ $(k = s_1, s_2, \tau)$ are the mixed model matrices obtained for the reparameterization of the marginal basis described in "Appendix A". The covariance matrix of random effects, $\mathbf{G}$, is such that:

$$\begin{aligned}
\mathbf{G}^{-1} = \text{blockdiag}\Bigg(&\mathbf{0}, \frac{1}{\sigma^2_{\nu_1}}\boldsymbol{\Lambda}_1, \frac{1}{\sigma^2_{\nu_2}}\boldsymbol{\Lambda}_2, \frac{1}{\sigma^2_{\nu_3}}\boldsymbol{\Lambda}_3, \frac{1}{\sigma^2_{\nu_4}}\boldsymbol{\Lambda}_4 + \frac{1}{\sigma^2_{\nu_5}}\boldsymbol{\Lambda}_5, \frac{1}{\sigma^2_{\nu_6}}\boldsymbol{\Lambda}_6 + \frac{1}{\sigma^2_{\nu_7}}\boldsymbol{\Lambda}_7, \\
&\frac{1}{\sigma^2_{\nu_8}}\boldsymbol{\Lambda}_8 = \frac{1}{\sigma^2_{\nu_9}}\boldsymbol{\Lambda}_9, \frac{1}{\sigma^2_{\nu_{10}}}\boldsymbol{\Lambda}_{10} + \frac{1}{\sigma^2_{\nu_{11}}}\boldsymbol{\Lambda}_{11} + \frac{1}{\sigma^2_{\nu_{12}}}\boldsymbol{\Lambda}_{12}\Bigg)
\end{aligned} \tag{13}$$

where

$$\begin{aligned}
&\boldsymbol{\Lambda}_1 = \widetilde{\boldsymbol{\Sigma}}_{s_1}, \quad \boldsymbol{\Lambda}_2 = \widetilde{\boldsymbol{\Sigma}}_{s_2}, \quad \boldsymbol{\Lambda}_3 = \widetilde{\boldsymbol{\Sigma}}_\tau \\
&\boldsymbol{\Lambda}_4 = \widetilde{\boldsymbol{\Sigma}}_{s_1} \otimes \mathbf{I}_{c_{s_2}-2}, \quad \boldsymbol{\Lambda}_5 = \mathbf{I}_{c_{s_1}-2} \otimes \widetilde{\boldsymbol{\Sigma}}_{s_2}, \quad \boldsymbol{\Lambda}_6 = \widetilde{\boldsymbol{\Sigma}}_{s_1} \otimes \mathbf{I}_2 \\
&\boldsymbol{\Lambda}_7 = \mathbf{I}_{c_{s_1}-q_{s_1}} \otimes \mathbf{I}_2, \quad \boldsymbol{\Lambda}_8 = \widetilde{\boldsymbol{\Sigma}}_{s_2} \otimes \mathbf{I}_{c_t-2} \quad \boldsymbol{\Lambda}_9 = \mathbf{I}_{c_{s_2}-2} \otimes \widetilde{\boldsymbol{\Sigma}}_\tau \\
&\boldsymbol{\Lambda}_{10} = \widetilde{\boldsymbol{\Sigma}}_{s_1} \otimes \mathbf{I}_{c_{s_2}-2} \otimes \mathbf{I}_{c_\tau-2}, \quad \boldsymbol{\Lambda}_{11} = \mathbf{I}_{c_{s_1}-2} \otimes \widetilde{\boldsymbol{\Sigma}}_{s_2} \otimes \mathbf{I}_{c_\tau-2}, \\
&\boldsymbol{\Lambda}_{12} = \mathbf{I}_{c_{s_1}-2} \otimes \mathbf{I}_{c_{s_2}-2} \otimes \widetilde{\boldsymbol{\Sigma}}_\tau
\end{aligned} \tag{14}$$

and $\widetilde{\boldsymbol{\Sigma}}$ matrices correspond to the non-zero eigenvectors of the singular value decomposition of penalty matrices. It is important to be able to decompose the precision matrix of the random effects as a linear combination over the variance parameters, since this is a necessary condition to apply the SAP algorithm.

