

Article

Pspatreg: R Package for Semiparametric Spatial Autoregressive Models

Román Mínguez ^{1,*} , Roberto Basile ²  and María Durbán ³ ¹ University of Castilla-La Mancha, Av. de los Alfares, 44, 16002 Cuenca, Spain² University of Rome Sapienza, 00185 Rome, Italy; roberto.basile@uniroma1.it³ Carlos III University of Madrid, 28903 Madrid, Spain; mdurban@est-econ.uc3m.es

* Correspondence: roman.minguez@uclm.es

Abstract: This article introduces the R package `pspatreg`, which is publicly available for download from the Comprehensive R Archive Network, for estimating semiparametric spatial econometric penalized spline (P-Spline) models. These models can incorporate a nonparametric spatiotemporal trend, a spatial lag of the dependent variable, independent variables, noise, and time-series autoregressive noise. The primary functions in this package cover the estimation of P-Spline spatial econometric models using either Restricted Maximum Likelihood (REML) or Maximum Likelihood (ML) methods, as well as the computation of marginal impacts for both parametric and nonparametric terms. Additionally, the package offers methods for the graphical display of estimated nonlinear functions and spatial or spatiotemporal trend maps. Applications to cross-sectional and panel spatial data are provided to illustrate the package's functionality.

Keywords: semiparametric regression; spatial econometrics; P-splines

MSC: 62-04; 62-08; 62G05



Citation: Mínguez, R.; Basile, R.; Durbán, M. Pspatreg: R Package for Semiparametric Spatial Autoregressive Models. *Mathematics* **2024**, *12*, 3598. <https://doi.org/10.3390/math12223598>

Academic Editors: Antonio Musolesi and Alexandra Soberon

Received: 27 September 2024

Revised: 6 November 2024

Accepted: 8 November 2024

Published: 17 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, a new research area in spatial econometrics has emerged in which spatial econometric specifications are combined with semiparametric models [1–4]. These models are highly flexible, enabling the inclusion of the following components within a single specification: (i) Spatial autoregressive terms (i.e., spatial lags of dependent and independent variables as well as spatial error 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; (iii) A spatiotemporal trend, i.e., a smooth interaction between spatial coordinates and the time trend, to capture both spatial and temporal heterogeneity. Although these models are potentially applicable in numerous fields, there is, to the best of our knowledge, no comprehensive computational implementation formally available in the commonly used programming languages. The R package `pspatreg`, available on CRAN and GitHub, provides this functionality along with a user-friendly input/output interface and extensive documentation.

The theoretical background of this package is thoroughly discussed in Mínguez et al. [4] and Basile et al. [2]. These authors recently proposed an approach that combines penalized regression spline methods (see [5], for a recent and complete review of these models) with standard spatial econometric models (such as the Spatial Autoregressive model, or SAR, the Spatial Error Model, or SEM, the Spatial in X-variables model, or SLX, the Spatial Durbin Model, or SDM, the Spatial Error Durbin Model, or SDEM, and the SARAR model, a combination of SAR and SEM); (see LeSage and Pace [6] and Elhorst [7] for a thorough revision). These models address four main econometric issues crucial to modeling spatiotemporal data: functional form bias, spatial dependence bias, spatial heterogeneity bias, and omitted time-related factors bias.

The proposed methodology also enables an Analysis of Variance (ANOVA) decomposition of the spatiotemporal trend into multiple components (spatial and temporal main effects, and second and third-order interactions between them); this was proposed by Lee and Durbán [8] and provides further insight into the dynamics of the data. As a result, the methodology is denoted by the acronym PS-ANOVA-SAR (SEM, SDM, SDEM, SLX, SARAR). The use of nested B-spline bases for interaction components of the spatiotemporal trend enhances the efficiency of the fitting procedure without compromising the model's accuracy. Additionally, we propose an extension of PS-ANOVA-SAR (SEM, SDM, SLX, SARAR), which incorporates a first-order autoregressive term process (AR1) in the noise to account for residual serial correlation. Further theoretical background for the package is available in the online appendix to this paper.

This paper first reviews the statistical methodology for identifying and estimating semiparametric spatial econometric models. Then, it introduces the main functions of the package, covering (a) the estimation and inference of P-Spline regression models, (b) the computation of marginal impacts for parametric and nonparametric terms, and (c) the graphical display of estimated nonlinear functions and spatial trend maps. We also compare `pspatreg`'s functionalities with those of other statistical software for estimating spatial semiparametric regression models. Finally, applications to cross-sectional and panel spatial data demonstrate the usage of these functions.

2. Semiparametric Spatial Econometric Models: A Review

Currently, the prevailing paradigm in spatial econometrics remains largely parametric. The first generation of spatial econometric models—primarily developed to handle cross-sectional data—focused on modeling spatial dependence through various linear specifications, including SAR, SEM, SDM, SLX, and SARAR [6,9]. In the past decade, these models have been extended to account for spatial panel data (or spatiotemporal data), which includes time series observations across geographical units. Elhorst [7] designates these as second generation spatial econometric models. By incorporating region-specific fixed or random effects, these models are particularly useful for controlling unobserved spatial heterogeneity—an essential consideration in empirical economic analyses. Ignoring such heterogeneity can lead to omitted variable biases and obscure causal inferences. More recent advances have included dynamic spatial panel data models and spatial VAR (Vector-Autoregressive) models, which address issues of time persistence and reverse causality.

Despite these substantial advancements, it is essential to recognize that any parametric model is restricted to specific forms of spatial parameter variation, such as spatial regimes. These models are ill-suited for capturing more generalized forms of spatial heterogeneity in model parameters, such as when parameter variation is continuous (smooth) over space or dependent on geographic coordinates and when the functional relationship between dependent and explanatory variables is unknown or non-monotonic. Departing from a strictly parametric approach, another branch of spatial econometric literature has introduced semiparametric methods, which offer more flexible estimation frameworks. This direction follows the recommendations of McMillen [10] to adopt smoother techniques that account for spatial heterogeneity while accommodating potential nonlinearities.

First, following the work of Brunson et al. [11], Cho et al. [12] introduced an approach combining geographically weighted regression (GWR) with spatial autoregression (SEM), known as GWR-SEM. In this approach, the spatial autoregressive error term mitigates spatial dependency, while GWR captures spatial heterogeneity by allowing coefficients to vary across observations. Similarly, Páez et al. [13] proposed an estimation method for cross-sectional data in which the covariance varies locally, effectively addressing spatial autocorrelation in the error terms. Another significant contribution in managing both spatial autocorrelation and parameter nonstationarity was made by Pace and LeSage [14], who proposed a spatial autoregressive local estimation based on a recursive maximum-likelihood estimation approach. This approach enables subsample-based estimation within each observation's neighborhood. More recently, Geniaux and Martinetti [15] introduced

the MGWR-SAR (Mixed Geographically Weighted Regression Simultaneous AutoRegressive Model) by integrating kernel smoothing with standard spatial lag models. This model allows regression parameters and spatial dependence coefficients to vary across space, accommodating cases where some parameters remain constant while others vary spatially.

Second, Montero et al. [1], Basile et al. [2], and Hoshino [3] incorporated penalized regression spline (PS) methods [5] into standard cross-sectional spatial autoregressive models (e.g., SAR, SEM, SDM, and SLX). These models capture spatial interaction effects alongside spatially autocorrelated unobserved heterogeneity, and allow for identifying nonlinear relationships between the dependent variable and the covariates.

Third, Mínguez et al. [4] extended the PS-SAR approach to spatiotemporal data, particularly suited for datasets with large cross-sectional and time series dimensions. Such data enable the estimation of both spatial and spatiotemporal trends in a nonparametric manner [8], capturing region-specific nonlinear time trends while accounting for spatial autocorrelation. This approach helps address questions such as the following: How do unobserved, time-related factors (e.g., economy-wide technological or demand shocks) differentially influence the long-term dynamics across sample units? And how does their inclusion impact the estimation of spatial interaction effects? In this sense, the PS-SAR model with spatiotemporal trends offers an alternative to parametric methods for distinguishing common factors effects (such as shared business cycle influences) from spatial dependence effects (local interactions generating spillover). Here, the former is often termed ‘strong’ cross-sectional dependence, while the latter represents ‘weak’ cross-sectional dependence [16].

In the following subsection, we provide a concise overview of P-Spline models both for cross-sectional and panel data, outlining various model specifications and estimation techniques.

P-Spline Models for Spatial and Spatiotemporal Data

Let $\{(y_i, x_{i1}, \dots, x_{ip}, z_{i1}, \dots, z_{ik})\}_{i=1}^N$ be a set of observations from a Gaussian response variable Y and two collections of p and k covariates X_1, \dots, X_p and Z_1, \dots, Z_k , respectively, and (s_{1i}, s_{2i}) are the spatial coordinates (latitude and longitude) of individual i .

The first specification of semiparametric models for spatial data implemented in the `pspatreg` package is the PS model. In matrix form, it reads as follows:

- PS model

$$\mathbf{y} = f(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{X}\boldsymbol{\beta} + \sum_{l=1}^k g_l(\mathbf{z}_l) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N), \quad (1)$$

where \mathbf{y} is a $N \times 1$ vector including the observations of the dependent variable, \mathbf{X} is a $N \times p$ matrix of parametric effects and $g_l(\mathbf{z}_l)$ are smooth functions that account for the non-linear effects of the covariates $\mathbf{z}_1, \dots, \mathbf{z}_k$. The term $f(\mathbf{s}_1, \mathbf{s}_2)$ is a two-dimensional smooth function of the spatial coordinates \mathbf{s}_1 and \mathbf{s}_2 that collects the potential spatial heterogeneity usually present when working with spatial data.

All the nonparametric terms in this and all subsequent models are modeled using Penalized-splines (P-Splines), as described in Eilers and Marx [17]. This methodology assumes that each unknown function to be estimated is smooth and can be represented as a linear combination of basis functions and a penalty on the coefficients is added to the likelihood to control the smoothness of the estimated function. A popular choice for the basis is the use of cubic B-spline [18]:

$$g_l(\mathbf{z}_l) = \sum_{j=1}^{\kappa_l} B_j(\mathbf{z}_l) \theta_j, \quad l = 1, \dots, k,$$

where κ_l is the number of knots used to construct the basis for each covariate. In the case of P-splines, the selection of the number and position of the knots is not an issue: the idea is to select a large enough number of equally spaced knots and let the penalty control for the smoothness. Ruppert [19] proposes a general rule of thumb for the number of knots:

use the $\min(40, n/4)$, where n is the number of unique values of the covariate. The penalty can take several forms, depending on the prior knowledge of the curve/surface. The most common choice is to use second-order differences on adjacent B-spline coefficients (see [4]). In the case of an univariate function, i.e., $g_l(\mathbf{z}_l)$, the penalty would be:

$$\text{Penalty} = \lambda \sum_j (\theta_j - 2\theta_{j-1} + \theta_{j-2})^2,$$

where λ is the smoothing parameter that control the roughness of the function to be estimated and prevents overfitting. This parameter will be estimated via maximum likelihood as the rest of the parameters in the model.