B.1: Estimation of the PS-ANOVA-SAR(AR1) model via the SAP algorithm

Fixed and random effects in model (12) are estimated (conditional on the correlation parameters and variance components) using the standard mixed model theory (see Searle et al. 1992):

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y} \tag{15}$$

$$\widehat{\boldsymbol{\alpha}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}((\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}), \tag{16}$$

where $\mathbf{V} = \frac{\sigma^2}{1-\phi^2}(\mathbf{I}_N \otimes \boldsymbol{\Omega}) + \mathbf{Z}\mathbf{G}\mathbf{Z}'$.

Variance components (and, therefore, smoothing parameters), and correlation parameters may be estimated by maximizing the residual log-likelihood (REML) of Patterson and Thompson (1971) (slightly modified by the Kronecker matrix product, $\mathbf{A}_N \otimes \mathbf{I}_T$):

$$\begin{aligned}\ell(\sigma^2_{\nu_i}, \sigma^2, \rho, \phi) = &-\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \\ &-\frac{1}{2}\left[(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y}\right]'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\left[(\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y}\right] \\ &+\log|\mathbf{A}_N \otimes \mathbf{I}_T| \end{aligned} \tag{17}$$

where the matrices $\mathbf{V}$, $\mathbf{X}$ and $\mathbf{Z}$ are obtained as described above (if linear and non-linear covariates have been added, $\mathbf{X}$ and $\mathbf{Z}$ matrices are augmented in a suitable additive way).

Maximization of this REML function is a very complex numerical problem, specially when the number of variance components/correlation parameters is large. Rodriguez-Alvarez et al. (2015) recently developed an algorithm named SAP (Separation of Anisotropic Penalties), which is based on the fact that the inverse variance-covariance matrix of the random effects, $\mathbf{G}^{-1}$, is a linear combination of precision matrices. This is the case for the PS-ANOVA-SAR(AR1) model, as we showed in (13). This expression allows us to get closed estimates for all the variance component parameters $\sigma^2_{\nu_i}$ and σ^2 very efficiently. We have adapted this algorithm to also include the estimation of ρ and ϕ parameters. The steps for applying the SAP algorithm to optimize (17) can be summarized as follows:

1. Initialization. Set
   - Set $k = 0$
   - $\hat{\boldsymbol{\beta}}^{(k)} = \mathbf{0}$; $\hat{\boldsymbol{\alpha}}^{(k)} = \mathbf{0}$
   - $\hat{\sigma}^{2,(k)}_{\nu_i} = 1 \quad i = 1, 2, \ldots, 12$
   - $\hat{\sigma}^{2,(k)} = \mathrm{var}(\mathbf{y})$
   - $\hat{\rho}^{(k)} = 0$
2. Compute $\hat{\mathbf{G}}^{(k)}$, $\hat{\mathbf{V}}^{(k)}$, $\hat{\mathbf{P}}^{(k)}$, $\hat{\mathbf{A}}^{(k)}_N$ matrices using next expressions:

$$\begin{aligned}\hat{\mathbf{G}}^{-1,(k)} &= \sum_{i=1}^{12} \frac{1}{\hat{\sigma}^{2,(k)}_{\nu_i}} \boldsymbol{\Lambda}^{(k)}_i \\ \hat{\mathbf{V}}^{(k)} &= \hat{\sigma}^{2,(k)}\mathbf{I}_{NT} + \mathbf{Z}\hat{\mathbf{G}}^{(k)}\mathbf{Z} \\ \hat{\mathbf{P}}^{(k)} &= \hat{\mathbf{V}}^{-1,(k)} - \hat{\mathbf{V}}^{-1,(k)}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1,(k)}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1,(k)} \\ \hat{\mathbf{A}}^{(k)}_N &= \mathbf{I}_N - \hat{\rho}^{(k)}\mathbf{W}_N\end{aligned}$$

3. Compute the estimates:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{(k)} &= (\mathbf{X}'\hat{\mathbf{V}}^{-1,(k)}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{V}}^{-1,(k)}\hat{\mathbf{A}}_N^{(k)}\mathbf{y}) \\
\hat{\boldsymbol{\alpha}}^{(k)} &= \hat{\mathbf{G}}^{(k)}\mathbf{Z}'\hat{\mathbf{V}}^{-1,(k)}(\hat{\mathbf{A}}_N^{(k)}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(k)}) \\
ed_i^{(k)} &= \text{trace}(\mathbf{Z}'\hat{\mathbf{P}}^{(k)}\mathbf{Z}\hat{\mathbf{G}}^{(k)}\frac{1}{\hat{\sigma}_{\nu_i}^{2,(k)}}\boldsymbol{\Lambda}_i\hat{\mathbf{G}}^{(k)}) \quad i = 1, 2, \ldots, 12
\end{aligned}$$

where $\boldsymbol{\Lambda}_i \quad i = 1, \ldots, 12$ is defined in (14).

4. Estimate the variance components:

$$\hat{\sigma}_{\nu_i}^{2,(k+1)} = \frac{\hat{\boldsymbol{\alpha}}^{(k)'}\boldsymbol{\Lambda}_i\hat{\boldsymbol{\alpha}}^{(k)}}{ed_i^{(k)}} \quad i = 1, \ldots, 12$$

Estimate the variance of the noise as:

$$\hat{\sigma}^{2,(k+1)} = \frac{(\hat{\mathbf{A}}_N^{(k)}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(k)} - \mathbf{Z}\hat{\boldsymbol{\alpha}}^{(k)})'(\hat{\mathbf{A}}_N^{(k)}\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^{(k)} - \mathbf{Z}\hat{\boldsymbol{\alpha}}^{(k)})}{N - \sum_i ed_i^{(k)} - \text{rank}(\mathbf{X}) - 2}$$

5. Estimate the spatial parameter $\hat{\rho}^{(k+1)}$ and serial correlation parameter $\hat{\phi}^{(k+1)}$ solving numerically the non-linear equations obtained by equating to zero the score of REML function with respect to ρ and ϕ respectively (this additional step is the only difference with respect to the usual SAP algorithm):

$$\begin{aligned}
\frac{\partial \ell(.)}{\partial \rho} &= -\frac{1}{2}\left[2\hat{\mathbf{P}}^{(k)}\left((\mathbf{A}_N \otimes \mathbf{I}_T)\mathbf{y}\right)\right]'\left(\frac{\partial(\mathbf{A}_N \otimes \mathbf{I}_T)}{\partial \rho}\mathbf{y}\right) \\
&\quad + \text{trace}\left((\mathbf{A}_N \otimes \mathbf{I}_T)^{-1}\frac{\partial(\mathbf{A}_N \otimes \mathbf{I}_T)}{\partial \rho}\right) \\
&= \mathbf{y}'(\mathbf{A}_N \otimes \mathbf{I}_T')\hat{\mathbf{P}}^{(k)}(\mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} - T\,\text{trace}(\mathbf{A}_N^{-1}\mathbf{W}_N) = 0 \\
\frac{\partial l(.)}{\partial \phi} &= -\frac{1}{2}\left\{\text{trace}\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \phi}\right) - \left[(\mathbf{A}_N \otimes \mathbf{I}_T)\,\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right]'\mathbf{V}^{-1}\right. \\
&\quad \left.\times \frac{\partial \mathbf{V}}{\partial \phi}\mathbf{V}^{-1}\left[(\mathbf{A}_N \otimes \mathbf{I}_T)\,\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right]\right\} = 0
\end{aligned}$$