In the multidimensional setting, i.e., $f(\mathbf{s}_1, \mathbf{s}_2)$, the basis is computed as a tensor product of the marginal basis functions in each dimension, and the penalty is a kronecker sum of penalties controlled by as many smoothing parameters as the dimension of the function to be estimated.

The PS model can be seen as the baseline model in the context of semiparametric spatial autoregressive models and can be extended to include spatial lags of the dependent or independent variables (both parametric and nonparametric) or the model’s noise in a way similar to the standard spatial econometric models [6,7]. The semiparametric spatial econometric models are called PS-SLX, PS-SAR, PS-SEM, PS-SDM, PS-SDEM, and PS-SARAR to reflect the combination of semiparametric models via P-splines with the most common spatial econometric models (already mentioned in Section 1).

- PS-SLX model

This specification includes a spatial lag of the parametric and nonparametric covariates. As usual in spatial econometrics, \mathbf{W} represents a $N \times N$ row-standardized spatial weights matrix:

$$\mathbf{y} = f(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\beta}^* + \sum_{i=1}^k g_i(\mathbf{z}_i) + \sum_{i=1}^k g_i^*(\mathbf{W}\mathbf{z}_i) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N). \quad (2)$$

- PS-SAR model (PS-spatial autorregresive model)

In this case, a spatial lag only for the dependent variable ($\mathbf{W}\mathbf{y}$) is added to the PS model. This term implies that the response variable is influenced by the values of other observations based on their geographic proximity, and it is controlled by the parameter ρ :

$$\mathbf{y} = f(\mathbf{s}_1, \mathbf{s}_2) + \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^k g_i(\mathbf{z}_i) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N). \quad (3)$$

- PS-SEM model (PS-spatial error model)

The systematic part of the PS-SEM model is the same as the PS model, but the errors (rather than the response variable) are affected by spatial autocorrelation:

$$\begin{aligned} \mathbf{y} &= f(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^k g_i(\mathbf{z}_i) + \mathbf{u} \\ \mathbf{u} &= \delta \mathbf{W}\mathbf{u} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N). \end{aligned} \quad (4)$$

- PS-SDM model (PS-spatial Durbin model)

A spatial lag of the dependent and independent variables (both parametric and nonparametric) is included in this case. This specification nests both PS-SAR and PS-SEM models (the last one as a nonlinear constraint of PS-SDM). A particular case of this model can be specified when the spatial lag is included only for a subset of the covariates (either parametric or nonparametric).

$$\mathbf{y} = f(\mathbf{s}_1, \mathbf{s}_2) + \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\beta}^* + \sum_{i=1}^k g_i(\mathbf{z}_i) + \sum_{i=1}^k g_i^*(\mathbf{W}\mathbf{z}_i) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N). \quad (5)$$

- PS-SDEM model (PS-spatial Durbin error model)

This model includes a spatial lag of the independent covariates (parametric and nonparametric) and the error model, and it can be seen as a combination of PS-SLX and PS-SEM:

$$\begin{aligned}
 \mathbf{y} &= f(\mathbf{s}_1, \mathbf{s}_2) + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{X}\boldsymbol{\beta}^* + \sum_{i=1}^k g_i(\mathbf{z}_i) + \sum_{i=1}^k g_i^*(\mathbf{W}\mathbf{z}_i) + \mathbf{u} \\
 \mathbf{u} &= \delta\mathbf{W}\mathbf{u} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N).
 \end{aligned}
 \tag{6}$$

- PS-SARAR model

This specification allows for the inclusion of a spatial lag of both the dependent variable and the model noise, combining models PS-SAR and PS-SEM:

$$\begin{aligned}
 \mathbf{y} &= f(\mathbf{s}_1, \mathbf{s}_2) + \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^k g_i(\mathbf{z}_i) + \mathbf{u} \\
 \mathbf{u} &= \delta\mathbf{W}\mathbf{u} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_N).
 \end{aligned}
 \tag{7}$$

All model parameters can be estimated using either Restricted Maximum Likelihood (REML) or Maximum Likelihood (ML). We exploit the reparameterization of the P-spline as a mixed model [20] via the singular value decomposition (SVD) of the penalty matrix. This reparameterization allows us to impose the necessary constraints so that all terms in the model are identifiable. Furthermore, it will enable us to add spatial (or other) random effects if needed. Once the nonparametric terms have been reparameterized, REML or ML is used to obtain estimates of variance components and random effects (for nonparametric covariates and spatial trends) and β parameters (for parametric covariates). Conditioning on these estimates, the spatial parameters ρ and δ , which are associated to spatial lags of the dependent variable or the noise, are estimated using numerical optimization. This process is iterated until convergence. The algorithm is summarized in the following steps (see Mínguez et al. [4] for more details):

1. Put the full model as a mixed model [21,22].
2. Use SAP (Separation of Anisotropic Penalties) algorithm [23] to obtain REML (or ML) estimates of the variance components, random effects, and β parameters of the mixed model. This algorithm is high-speed and it does not require numerical optimization.
3. Conditioning on previous estimates, use numerical optimization to estimate spatial parameters, i.e., ρ and/or δ , associated to spatial lags of the dependent variable and/or the model’s noise. We use the `bobyqa()` optimizer implemented in `minqa` package [24].
4. Iterate 2 and 3 steps until convergence.

At convergence, the standard errors of $\hat{\beta}$ are obtained, conditioned on the estimated variance components, as usual in mixed models (see, for example, pp. 396 in [25]). The standard errors of $\hat{\rho}$ and $\hat{\delta}$ are obtained by computing the inverse of the information matrix at convergence.

When using spatiotemporal data, we can extend the spatial trend incorporating a temporal coordinate τ with dimensions $T \times 1$. In this way, the spatiotemporal trend is defined as $f(\mathbf{s}_1, \mathbf{s}_2, \tau)$. Consequently, the row-dimensions of \mathbf{y} , \mathbf{X} and \mathbf{z}_i are NT instead of N . Additionally, it is necessary to extend the spatial lags over time by using the Kronecker product of the spatial weighting matrix, \mathbf{W}_N , and the identity matrix, \mathbf{I}_T , where N and T represent the lengths of the spatial and temporal coordinates, respectively. As an example, the PS-SAR model for spatiotemporal data is the following:

$$\mathbf{y} = f(\mathbf{s}_1, \mathbf{s}_2, \tau) + (\rho\mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^k g_i(\mathbf{z}_i) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{R}_{TN}). \tag{8}$$

In this case the error term, ϵ , can be generated either by a white noise ($\mathbf{R}_{TN} = \mathbf{I}_{TN}$) or an autoregressive process, AR(1). The last case is intended to gather the temporal correlation usually present when working with spatiotemporal data.

Typically, estimating the spatiotemporal trend requires a large sample of data. However, sometimes, there may be data gaps for the combination of spatial and temporal coordinates. In addition, there may be computing power issues to estimate the spatiotemporal trend. To speed up the computational process, we can perform a functional analysis of variance decomposition (ANOVA) of the spatio-temporal trend into the main function, second-order interaction, and third-order interaction functions [26]. The ANOVA decomposition of the spatio-temporal trend is defined as follows:

$$f(\mathbf{s}_1, \mathbf{s}_2, \tau) = f_1(\mathbf{s}_1) + f_2(\mathbf{s}_2) + f_t(\tau) + f_{12}(\mathbf{s}_1, \mathbf{s}_2) + f_{1t}(\mathbf{s}_1, \tau) + f_{2t}(\mathbf{s}_2, \tau) + f_{12t}(\mathbf{s}_1, \mathbf{s}_2, \tau)$$

where the main and interaction functions are the following:

- Main functions: $f_1(\mathbf{s}_1)$, $f_2(\mathbf{s}_2)$ and $f_t(\tau)$.
- Second-order interactions: $f_{12}(\mathbf{s}_1, \mathbf{s}_2)$, $f_{1t}(\mathbf{s}_1, \tau)$ and $f_{2t}(\mathbf{s}_2, \tau)$.
- Third-order interaction: $f_{12t}(\mathbf{s}_1, \mathbf{s}_2, \tau)$.

Each function in this ANOVA decomposition can be estimated using P-Splines with different penalty coefficients. Furthermore, the maximum number of knots can be different for each term, allowing more degrees of freedom (i.e., more flexibility) for the main functions and fewer degrees of freedom for the interaction functions. For this purpose, we follow Lee et al. [26] who proposed that the basis generating the interaction terms is a nested basis of the main terms in the ANOVA decomposition. Accordingly, the maximum number of knots in the spline basis generating the interaction terms is a divisor of the number of knots in the spline basis generating the main terms. Therefore, the space spanned for the splines basis in the higher interaction terms is a subset of the subspace spanned for the splines basis generating the lower terms. This is usually a good compromise between flexibility and simplification of the computations. Of course, the ANOVA decomposition can also be performed for the pure spatial trend, allowing only two main functions, $f_1(\mathbf{s}_1)$ and $f_2(\mathbf{s}_2)$, and a second-order interaction function, $f_{12}(\mathbf{s}_1, \mathbf{s}_2)$, for the spatial coordinates.

3. Main Functions of the Package Pspatreg

3.1. The Function Pspatfit()

The main function in the `pspatreg` package is `pspatfit()`, which estimates spatiotemporal penalized spline spatial regression models using either Restricted Maximum Likelihood (REML) or Maximum Likelihood (ML) methods. In its generic form `pspatfit()` appears as the following:

```
pspatfit(formula, data, na.action, listw = NULL, type = "sim",
method = "Chebyshev", Durbin = NULL, zero.policy = NULL, interval = NULL,
trs = NULL, cor = "none", dynamic = FALSE, control = list())
```

The `pspatfit()` function returns an object of the `pspatreg` class containing a list of coefficients of the parametric terms and their standard errors, estimated coefficients corresponding to random effects in a mixed model and their standard errors, equivalent degrees of freedom, residuals, fitted values, etc. A wide range of standard methods is also available for the `pspatreg` class, including `print()`, `summary()`, `coef()`, `vcov()`, `anova()`, `fitted()`, `residuals()` and `plot()`.