where:

$$\frac{\partial \mathbf{V}}{\partial \phi} = \frac{\partial\left\{\mathbf{Z}\mathbf{G}\mathbf{Z}' + \frac{\sigma^2}{1-\phi^2}(\mathbf{I}_N \otimes \boldsymbol{\Omega})\right\}}{\partial \phi} = \left(\mathbf{I}_N \otimes \frac{\partial\left[(\frac{\sigma^2}{1-\phi^2})\boldsymbol{\Omega}\right]}{\partial \phi}\right)$$

and

6. Set $k = k + 1$ and go to step (2) until convergence.

Once the convergence is obtained, the effective degrees of freedom of the model can be estimated as:

$$\text{edf} = \sum_i ed_i^{(k)} + \text{rank}(\mathbf{X}) + 2$$

This quantity is increased by two units with respect to spatio-temporal smooth models because of the need to estimate ρ and ϕ parameters.

To obtain the covariance matrix of the estimates, we need the hessian matrix of REML function with respect to ρ and ϕ parameters given by the expressions:

$$\begin{aligned}
\frac{\partial^2 l(.)}{\partial \rho^2} &= -\mathbf{y}'\left(\mathbf{W}_N' \otimes \mathbf{I}_T\right)\mathbf{P}\left(\mathbf{W}_N \otimes \mathbf{I}_T\right)\mathbf{y} - T\,\text{trace}\left((\mathbf{A}_N^{-1}\mathbf{W}_N)^2\right) \\
\frac{\partial^2 l(.)}{\partial \phi^2} &= -\frac{1}{2}\left\{\frac{\partial \text{trace}\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \phi}\right)}{\partial \phi} - \left[(\mathbf{A}_N \otimes \mathbf{I}_T)\,\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right]'\right. \\
&\quad \left. \times \frac{\partial\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{V}^{-1}\right)}{\partial \phi}\left[(\mathbf{A}_N \otimes \mathbf{I}_T)\,\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right]\right\} \\
\frac{\partial^2 l(.)}{\partial \phi \partial \rho} &= \mathbf{y}'\left(\mathbf{W}_N' \otimes \mathbf{I}_T\right)\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{V}^{-1}\left[(\mathbf{A}_N \otimes \mathbf{I}_T)\,\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right]
\end{aligned}$$

where:

$$\begin{aligned}
&\frac{\partial \text{trace}\left(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \phi}\right)}{\partial \phi} = \text{trace}\left(\frac{\partial(\mathbf{P}\frac{\partial \mathbf{V}}{\partial \phi})}{\partial \phi}\right) = \text{trace}\left(\frac{\partial \mathbf{P}}{\partial \phi}\frac{\partial \mathbf{V}}{\partial \phi} + \mathbf{P}\frac{\partial^2 \mathbf{V}}{\partial \phi^2}\right) \\
&\frac{\partial \mathbf{V}}{\partial \phi} = \left(\mathbf{I}_N \otimes \frac{\partial\left[(\frac{\sigma_\epsilon^2}{1-\phi^2})\boldsymbol{\Omega}\right]}{\partial \phi}\right) \qquad \frac{\partial^2 \mathbf{V}}{\partial \phi^2} = \left(\mathbf{I}_N \otimes \frac{\partial^2\left[(\frac{\sigma_\epsilon^2}{1-\phi^2})\boldsymbol{\Omega}\right]}{\partial \phi^2}\right) \\
&\frac{\partial \mathbf{P}}{\partial \phi} = -\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{V}^{-1} - \left(-\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\right. \\
&\quad + \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \\
&\quad \left. -\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \phi}\mathbf{V}^{-1}\right)
\end{aligned}$$

These expressions can be evaluated at maximum of REML function to obtain the negative of the hessian matrix. The inverse of this matrix provides the asymptotic covariance matrix in the usual way.