3.1.1. The Argument Formula

The argument `formula`, within the function `pspatfit()`, is a formula similar to the GAM specification, including parametric and nonparametric terms. Parametric covariates are usually included. P-spline smooth terms are specified using `pspl()` and `pspt()` for the nonparametric covariates and spatial or spatiotemporal trends, respectively. For example:

```
formula <- y ~ x1 + x2 + pspl(x3, nknots = 15) + pspl(x4, nknots = 20) +
  pspt(long, lat, year, nknots = c(18,18,8),
    psanova = TRUE, nest_sp1 = c(1, 2, 3),
    nest_sp2 = c(1, 2, 3), nest_time = c(1, 2, 2))
```

In the example above, the model includes two parametric terms (x_1 and x_2), two nonparametric terms (x_3 and x_4), and a spatiotemporal trend (with *long* and *lat* as spatial coordinates, and *year* as the temporal coordinate). The dimension of the basis function in both `pspl()` and `pspt()` is defined by `nknots`. This term should not be less than the dimension of the null space of the penalty for the term (see `?null.space.dimension` and `?choose.k` in the `mgcv` package to know how to choose `nknots`). The default number of `nknots` in `pspl()` is 10. However, in this example, we have chosen 15 `nknots` for $g_1(x_3)$ and 20 `nknots` for $g_2(x_4)$. The default number of `nknots` in `pspt()` is `c(10,10,5)`, but we have chosen `c(18,18,8)`.

In this example, we also apply an ANOVA decomposition of the spatiotemporal trend (choosing `psanova = TRUE` within `pspt()`). Each effect has its degree of smoothing, allowing greater flexibility for the spatiotemporal trend. Calculating up to third-order interactions can be computationally expensive. To overcome this problem, we can select subgroups of interaction effects for the second and third-order effects. To define these subgroups, we use three parameters available in `pspt()`: `nest_sp1`, `nest_sp2`, and `nest_time`. These parameters indicate the divisors of the `nknots` chosen for the main functions. For example, if we set `nest_sp1 = c(1,2,3)`, we will have all knots for the s_1 main function, $18/2$ for each second-order interaction function with s_1 , and $18/3$ knots for the third-order interaction function with s_1 (in most empirical cases, the main functions are more flexible than interaction functions and, therefore, the number of knots in B-Spline bases for interaction effects need not to be as large as the number of knots for the main functions [26]).

If we want to exclude a function from the ANOVA decomposition, we have to set the parameters `f12_int`, `f1t_int`, `f2t_int`, `f12t_int` to `FALSE` (for interaction functions) or `f1_main`, `f2_main`, `ft_main` to `FALSE` for main functions.

3.1.2. Data Structure

The argument `data` must contain a `data.frame` or a `sf` object, including all the variables for both parametric and nonparametric terms of the model. If a `pspt()` term is included in `formula`, the data must contain the spatial and temporal coordinates specified in `pspt()` (e.g., latitude, longitude, and year). In this case, the coordinates must be ordered in the `data.frame` or `sf`, with time as the fast index and spatial coordinates as the slow index.

Both `data.frame` (or `tbl_df`) and `sf` class objects can be used as data inputs, but `sf` objects are highly recommended as they allow the user to map spatial trends.

3.1.3. Spatial Arguments

The argument `type` allows us to select different spatial econometric model specifications: `"sar"`, `"sem"`, `"sdm"`, `"sdem"`, `"sarr"`, or `"slx"`. When creating a `"slx"`, `"sdem"` or `"sdm"` model, we need to include the formula of the Durbin part in the `Durbin` argument (see `?spSURml` in `spSUR` package for more details). In the same spirit as the well-established standard packages for spatial econometrics (i.e., `spatialreg`, `spdep`), the argument `listw` is used to include a list of neighbors, and the argument `method` is used to choose the method to compute the jacobian of the spatial lag. When dealing with spatiotemporal data, the argument `cor` is used to specify temporal correlation in the model's noise. More specifically, for spatial panel data, the argument `index` specifies the indices for cross-sectional (spatial) units and temporal coordinates. Meanwhile, the arguments `demean` and `eff_demean` are used to demean the panel data, and the argument `dynamic` is used to specify a dynamic spatial panel data model. Finally, the `control` argument contains some options related to the optimization process and other technical details. The function help (`?pspatfit`)

provides extensive documentation on every argument and examples of spatial and spatiotemporal cases.

3.2. Plotting Smooth Terms

Plotting the estimated nonparametric smooth terms is essential in semiparametric regression analyses. First, once a model has been estimated using `pspatfit()`, the function `fit_terms()` computes fitted nonparametric smooth terms, which can be plotted with the function `plot_terms()`. Furthermore, the functions `plot_sp2d()` and `plot_sp3d()` are used to plot maps of spatial and spatiotemporal trends, respectively, while `plot_sptime()` is used to plot a different time trend for each spatial unit.

3.3. Computing Impact Functions

In the case of a semiparametric model without the spatial lag of the dependent variable (PS model), when all regressors are manipulated independently of the errors, the $\mathbf{X}\hat{\beta}$ (for parametric covariates) or $\hat{g}(\mathbf{z})$ (for nonparametric covariates) can be interpreted as the conditional expectation (net of the effect of the other regressors) of \mathbf{y} given either \mathbf{x} or \mathbf{z} , respectively. Blundell and Powell [27] use the term Average Structural Function (ASF) to refer to these functions. However, in PS-SAR, PS-SDM, and PS-SARAR models, when ρ is different from zero, the estimated smooth functions for nonparametric covariates cannot be interpreted as ASF. Taking advantage of the results obtained for parametric covariates in spatial econometric models (see LeSage and Pace [6]), we can extend the idea of impacts to nonparametric terms. Specifically, for PS-SAR or PS-SARAR models, we compute the total impact function of a nonparametric covariate \mathbf{z} as follows:

$$\hat{g}^T(\mathbf{z}) = [\mathbf{I}_N - \hat{\rho}\mathbf{W}]^{-1} \sum_k B_k(\mathbf{z})\hat{\theta}_k$$

where $B_k(\mathbf{z})$ are P-spline basis functions, and $\hat{\theta}_k$ the corresponding estimated parameters.

We can also compute the direct impact function of smooth terms for the PS-SAR or PS-SARAR cases as follows:

$$\hat{g}^D(\mathbf{z}) = \text{diag}([\mathbf{I}_N - \hat{\rho}\mathbf{W}]^{-1}) \sum_k B_k(\mathbf{z})\hat{\theta}_k$$

where $\text{diag}(\mathbf{A})$ is the diagonal matrix, including the elements of the main diagonal in \mathbf{A} .

Eventually, the indirect (or spillover) impacts function for the previous cases can be obtained as follows:

$$\hat{g}^I = \hat{g}^T - \hat{g}^D$$

Similar expressions can be provided for the total, direct, and indirect impact functions of the ps-sdm including $\sum_k B_k(\mathbf{z})\hat{\theta}_k + \sum_{k^*} B_{k^*}^*(\mathbf{W}\mathbf{z})\hat{\theta}_{k^*}$ instead of $\sum_k B_k(\mathbf{z})\hat{\theta}_k$ in the previous expressions. Eventually, the expressions for ps-slx and ps-sdem models are similar to ps-sdm case but excluding $[\mathbf{I}_N - \hat{\rho}\mathbf{W}]^{-1}$.

The function `impactspar()` in the `pspatreg` package computes direct, indirect, and total impacts for continuous parametric covariates using the standard procedure (see [6] for details). On the other hand, the function `impactsnopar()` computes direct, indirect, and total impact functions for continuous nonparametric covariates, while the function `plot_impactsnopar()` is used to plot these impact functions. It is worth noticing that total, direct, and indirect impact functions are never smooth over the domain of the variable \mathbf{z} due to the presence of the spatial multiplier matrix in the algorithm for their computation. Indeed, a wiggly profile of direct, indirect, and total impacts would appear even if the model were linear. Therefore, in the spirit of the semiparametric approach, we include the possibility of applying a smoother to obtain smooth curves using the argument `smooth = TRUE` (default) in `plot_impactsnopar()`.

4. What Can Pspatreg Offer More than Other Software?

Several packages have recently been proposed to perform spatial econometrics in R (see [28], for a survey). Focusing on packages and methods dealing with polygonal (or areal) spatial data, the first package was `spdep` [29,30]. It is primarily designed for cross-sectional spatial data and for modeling spatial dependence (SAR, SEM, or SARAR specifications) through Maximum Likelihood (ML) or Generalized Method of Moments (GMM). The estimation functions from `spdep` have been moved to the `spatialreg` package [28]. Other spatial econometric methods for cross-sectional data have been implemented in the following packages: `sphet` [31] for estimating and testing spatial models with heteroskedastic innovations, `spfilter` [32] for filtering out spatial dependence in linear models, `spgwr` [33] for estimating geographically weighted regression models, and `spsur` [34] for estimating seemingly unrelated regression equations. Other R packages for spatial econometric analysis have been developed following several theoretical contributions to the literature on static and dynamic spatial panel data models (see [7], for a review). In particular, `splm` [35] and `SDPDM` [36] implement estimation methods for static and dynamic spatial panel data. All of these R packages focus on parametric methods (except `spgwr`, of course), leaving aside issues related to nonlinearities in functional form and the estimation of spatial or spatiotemporal trends. The Appendix to this article reports comparisons of the estimates for pure parametric models between `spatialreg` (for spatial data) and `splm` (for spatial panel data). As expected, the estimates are practically the same in the case of pure parametric models.

On the other hand, `mgcv` [22] is the main reference for dealing with semiparametric and nonparametric models using splines in R. Nevertheless, to our knowledge, neither this package nor other good alternative packages using splines, such as `SpATS` [37] or `JOPS` [38], allow the treatment of semiparametric models when spatial lags either in dependent or independent variables or the noise are included in the model. In this sense, our proposal aims to combine both perspectives, allowing the modeling of semiparametric spatial or spatiotemporal specifications. However, `pspatreg` at the moment only allows for the estimation of models with a Gaussian dependent variable. Additionally, the `SDPDM` package allows for the specification of more general dynamic spatial panels than `pspatreg` and supports estimation using both ML and Bayesian methods. Finally, the `spatialreg` package is more generalist than `pspatreg` and, consequently, provides a much more extensive array of functions for handling spatial data. The following table summarizes the models mentioned in the previous section and the possible packages used to deal with them (see Table 1).

Table 1. Summary of semiparametric models and R packages to deal with them. In the table, *par. (2d)* is used for parametric models dealing with spatial data and *par. (panel)* is used for parametric models dealing with spatial panel data. The term *full* is used when the package can estimate both the full spatial and spatiotemporal semiparametric models.