Eventually the covariance matrix of ρ and ϕ, jointly with the covariance matrix of the regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ given by $Cov(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{C}^{-1}$ (see Sect. 3), can be used to obtain the simulated distributions of total, direct and indirect effects as explained in Sect. 2. As usual, REML estimates are asymptotically unbiased and gaussian distributed.

References

Aragon Y, Haughton D, Haughton J, Leconte E, Malin E, Ruiz-Gazen A, Thomas-Agnan C (2003) Explaining the pattern of regional unemployment: the case of the Midi-Pyrénées region. Pap Reg Sci 82(2):155–174

Bailey N, Holly S, Pesaran MH (2016) A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. J Appl Econom 31(1):249–280

Bai J, Li K (2013) Spatial panel data models with common shocks. MPRA paper 52786, University of Munich, Germany

Basile R, Girardi A, Mantuano M (2012) Migration and regional unemployment in Italy. Open Urb Stud J 5:1–13

Basile R, Durbán M, Mínguez R, Montero JM, Mur J (2014) Modeling regional economic dynamics: spatial dependence, spatial heterogeneity and nonlinearities. J Econ Dyn Control 48:229–245

Blanchard OJ, Katz LF, Hall RE, Eichengreen B (1992) Regional evolutions. Brook Pap Econ Act 1992(1):1–75

Burridge P, Gordon IR (1981) Unemployment in the British metropolitan labour areas. Oxf Econ Pap 33(2):274–97

Chudik A, Pesaran MH (2015) Common correlated effects estimation of heterogeneous dynamic panel data models with weakly exogenous regressors. J Econom 188(2):393–420

Chudik A, Pesaran MH, Tosetti E (2011) Weak and strong cross-section dependence and estimation of large panels. Econom J 14(1):C45–C90

Claeskens G, Krivobokova T, Opsomer J (2007) Asymptotic properties of penalized spline estimators. Biometrika 96:529–544

Cracolici MF, Cuffaro M, Nijkamp P (2007) Geographical distribution of unemployment: an analysis of provincial differences in italy. Growth Change 38(4):649–670

Currie ID, Durbán M (2002) Flexible smoothing with P-splines: a unified approach. Stat Model 2:333–349

Currie I, Durbán M, Eilers P (2006) Generalized linear array models with applications to multidimensional somoothing. J R Stat Soc B 68:1–22

De Boor C (1977) Package for calculating with B-splines. J Numer Anal 14:441–472

Decressin J, Fatas A (1995) Regional labor market dynamics in Europe. Eur Econ Rev 39(9):1627–1655

Eilers P, Marx B (1996) Flexible smoothing with B-splines and penalties. Stat Sci 11:89–121

Eilers P, Currie I, Durbán M (2006) Fast and compact smoothing on large multidimensional grids. Comput Stat Data Anal 50(1):61–76

Eilers PH, Marx BD, Durbán M (2015) Twenty years of p-splines. SORT 39(2):149–186

Elhorst JP (1995) Convergence and divergence among European Union. Pion, London, pp 190–200

Elhorst J (2014) Spatial econometrics. From cross-sectional data to spatial panels. SpringerBriefs in regional science. Springer, Berlin

Ertur C, Musolesi A (2016) Weak and strong cross-sectional dependence: a panel data analysis of international technology diffusion. J Appl Econom 32(3):477–503

Han X, Lee L-F (2016) Bayesian analysis of spatial panel autoregressive models with time-varying endogenous spatial weight matrices, common factors, and random coefficients. J Bus Econ Stat 34:642–660

Hoshino T (2018) Semiparametric spatial autoregressive models with endogenous regressors: with an application to crime data. J Bus Econ Stat 36:160–172

Kapetanios G, Pesaran MH, Yamagata T (2011) Panels with non-stationary multifactor error structures. J Econom 160(2):326–348

Lee D (2010) Smoothing mixed models for spatial and spatio-temporal data. Ph.D. thesis, University Carlos-III