Model	<code>spatialreg/sphet</code>	<code>splm/SDPDM</code>	<code>pspatreg</code>	<code>mgcv/SpATS/JOPS</code>
ps	par. (2d)	par. (panel)	full	full
ps-slx	par. (2d)	par. (panel)	full	full
ps-sar	par. (2d)	par. (panel)	full	-
ps-sem	par. (2d)	par. (panel)/-	full	-
ps-sdm	par. (2d)	par. (panel)	full	-
ps-sdem	par. (2d)	par. (panel)/-	full	-
ps-sarar	par. (2d)	par. (panel)/-	full	-

Other functions and toolboxes to work with spatial econometric models and spatial panel models are available in Stata [39–41] and Matlab [42–44]. However, as far as we know, they only deal with purely parametric specifications in the same way as `spatialreg` and `splm`.

5. Examples for Cross-Sectional Data (2d)

We start our demo by applying `pspatreg` to the analysis of cross-sectional data (2d). The purpose of this example is to illustrate the significant advantages of estimating semi-parametric spatial models over traditional parametric ones. In a cross-sectional setting, traditional models not only impose linearity in the functional form but also, due to the lack of a time dimension in the data, fail to account for unobserved heterogeneity. In contrast, using the `pspatreg` package, we can address spatial heterogeneity (through a spatial trend function) and account for nonlinearities in the relationship between the dependent variable and covariates. We demonstrate that controlling for both sources of model misspecification allows for more accurate and reliable identification of spatial spillovers.

In particular, we use the dataset *ames*, included in the `AmesHousing` package [45], which contains data on 2930 properties in Ames, Iowa. The dataset includes information on house characteristics (bedrooms, garages, fireplaces, pools, porches, etc.), location characteristics (neighborhood), lot information (zoning, shape, size, etc.), ratings of condition and quality, and sale price (from 2006 to 2010).

5.1. Reading the Data

First, *ames* should be transformed into a spatial object of class `sf`:

```
library(sf)
library(spatialreg)
library(pspatreg)
ames_sf <- st_as_sf(AmesHousing::make_ames(), coords = c("Longitude",
"Latitude"))
```

The dependent variable in the regression analysis is *Sale_Price*, while we selected the following variables as covariates:

- *Lot_Area*: Lot size in square feet.
- *Total_Bsmt_SF*: Total square feet of basement area.
- *Garage_Cars*: Size of garage in car capacity.
- *Gr_Liv_Area*: Above grade (ground) living area square feet.
- *Fireplaces*: Number of fireplaces.

Due to the skewed distributions of the dependent variable *Sale_Price* and continuous covariates, we use the log-transformation of them (except *Garage_Cars* and *Fireplaces*):

```
ames_sf$lnSale_Price <- log(ames_sf$Sale_Price)
ames_sf$lnLot_Area <- log(ames_sf$Lot_Area)
ames_sf$lnTotal_Bsmt_SF <- log(ames_sf$Total_Bsmt_SF + 1)
ames_sf$lnGr_Liv_Area <- log(ames_sf$Gr_Liv_Area)
```

5.2. Constructing the Spatial Weights Matrix

Creating spatial weights is a necessary step when using areal data. This involves selecting a criterion to define neighbors and assigning weights to each. In particular, we have used the criterion of nearest neighbors after eliminating duplicate coordinate values:

```
coord <- st_coordinates(ames_sf)
ames_sf$Longitude <- coord[, 1]
ames_sf$Latitude <- coord[, 2]
ames_sf1 <- ames_sf[duplicated(ames_sf$Longitude) == FALSE, ]
library(spdep)
nbames <- knn2nb(knearneigh(st_coordinates(ames_sf1), k = 5,
longlat = TRUE), sym = TRUE)
lwames <- nb2listw(nbames, style = "W", zero.policy = FALSE)
```

It should be noted that since the argument `sym` has been set to `TRUE`, the number of neighbors differs for each spatial unit (see the output of `summary(lwames)` for details).

5.3. Estimating Parametric Spatial Linear Models

We first estimate parametric spatial linear models using the function `pspatfit()` included in the `pspatreg` package and compare them with the results obtained using the functions provided by the `spatialreg` package.

The linear SAR model for cross-sectional data can be written as follows:

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N).$$

To estimate this model, we use `pspatfit()` with the argument `type="sar"`:

```
flin <- lnSale_Price ~ lnLot_Area + lnTotal_Bsmt_SF + lnGr_Liv_Area +
  Garage_Cars + Fireplaces
linsar <- pspatfit(flin, data = ames_sf1, listw = lwames,
  method = "Chebyshev", type = "sar")
summary(linsar)

##
## Call
## pspatfit(formula = flin, data = ames_sf1, listw = lwames, type = "sar",
## method = "Chebyshev")
##
## Parametric Terms
##
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.4596366 0.0991499 24.8073 < 2.2e-16 ***
## lnLot_Area 0.0296596 0.0071596 4.1427 3.536e-05 ***
## lnTotal_Bsmt_SF 0.0488238 0.0029233 16.7019 < 2.2e-16 ***
## lnGr_Liv_Area 0.3773085 0.0130689 28.8706 < 2.2e-16 ***
## Garage_Cars 0.0943737 0.0051695 18.2559 < 2.2e-16 ***
## Fireplaces 0.0486969 0.0058596 8.3106 < 2.2e-16 ***
## rho 0.5019890 0.0123499 40.6471 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Goodness-of-Fit
##
## EDF Total: 7
## Sigma: 0.201487
## AIC: -6745.73
## BIC: -6704.22
```

As expected, all $\hat{\beta}'s$ are significant and positive. The estimated spatial autoregressive parameter (0.50) is also positive and significant. The output of `pspatfit()` is an object of class `pspatreg` for which the usual methods are defined: `coef()`, `fitted()`, `residuals()`, `loglik()`, `anova()`, `vcov()` and `print()`.

Once the spatial model has been estimated, the function `impactspar()` allows the average direct, indirect, and total impacts to be obtained for the parametric covariates [6]:

```
Wames <- as(lwames, "CsparseMatrix")
trMatc <- trW(Wames, type="mult")
set.seed(123456)
implinsar <- impactspar(linsar, tr = trMatc)
summary(implinsar)
```

```

##
## Total Parametric Impacts (sar)
##           Estimate Std. Error   t value Pr(>|t|)
## lnLot_Area      0.0603784  0.0146613  4.1182030    0
## lnTotal_Bsmt_SF 0.0982732  0.0063178 15.5550686    0
## lnGr_Liv_Area   0.7581528  0.0327675 23.1373076    0
## Garage_Cars     0.1896385  0.0114581 16.5505853    0
## Fireplaces      0.0977258  0.0119669  8.1663537    0
##
## Direct Parametric Impacts (sar)
##           Estimate Std. Error   t value Pr(>|t|)
## lnLot_Area      0.0319102  0.0076702  4.1602840    0
## lnTotal_Bsmt_SF 0.0519511  0.0030954 16.7831818    0
## lnGr_Liv_Area   0.4007849  0.0142408 28.1434593    0
## Garage_Cars     0.1002486  0.0055287 18.1322833    0
## Fireplaces      0.0516534  0.0061226  8.4365332    0
##
## Indirect Parametric Impacts (sar)
##           Estimate Std. Error   t value Pr(>|t|)
## lnLot_Area      0.0284682  0.0070539  4.0357950 1e-04
## lnTotal_Bsmt_SF 0.0463221  0.0035330 13.1113879 0e+00
## lnGr_Liv_Area   0.3573679  0.0213565 16.7334190 0e+00
## Garage_Cars     0.0893899  0.0065451 13.6575927 0e+00
## Fireplaces      0.0460724  0.0060279  7.6431596 0e+00

```

As expected, all marginal effects are highly significant, and spillovers are relatively high. We compare these results with those obtained using ML estimates with the function `lagsarlm()` from the package `spatialreg`:

```

linsarspt <- lagsarlm(flin, data = ames_sf1, listw = lwames, method = "Chebyshev")
impacts(linsarspt, tr = trMatc)

```

```

## Impact measures (lag, trace):
##           Direct   Indirect   Total
## lnLot_Area      0.03148261 0.02804245 0.05952506
## lnTotal_Bsmt_SF 0.05184279 0.04617784 0.09802063
## lnGr_Liv_Area   0.40063291 0.35685505 0.75748796
## Garage_Cars     0.10012298 0.08918237 0.18930535
## Fireplaces      0.05168978 0.04604155 0.09773133

```

As expected, the marginal impacts obtained with both packages (`pspatreg` and `spatialreg`) are virtually the same for parametric covariates. Minor numerical differences may be due to the estimation method (REML versus ML) as well as the optimization algorithms used. Similar comparisons can be checked for the rest of the spatial linear econometric models (SLX, SEM, SDM, and SARAR). The appendix provides some additional evidence of the SLX, SEM, and SDM models being compared.

The equality of the results obtained with the two packages implies that the new package `pspatreg` can be used to compare the performance of parametric and semi-parametric models without resorting to alternative packages for estimating parametric models. Specifically, comparison tests for parametric and semi-parametric models (via AIC, BIC, and LR-Test) can be easily conducted using only `pspatreg`.

5.4. Estimating Semiparametric Nonlinear Models With and Without a Spatial Trend

We now provide examples of the estimation of semiparametric models. Let us start with a simple semiparametric model without spatial trends and spatially lagged terms that is, an additive model (or PS model in our terminology):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^k g_i(\mathbf{z}_i) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N),$$

In particular, we introduce the discrete variables `Fireplaces` and `Garage_Cars` as linear terms and the continuous variables `lnLot_Area`, `lnTotal_Bsmt_SF`, and `lnGr_Liv_Area` as smooth covariates, adding `pspl()` terms with 20 knots to the formula:

```
fps <- lnSale_Price ~ Fireplaces + Garage_Cars +
  pspl(lnLot_Area, nknots = 20) + pspl(lnTotal_Bsmt_SF, nknots = 20) +
  pspl(lnGr_Liv_Area, nknots = 20)
ps <- pspatfit(fps, data = ames_sf1)
summary(ps)
```

```
##
## Call
## pspatfit(formula = fps, data = ames_sf1)
##
## Parametric Terms
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.5806060  0.0679093  170.5304 < 2.2e-16 ***
## Fireplaces      0.0709426  0.0069756   10.1700 < 2.2e-16 ***
## Garage_Cars     0.1588650  0.0063789   24.9047 < 2.2e-16 ***
## pspl(lnLot_Area) 0.0156563  0.0122900    1.2739  0.20280
## pspl(lnTotal_Bsmt_SF) 0.1213197  0.0239065    5.0748 4.139e-07 ***
## pspl(lnGr_Liv_Area) 0.0516587  0.0237088    2.1789  0.02943 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Non-Parametric Terms
##
##              EDF
## pspl(lnLot_Area)    9.5779
## pspl(lnTotal_Bsmt_SF) 3.6633
## pspl(lnGr_Liv_Area) 12.8471
##
## Goodness-of-Fit
##
## EDF Total: 32.0883
## Sigma: 0.20342
## AIC: -5896.24
## BIC: -5705.98
```

The EDF numbers suggest that the three continuous variables enter the model nonlinearly. It should be noted that smooth terms in P-splines modeling also have a linear part because of the penalty term (in the output, this linear part is included as a parametric term).

Now, we introduce the spatial lag of the dependent variable, thus specifying a semi-parametric SAR model (PS-SAR):

$$\mathbf{y} = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^k g_i(\mathbf{z}_i) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N),$$

The PS-SAR model can be estimated using `pspatfit()` with the previous formula and choosing argument `type = "sar"`.


```

ps_sar <- pspatfit(fps, data = ames_sf1, listw = lwames,
method = "Chebyshev", type = "sar")
summary(ps_sar)

##
## Call
## pspatfit(formula = fps, data = ames_sf1, listw = lwames, type = "sar",
## method = "Chebyshev")
##
## Parametric Terms
##
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.1642391 0.0492228 125.2313 < 2.2e-16 ***
## Fireplaces 0.0469980 0.0056748 8.2818 < 2.2e-16 ***
## Garage_Cars 0.0855018 0.0051776 16.5139 < 2.2e-16 ***
## pspl(lnLot_Area) 0.0200760 0.0090734 2.2126 0.02701 *
## pspl(lnTotal_Bsmt_SF) 0.0886416 0.0149158 5.9428 3.155e-09 ***
## pspl(lnGr_Liv_Area) 0.0412042 0.0196408 2.0979 0.03601 *
## rho 0.4652605 0.0125712 37.0099 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Non-Parametric Terms
## EDF
## pspl(lnLot_Area) 6.7873
## pspl(lnTotal_Bsmt_SF) 3.1119
## pspl(lnGr_Liv_Area) 13.8152
##
## Goodness-of-Fit
##
## EDF Total: 30.7144
## Sigma: 0.184702
## AIC: -6927.36
## BIC: -6745.25

```

The spatial spillover parameter is now 0.46, lower than the one estimated with the linear SAR (0.50), confirming the trade-off between nonlinearities and spatial dependence [2].

Let us now introduce also a spatial trend 2d (without the ANOVA decomposition) in order to control for unobserved spatial heterogeneity:

$$y = f(s_1, s_2) + \rho W y + X\beta + \sum_{i=1}^k g_i(z_i) + \epsilon \quad \epsilon \sim N(0, \sigma^2 \mathbf{I}_N),$$

To estimate this model, it is necessary to introduce a `pspt()` term in the formula. In the output, the terms named *Xspt* correspond to the linear part of the spatial trend:

```

fpsp2d <- update(fps, . ~ . + pspt(Longitude, Latitude, nknots = c(10, 10),
psanova = FALSE))
psp2d_sar <- pspatfit(fpsp2d, data = ames_sf1, listw = lwames,
method = "Chebyshev", type = "sar")
summary(psp2d_sar)

##
## Call
## pspatfit(formula = fpsp2d, data = ames_sf1, listw = lwames, type = "sar",
## method = "Chebyshev")

```

```

##
## Parametric Terms
##
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.47756924 0.12220364 69.3725 < 2.2e-16 ***
## Fireplaces 0.05454867 0.00579113 9.4193 < 2.2e-16 ***
## Garage_Cars 0.06563134 0.00567565 11.5637 < 2.2e-16 ***
## Xspt.2 -0.10795583 0.11751199 -0.9187 0.3583452
## Xspt.3 -0.00080787 0.11692482 -0.0069 0.9944877
## Xspt.4 -0.02287410 0.11385028 -0.2009 0.8407810
## pspl(lnLot_Area) 0.03139985 0.00943208 3.3290 0.0008831 ***
## pspl(lnTotal_Bsmt_SF) 0.08678510 0.01744709 4.9742 6.962e-07 ***
## pspl(lnGr_Liv_Area) 0.05583719 0.02002986 2.7877 0.0053455 **
## rho 0.29097232 0.01749159 16.6350 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Non-Parametric Terms
##
## EDF
## pspl(lnLot_Area) 6.6923
## pspl(lnTotal_Bsmt_SF) 3.4601
## pspl(lnGr_Liv_Area) 14.1193
##
## Non-Parametric Spatio-Temporal Trend
##
## EDF
## f(sp1, sp2) 36.782
##
## Goodness-of-Fit
##
## EDF Total: 71.054
## Sigma: 0.182477
## AIC: -7001.36
## BIC: -6580.08

```

The estimated spatial spillover parameter $\hat{\rho}$ (0.29) is much lower than the previous ones, suggesting that the SAR model without spatial trend (linear and nonlinear) captures spatial autocorrelated unobserved heterogeneity. Furthermore, the EDF of the spatial trend (41.105) points out the need to include a nonlinear spatial trend. Later, we compare different specifications based on information criteria. For semiparametric models, in addition to calculating the spatial impacts for the parametric covariates (as we have done previously for the linear SAR model), it is also possible to calculate impact functions for the nonparametric covariates. We can plot the estimated smooth impact functions using the abovementioned algorithms for the three nonparametric terms. The function `impactsnpar()` computes impact functions for all smooth functions and allows us to obtain the plots. We only show the impact functions of *lnGr_Liv_Area* to reduce the size of the output (by default, all the nonparametric covariates are included) (see Figure 1):

```

impvsnpar <- impactsnpar(psp2d_sar, variables = c("lnGr_Liv_Area"),
                        listw = lwames, viewplot = TRUE, smooth = TRUE)

```

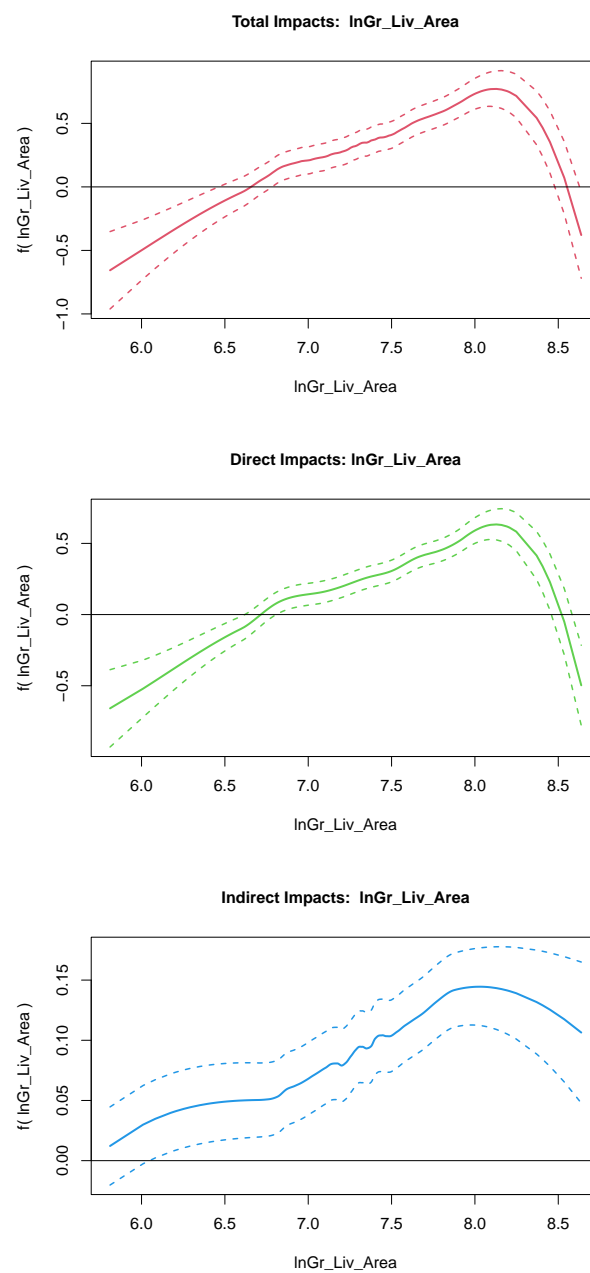


Figure 1. Impact functions of nonparametric covariate $\ln Gr_Liv_Area$.

Now, an example with the ANOVA decomposition of the spatial trend:

$$\mathbf{y} = f_1(\mathbf{s}_1) + f_2(\mathbf{s}_2) + f_{12}(\mathbf{s}_1, \mathbf{s}_2) + \rho \mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^k g_i(\mathbf{z}_i) + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_N),$$

In this case, we set the option `psanova = TRUE` within the term `pspt()` for the spatial trend (we omit the output of this model):

```
fpsp2dan <- update(fps, . ~ . + pspt(Longitude, Latitude,
                                   nknots = c(10, 10), psanova = TRUE))
psp2dan_sar <- pspatfit(fpsp2dan, data = ames_sf1, listw = lwames,
                       method = "Chebyshev", type = "sar")
summary(psp2dan_sar)
```

It is possible to compare the goodness-of-fit of different estimated models using `anova()` method for `pspatreg` class (`rlogLik` provides the value of Restricted Likelihood in the optimum). We set the argument `texttrlogLik=FALSE` because the models are not nested.

```
anova(linsar, ps_sar, psp2d_sar, psp2dan_sar, lrtest = FALSE)
```

```
##           logLik rlogLik   edf    AIC    BIC
## linsar      3379.9  3348.4  7.000 -6745.7 -6641.2
## ps_sar      3494.4  3468.0 30.714 -6927.4 -6692.8
## psp2d_sar   3571.7  3539.7 71.054 -7001.4 -6517.9
## psp2dan_sar 3573.3  3541.4 76.311 -6994.1 -6479.9
```

`psp2d_sar` model reaches the best AIC, but the best BIC corresponds to `ps_sar` model. This is due to the greater penalty for the increase in degrees of freedom by the BIC criterion. It also seems that the ANOVA decomposition of the trend has a worse performance.

In the previous model comparison, we set `lrtest = FALSE` because not all models were nested. It is also possible to perform a likelihood ratio test when the models are nested:

```
anova(ps_sar, psp2d_sar, lrtest = TRUE)
```

```
##           logLik rlogLik   edf    AIC    BIC LRtest    p.val
## ps_sar      3494.4  3468.0 30.714 -6927.4 -6692.8
## psp2d_sar   3571.7  3539.7 71.054 -7001.4 -6517.9  143.4 1.8124e-13
```

In this case, the LR test rejects the null, i.e., `ps_sar`, in favor of `psp2d_sar` model.