Lee D, Durbán M (2011) P-spline ANOVA type interaction models for spatio-temporal smoothing. Stat Model 11:49–69

Lee L-F, Yu J (2010) Estimation of spatial autoregressive panel data models with fixed effects. J Econom 154(2):165–185

Lee DJ, Durban M, Eilers P (2013) Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. Comput Stat Data Anal 61:22–37
LeSage J, Pace K (2009) Introduction to spatial econometrics. CRC Press, Boca Raton
Lin X, Zhang D (1999) Inference in generalized additive mixed models by using smoothing splines. J R Stat Soc B 61:381–400
Lottmann F (2012) Spatial dependencies in German matching functions. Reg Sci Urb Econ 42(1):27–41
Marston ST (1985) Two views of the geographic distribution of unemployment. Q J Econ 100(1):57–79
Millo G (2015) Testing for serial correlation in spatial panels. Mimeo
Molho I (1995) Spatial autocorrelation in british unemployment. J Reg Sci 35(4):641–658
Montero J, Mínguez R, Durbán M (2012) SAR models with nonparametric spatial trends. A P-spline approach. Estadística Española 54(177):89–111
Overman HG, Puga D (2002) Unemployment clusters across Europe's regions and countries. Econ policy 17(34):115–148
Partridge MD, Rickman DS (1997) The dispersion of US state unemployment rates: the role of market and non-market equilibrium factors. Reg Stud 31(6):593–606
Patacchini E, Zenou Y (2007) Spatial dependence in local unemployment rates. J Econ Geogr 7(2):169–191
Patterson H, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554
Perperoglou A, Eilers PHC (2009) Penalized regression and individual deviance effects. Comput Stat 25:341–361
Pesaran MH (2004) General diagnostic tests for cross section dependence in panels. Technical report, CESifo working paper series
Pesaran MH (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74(4):967–1012
Pesaran MH (2007) A simple panel unit root test in the presence of cross-section dependence. J Appl Econom 22(2):265–312
Pesaran MH (2015) Testing weak cross-sectional dependence in large panels. Econom Rev 34(6–10):1089–1117
Pesaran MH, Tosetti E (2011) Large panels with common factors and spatial correlation. J Econom 161(2):182–202
Ríos V (2014) What drives regional unemployment convergence? In: ERSA conference papers Ersa14p924, European Regional Science Association
Rodriguez-Alvarez MX, Kneib T, Durban M, Lee D, Eilers P (2015) Fast smoothing parameter separation in multidimensional generalized P-splines: the SAP algorithm. Stat Comput 25(5):941–957
Rodrìguez-Álvarez MX, Boer MP, Eeuwijk FAV, Eilers PH (2018) Correcting for spatial heterogeneity in plant breeding experiments with P-splines. Spat Stat 23:52–71
Searle S, Casella G, McCulloch C (1992) Variance components. Wiley, New York
Shi W, Lee L-F (2018) A spatial panel data model with time varying endogenous weights matrices and common factors. Reg Sci Urb Econ 72:6–34
Su L, Jin S (2012) Sieve estimation of panel data models with cross section dependence. J Econom 169(1):34–47
Taylor J, Bradley S (1997) Unemployment in Europe: a comparative analysis of regional disparities in Germany, Italy and the UK. Kyklos 50(2):221–245
Thirlwall AP (1966) Regional unemployment as a cyclical phenomenon. Scott J Polit Econ 13(2):205–219
Vega SH, Elhorst JP (2016) A regional unemployment model simultaneously accounting for serial dynamics, spatial dependence and common factors. Reg Sci Urb Econ 60:85–95
Wood S (2006) On confidence intervals for generalized additive models based on penalized regression splines. Aust N Z J Stat 48:445–464
Zeilstra AS, Elhorst JP (2014) Integrated analysis of regional and national unemployment differentials in the European Union. Reg Stud 48(10):1739–1755