5.5. Examples of Plotting Spatial Trends for Spatial Point Coordinates

Now, we show how to plot spatial trends using spatial coordinates. Notice that the database is a `sf` object and excludes duplicated spatial points. We can only plot the 2d smooth trend of latitude and longitude for the model without the ANOVA decomposition. Next, we show the spatial trend for `psp2d_sar` (see Figure 2):

```
plot_sp2d(psp2d_sar, data = ames_sf1)
```

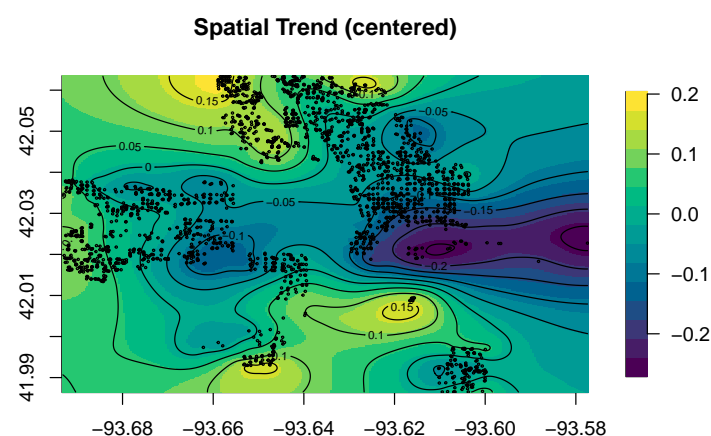


Figure 2. Spatial trend for `psp2d_sar` model.

For the model with the ANOVA decomposition, we can plot the whole 2d spatial trend (as previously) or its decomposition in main and interaction effects. We see the full spatial trend and its ANOVA decomposition for `psp2dan_sar` (see Figure 3):

```
plot_sp2d(psp2dan_sar, data = ames_sf1, addmain = TRUE, addint = TRUE)
```

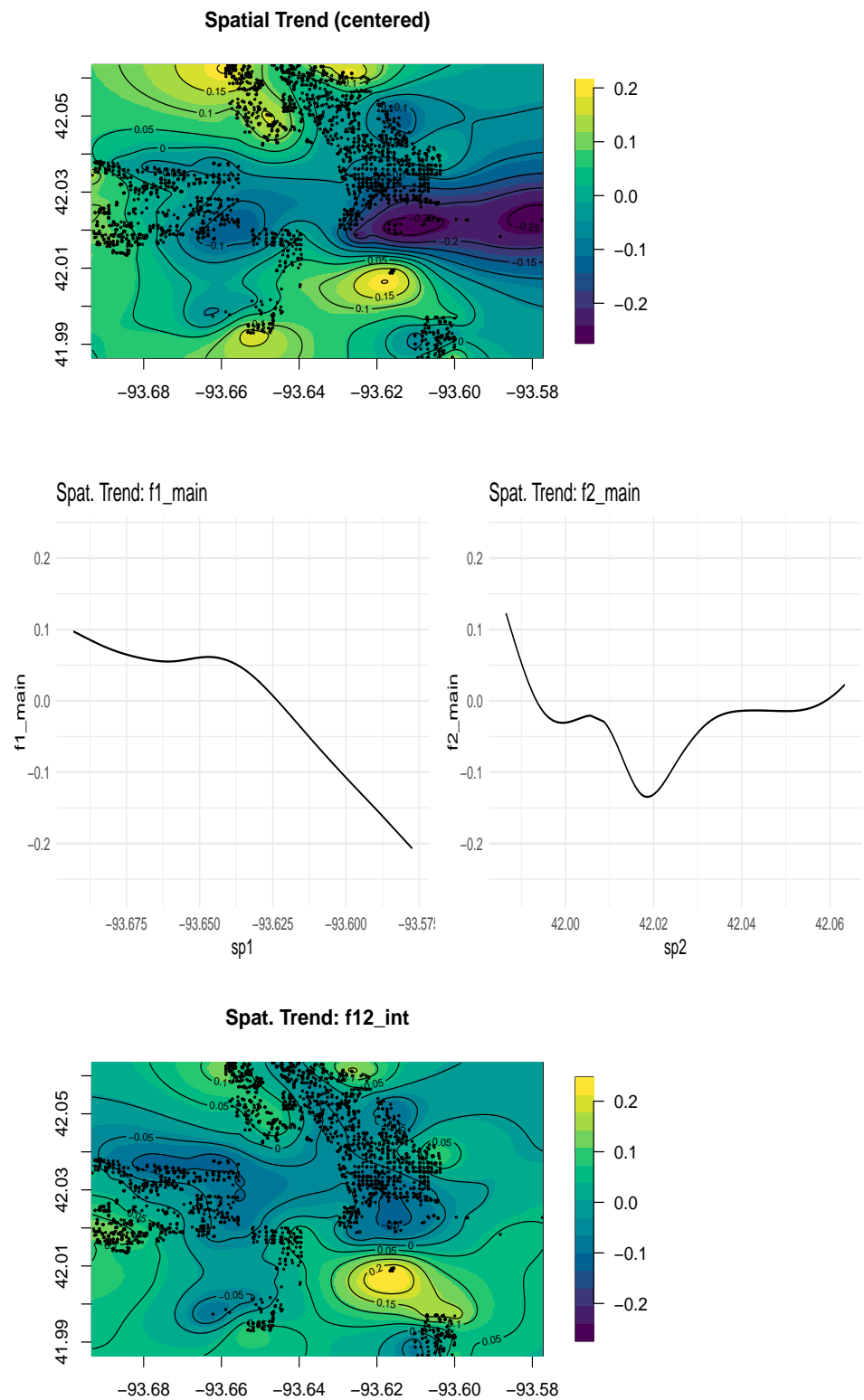


Figure 3. ANOVA decomposition of spatial trend for *psp2dan_sar* model.

6. Examples for Spatial Panel Data (3d)

This section focuses on the semiparametric P-Spline model for spatial panel data. The model may include a smooth spatiotemporal trend, a spatial lag of dependent and independent variables, a time lag of the dependent variable and its spatial lag, and a time-series autoregressive noise. Specifically, we consider a spatiotemporal ANOVA model that decomposes the trend into spatial and temporal main effects and second- and third-order interactions between them.

The empirical illustration is based on data on regional unemployment in Italy. This example shows that this model is a valid alternative to parametric methods aimed at disentangling strong and weak cross-sectional dependence when spatial and temporal heterogeneity is smoothly distributed (see [4]). In particular, while strong cross-sectional dependence is typically controlled within a parametric framework by including cross-sectional averages of dependent and independent variables in the model, in a semiparametric framework, this objective can instead be achieved by incorporating a spatiotemporal trend, as illustrated below. The section is organized as follows:

- Description of the dataset, spatial weights matrix, and model specifications.
- Estimation results of linear spatial models and comparison with those obtained with `splm`.
- Estimation results of semiparametric spatial models.

6.1. Dataset, Spatial Weights Matrix, and Model Specifications

The `pspatreg` provides the panel data `unemp_it` (a `data.frame` object) and the spatial weights matrix `Wsp_it` (a 103 by 103 square matrix). The raw data—a balanced panel with 103 Italian provinces observed for each year between 1996 and 2019—can be transformed into a spatial polygonal dataset of class `sf` after joining the `data.frame` object with the shapefile of Italian provinces (also included when loading `unemp_it`):

```
data(unemp_it, package = "pspatreg")
unemp_it_sf <- sf::st_as_sf(dplyr::left_join(unemp_it, map_it,
                                           by = c("prov" = "COD_PRO")))
```

The matrix `Wsp_it` is a standardized inverse distance **W** matrix. First, we transform the matrix into a list of neighbor objects:

```
lwsp_it <- spdep::mat2listw(Wsp_it)
```

6.2. Linear Model: Comparison with `Splm`

We begin by estimating fully parametric spatial linear autoregressive panel models using the function `pspatfit()` from the `pspatreg` package and compare these results with those obtained from the `splm` package, commonly used for parametric spatial panels.

In this analysis, we consider a linear SAR model with fixed spatial and temporal effects. The following equation presents the model in scalar form (a similar approach can be applied to SDM, SEM, SARAR, and SDEM specifications):

$$y_{it} = \rho \sum_{j=1}^N w_{ij} y_{jt} + \sum_{k=1}^K \beta_k x_{k,it} + \alpha_i + \theta_t + \epsilon_{it} \quad \epsilon_{it} \sim i.i.d.(0, \sigma_\epsilon^2)$$

As usual, the `pspatfit()` function in `pspatreg` can be used to estimate the model. The `index` argument specifies the spatial and temporal indices, `demean` adjusts the data for fixed effects, and setting `eff_demean = "twoways"` enables both spatial and temporal fixed effects.

```

flin2 <- unrate ~ empgrowth + partrate + agri + cons + serv
plinsar <- pspatfit(formula = flin2, data = unemp_it, listw = lwsp_it,
                    demean = TRUE, eff_demean = "twoways", type = "sar",
                    index = c("prov", "year"))
summary(plinsar)

##
## Call
## pspatfit(formula = flin2, data = unemp_it, listw = lwsp_it, type = "sar",
##          demean = TRUE, eff_demean = "twoways", index = c("prov",
##                  "year"))
##
## Parametric Terms
##           Estimate Std. Error t value Pr(>|t|)
## empgrowth -0.129739  0.014091 -9.2074 < 2.2e-16 ***
## partrate   0.393087  0.023315 16.8595 < 2.2e-16 ***
## agri       -0.036052  0.027267 -1.3222 0.1862167
## cons      -0.166196  0.044510 -3.7339 0.0001928 ***
## serv       0.012378  0.020597  0.6009 0.5479300
## rho        0.265671  0.018858 14.0880 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Goodness-of-Fit
##
## EDF Total:      6
## Sigma: 1.86929
## AIC: 5396.53
## BIC: 5431.41

```

Now, we compare the results obtained using the splm package.

```

plinsarsplm <- splm::splm(flin2, data = unemp_it, index = c("prov", "year"),
                        listw = lwsp_it, model="within", effect = "twoways",
                        spatial.error = "none", lag = TRUE)
round(data.frame(pspatreg = c(plinsar$rho, plinsar$bfixed),
                  splm = c(plinsarsplm$coefficients1,
                          plinsarsplm$coefficients[-1])),3)

```

```

##           pspatreg  splm
## rho           0.266  0.266
## fixed_empgrowth -0.130 -0.130
## fixed_partrate   0.393  0.392
## fixed_agri      -0.036 -0.037
## fixed_cons     -0.166 -0.167
## fixed_serv       0.012  0.012

```

The two packages yield identical results, at least at the third-digit level. Similar outcomes are observed for other spatial econometric models (SEM, SDM, and SARAR). Further details are provided in the online appendix.

6.3. Semiparametric Spatial Panel Models

We specify an additive semiparametric model with three parametric linear terms (for partrate, agri, and cons) and two nonparametric smooth terms (serv and empgrowth).

```
fps <- unrata ~ partrate + agri + cons + pspl(serv, nknots = 15) +
  pspl(empgrowth, nknots = 20)
```

In this model, we control for spatiotemporal heterogeneity by including a PS-ANOVA spatiotemporal trend. The interaction terms in the ANOVA decomposition (f12, f1t, f2t, and f12t) are incorporated with a nested basis to simplify computations. It is worth noticing that the arguments nest_sp1, nest_sp2, and nest_time) must be divisors of nknots. We estimate this model for a SAR specification using the pspatfit() function:

```
fpsp3dan <- update(fps, . ~ . + pspt(long, lat, year, nknots = c(18,18,8),
  psanova = TRUE, nest_sp1 = c(1, 2, 3),
  nest_sp2 = c(1, 2, 3), nest_time = c(1, 2, 2)))
ps3dan_sar <- pspatfit(fpsp3dan, data = unemp_it, listw = lwsp_it,
  method = "Chebyshev", type = "sar")
summary(ps3dan_sar)
```

```
##
## Call
## pspatfit(formula = fpsp3dan, data = unemp_it, listw = lwsp_it,
##   type = "sar", method = "Chebyshev")
##
## Parametric Terms
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.9757174  8.4450367  0.4708  0.6378
## partrate    0.1495539  0.0219457  6.8147 1.206e-11 ***
## agri        -0.0205019  0.0197572 -1.0377  0.2995
## cons        -0.0387163  0.0404494 -0.9572  0.3386
## f1_main.1   -1.8380455 13.2621302 -0.1386  0.8898
## f2_main.1  -15.5414740 10.9648616 -1.4174  0.1565
## ft_main.1    2.4143619  4.8934373  0.4934  0.6218
## f12_int.1   -8.7938027 12.3559132 -0.7117  0.4767
## f1t_int.1    4.1090210  6.4688949  0.6352  0.5254
## f2t_int.1   -2.1236074  6.7774834 -0.3133  0.7541
## f12t_int.1   1.4150889  7.6549475  0.1849  0.8534
## pspl(serv)   0.0915770  0.1681959  0.5445  0.5862
## pspl(empgrowth) -0.2880005  0.0655609 -4.3929 1.170e-05 ***
## rho          0.0064252  0.0193087  0.3328  0.7393
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Non-Parametric Terms
##           EDF
## pspl(serv)  5.6796
## pspl(empgrowth) 2.6253
##
## Non-Parametric Spatio-Temporal Trend
##           EDF
## f1    10.667
## f2     9.733
## ft     7.782
## f12   45.421
## f1t    4.239
## f2t   21.480
## f12t  82.057
```

```
##
## Goodness-of-Fit
##
## EDF Total: 203.683
## Sigma: 1.54113
## AIC: 5977.69
## BIC: 7161.65
```

The included trend effectively captures the spatiotemporal heterogeneity, resulting in a significantly smaller estimate of the spatial parameter (0.0064) than the previous estimate from the model without the trend (0.266).

In some applications, it might be convenient to exclude some terms in the ANOVA decomposition of the spatiotemporal trend to speed up computations. For example, to specify a model excluding the second-order interaction between spatial and temporal coordinates, we have to write the following formula:

```
fpsp3danr <- update(fps, . ~ . + pspt(long, lat, year, nknots = c(18,18,8),
  psanova = TRUE, nest_sp1 = c(1, 2, 3),
  nest_sp2 = c(1, 2, 3), nest_time = c(1, 2, 2),
  f1t = FALSE, f2t = FALSE))
```

In the context of spatial panel models, it is also possible to account for temporal autocorrelation by incorporating an AR(1) autoregressive error term. To do so, we need to set the argument `cor = "ar1"` within the `pspatfit()` function:

```
ps3dan_sarar1 <- pspatfit(fpsp3dan, data = unemp_it, listw = lwsp_it,
  method = "Chebyshev", type = "sar", cor = "ar1")
```

```
## Error in .solve.checkCond(a, tol) :
## 'a' is computationally singular, rcond(a)=3.71105e-17
```

```
summary(ps3dan_sarar1)
```

```
##
## Call
## pspatfit(formula = fpsp3dan, data = unemp_it, listw = lwsp_it,
##   type = "sar", method = "Chebyshev", cor = "ar1")
##
## Parametric Terms
##
```

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	3.9738e+00	8.2602e-05	4.8107e+04	< 2.2e-16 ***
## partrate	1.5220e-01	1.5408e-03	9.8783e+01	< 2.2e-16 ***
## agri	-2.9875e-02	6.5566e-03	-4.5565e+00	5.472e-06 ***
## cons	-3.2782e-02	1.1767e-03	-2.7858e+01	< 2.2e-16 ***
## f1_main.1	-1.2531e-01	5.8191e-04	-2.1535e+02	< 2.2e-16 ***
## f2_main.1	-1.2409e+01	7.1853e-04	-1.7271e+04	< 2.2e-16 ***
## ft_main.1	2.0255e+00	4.2745e-04	4.7385e+03	< 2.2e-16 ***
## f12_int.1	-6.1988e+00	3.7017e-04	-1.6746e+04	< 2.2e-16 ***
## f1t_int.1	1.7009e+00	5.9376e-06	2.8645e+05	< 2.2e-16 ***
## f2t_int.1	-3.8990e+00	2.0340e-05	-1.9169e+05	< 2.2e-16 ***
## f12t_int.1	2.5620e+00	2.4928e-04	1.0278e+04	< 2.2e-16 ***
## pspl(serv)	1.3822e-01	2.7308e-04	5.0615e+02	< 2.2e-16 ***
## pspl(empgrowth)	-3.1413e-01	7.0032e-05	-4.4855e+03	< 2.2e-16 ***
## rho	5.0744e-02	2.1664e-02	2.3423e+00	0.01925 *
## phi	3.2957e-01	1.3700e-02	2.4056e+01	< 2.2e-16 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Non-Parametric Terms
##                EDF
## pspl(serv)      4.3076
## pspl(empgrowth) 2.4433
##
## Non-Parametric Spatio-Temporal Trend
##                EDF
## f1      10.190
## f2       9.808
## ft       7.900
## f12     42.240
## f1t       3.605
## f2t     22.766
## f12t    45.466
##
## Goodness-of-Fit
##
## EDF Total: 162.726
## Sigma: 1.61722
## AIC: 5446.93
## BIC: 6392.82
```

As expected, the temporal correlation parameter (0.31) is positive and significant. Incorporating the AR(1) term into the model has resulted in a substantial reduction in the EDFs of the spatiotemporal trend, particularly for the `f12t` interaction term. Notably, in this specification, the spatial parameter (0.044) has increased and is now nearly significant. When comparing the goodness-of-fit of the two models, it appears that the model incorporating temporal correlation performs better in terms of both AIC and BIC:

```
anova(ps3dan_sar, ps3dan_sarar1, lrtest = FALSE)
```

```
##                logLik rlogLik   edf   AIC   BIC
## ps3dan_sar      -2785.2 -2785.9 203.68 5977.7 7145.6
## ps3dan_sarar1 -2560.7 -2565.1 162.73 5446.9 6390.4
```

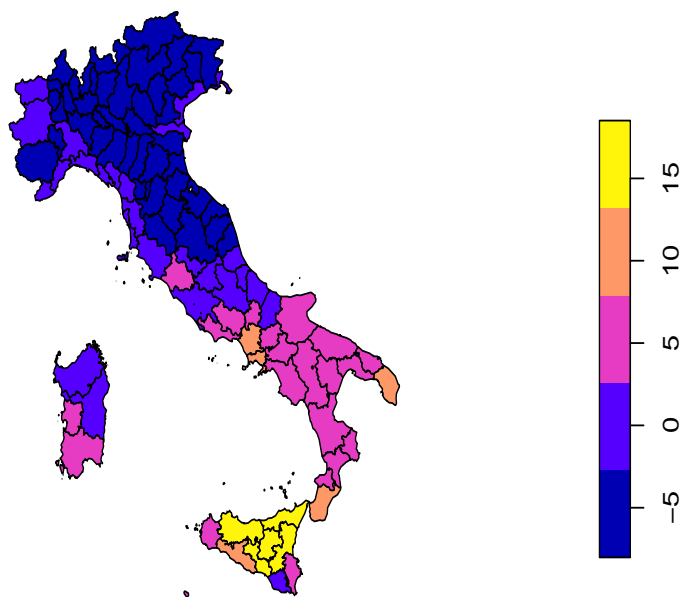
We can compute marginal impacts for either parametric or nonparametric covariates in the same way as in the 2d spatial case using `impactspar()` or `impactsnopar()` functions. See the online appendix for more details.

6.4. Plot of Spatial and Temporal Trends (3d)

Once we have estimated the model, we can plot spatial trend maps for different years. As an example, we plot the spatial trend for the first and last years (1996 and 2019) using the `plot_sp3d()` function for the `ps3dan_sarar1` object (see Figure 4):

```
plot_sp3d(ps3dan_sarar1, data = unemp_it_sf, time_var = "year",
          time_index = c(1996, 2019), addmain = FALSE, addint = FALSE)
```


Spatial Trend (centered) for : 1996



Spatial Trend (centered) for : 2019

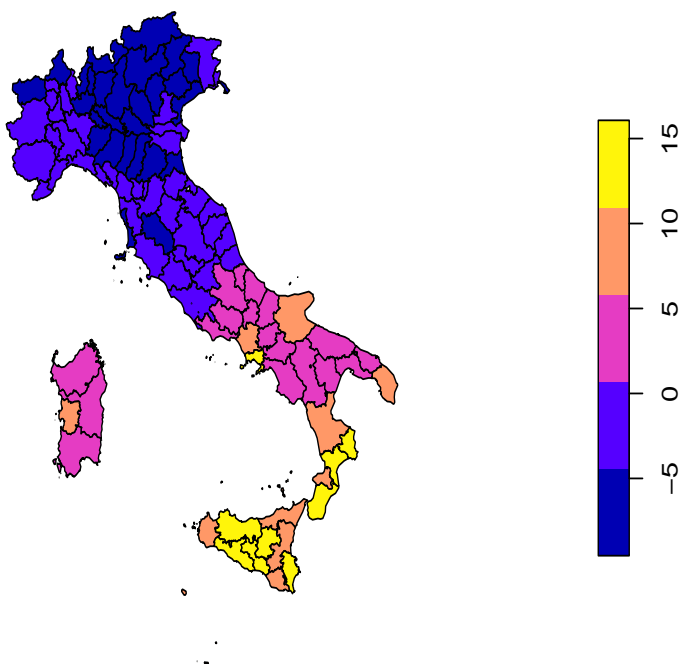


Figure 4. Spatial trends for *ps3dan_sarar1* model in 1996 and 2019.

With the ANOVA decomposition, it is also possible to plot the spatial maps of the main effects (latitude and longitude) and second-order spatial effects setting arguments `addmain` and `addint` to `TRUE`, i.e., (we do not include the output for the sake of brevity):

```
plot_sp3d(ps3dan_sarar1, data = unemp_it_sf, time_var = "year",
          time_index = c(1996, 2019), addmain = TRUE, addint = TRUE)
```

On the other hand, it is also possible to plot distinct temporal trends for each province using the function `plot_sptime()` (see Figure 5):

```
plot_sptime(ps3dan_sarar1, data = unemp_it, time_var = "year", reg_var = "prov")
```

Temporal Trend for each region

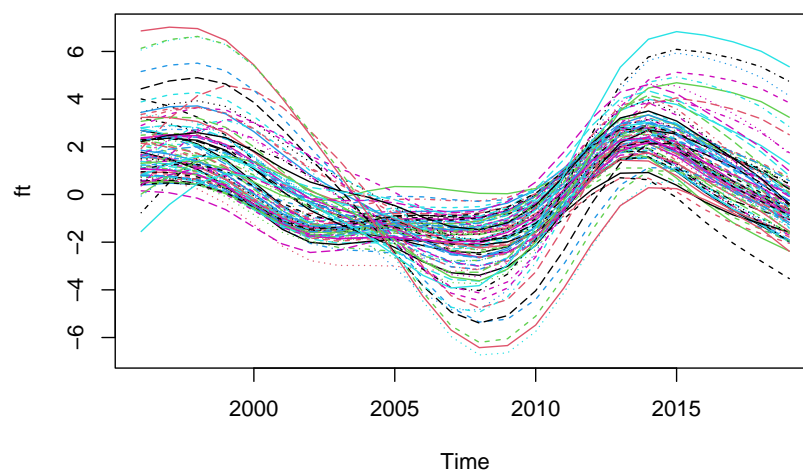


Figure 5. Temporal trend for each region for *ps3dan_sarar1* model.

The figure suggests that, although a common cycle exists in the temporal pattern, the time trend varies heterogeneously across different provinces.

7. Summary

This paper presents *pspatreg*, a package for estimating and inferencing semiparametric spatial econometric models using P-splines. The package includes functions for various spatial econometric specifications (SLX, SAR, SEM, SDM, SDEM, and SARAR), which incorporate parametric and nonparametric (smooth) terms, as well as spatial or spatiotemporal trends. For illustrative purposes, the paper provides empirical cases that encompass both spatial data (2d) and spatial panel data (3d).

In addition to the estimation functions and common methodologies associated with the *pspatreg* class, the examples include calculations of both parametric and nonparametric spatial effects, along with presentations of spatial and temporal trends. Furthermore, for purely parametric spatial econometric models, the empirical cases compare with numerical estimates obtained from the *pspatreg* package and with the well-known spatial econometric packages, namely *spatialreg* and *splm*.

Author Contributions: Conceptualization, R.B. and M.D.; Methodology, R.M., R.B. and M.D.; Software, R.M.; Validation, R.B. and M.D.; Formal analysis, R.B. and M.D.; Writing—original draft, R.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors gratefully acknowledge the financial support provided for this research. The first author extends sincere thanks for the funding received from the grants PID2022-136252NB-I00, funded by MCIN/AEI and the European Union-Next Generation, and 2022-GRIN-34336, funded by the European Union-Next Generation and UCLM. The second author expresses gratitude for the support from the European Union—Next Generation EU through the PRIN 2022 project titled “Reloading city: a new systemic approach to urban and territorial regeneration” (CUP: E53D23010120006). The third author acknowledges the grant PID2022-137243OB-I00, funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Montero, J.M.; Mínguez, R.; Durbán, M. SAR models with nonparametric spatial trends. A P-spline approach. *Estad. Esp.* **2012**, *54*, 89–111.
2. Basile, R.; Durbán, M.; Mínguez, R.; Montero, J.M.; Mur, J. Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities. *J. Econ. Dyn. Control.* **2014**, *48*, 229–245. [[CrossRef](#)]
3. Hoshino, T. Semiparametric Spatial Autoregressive Models with Endogenous Regressors: With an Application to Crime Data. *J. Bus. Econ. Stat.* **2018**, *36*, 160–172. [[CrossRef](#)]
4. Mínguez, R.; Basile, R.; Durbán, M. An alternative semiparametric model for spatial panel data. *Stat. Methods Appl.* **2020**, *29*, 669–708. [[CrossRef](#)]
5. Eilers, P.H.; Marx, B.D.; Durbán, M. Twenty years of P-splines. *SORT-Stat. Oper. Res. Trans.* **2015**, *39*, 149–186.
6. LeSage, J.P.; Pace, R.K. *Introduction to Spatial Econometrics*; CRC Press: Boca Raton, FL, USA, 2009. [[CrossRef](#)]
7. Elhorst, J.P. *Spatial Econometrics. From Cross-Sectional Data to Spatial Panels*; SpringerBriefs in Regional Science; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2014. [[CrossRef](#)]
8. Lee, D.; Durbán, M. P-Spline ANOVA Type Interaction Models for Spatio-Temporal Smoothing. *Stat. Model.* **2011**, *11*, 49–69. [[CrossRef](#)]
9. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer Academic Pub.: London, UK, 1988; Volume 4.
10. McMillen, D.P. Perspectives on spatial econometrics: Linear smoothing with structured models. *J. Reg. Sci.* **2012**, *52*, 192–209. [[CrossRef](#)]
11. Brunsdon, C.; Fotheringham, A.S.; Charlton, M.E. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geogr. Anal.* **1996**, *28*, 281–298. [[CrossRef](#)]
12. Cho, S.H.; Lambert, D.M.; Roberts, R.K.; Kim, S.G. Moderating urban sprawl: Is there a balance between shared open space and housing parcel size? *J. Econ. Geogr.* **2010**, *10*, 763–783. [[CrossRef](#)]
13. Páez, A.; Uchida, T.; Miyamoto, K. A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environ. Plan. A* **2002**, *34*, 733–754. [[CrossRef](#)]
14. Pace, R.K.; LeSage, J.P. Spatial autoregressive local estimation. In *Spatial Econometrics and Spatial Statistics*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 31–51.
15. Geniaux, G.; Martinetti, D. A new method for dealing simultaneously with spatial autocorrelation and spatial heterogeneity in regression models. *Reg. Sci. Urban Econ.* **2017**, *Forthcoming*. [[CrossRef](#)]
16. Chudik, A.; Pesaran, M.H.; Tosetti, E. Weak and strong cross-section dependence and estimation of large panels. *Econom. J.* **2011**, *14*, C45–C90. [[CrossRef](#)]
17. Eilers, P.H.C.; Marx, B.D. Flexible Smoothing with B-Splines and Penalties. *Stat. Sci.* **1996**, *11*, 89–121. [[CrossRef](#)]
18. De Boor, C. Package for calculating with B-splines. *J. Numer. Anal.* **1977**, *14*, 441–472. [[CrossRef](#)]
19. Ruppert, D. Selecting the number of knots for penalized splines. *J. Comput. Graph. Stat.* **2002**, *11*, 735–757. [[CrossRef](#)]
20. Lee, D. Smoothing Mixed Models for Spatial and Spatio-Temporal Data. Ph.D. Thesis, University Carlos-III, Getafe, Spain, 2010.
21. Eilers, P.H.; Marx, B.D. *Practical Smoothing: The Joys of P-Splines*; Cambridge University Press: Cambridge, UK, 2021. [[CrossRef](#)]
22. Wood, S. *Generalized Additive Models: An Introduction with R*, 2nd ed.; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017. [[CrossRef](#)]
23. Rodríguez-Álvarez, M.X.; Lee, D.J.; Kneib, T.; Durbán, M.; Eilers, P.H.C. Fast smoothing parameter separation in multidimensional generalized P-splines: The SAP algorithm. *Stat. Comput.* **2015**, *25*, 941–957. [[CrossRef](#)]
24. Bates, D.; Mullen, K.M.; Nash, J.C.; Varadhan, R. *Minqa: Derivative-Free Optimization Algorithms by Quadratic Approximation*; R Package Version 1.2.5; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: https://www.google.com.hk/url?sa=t&source=web&rct=j&opi=89978449&url=https://rdrr.io/cran/minqa/&ved=2ahUKewitx5zS-NOJAxXTrIYBHTnvMk8QFnoECBQQAQ&usq=AOvVaw3G00dnjwc1_isOQMAWhjNu (accessed on 1 September 2024).
25. Fahrmeir, L.; Kneib, T.; Lang, S.; Marx, B.D. *Regression*; Springer: Berlin/Heidelberg, Germany, 2022. [[CrossRef](#)]
26. Lee, D.J.; Durbán, M.; Eilers, P.H.C. Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. *Comput. Stat. Data Anal.* **2013**, *61*, 22–37. [[CrossRef](#)]
27. Blundell, R.; Powell, J. Advances in Economics and Econometrics Theory and Application. In *Chapter Endogeneity in Nonparametric and Semiparametric Regression Models*; Cambridge University Press: Cambridge, UK, 2003.
28. Bivand, R.; Millo, G.; Piras, G. A Review of Software for Spatial Econometrics in R. *Mathematics* **2021**, *9*, 1276. [[CrossRef](#)]
29. Bivand, R.S.; Pebesma, E.; Gomez-Rubio, V. *Applied Spatial Data Analysis with R*, 2nd ed.; Springer: New York, NY, USA, 2013.
30. Bivand, R. R Packages for Analyzing Spatial Data: A Comparative Case Study with Areal Data. *Geogr. Anal.* **2022**, *54*, 488–518. [[CrossRef](#)]
31. Piras, G. sphet: Spatial Models with Heteroskedastic Innovations in R. *J. Stat. Softw.* **2010**, *35*, 1–21. [[CrossRef](#)]
32. Juhl, S. spfilter: An R Package for Semiparametric Spatial Filtering with Eigenvectors in (Generalized) Linear Models. *R J.* **2021**, *13*, 450–459. [[CrossRef](#)]
33. Bivand, R.; Yu, D. *spgwr: Geographically Weighted Regression*; R Package Version 0.6-35; R Foundation for Statistical Computing: Vienna, Austria, 2022.

34. Mínguez, R.; López, F.A.; Mur, J. *spsur*: An R Package for Dealing with Spatial Seemingly Unrelated Regression Models. *J. Stat. Softw.* **2022**, *104*, 1–43. [[CrossRef](#)]
35. Millo, G.; Piras, G. *splm*: Spatial Panel Data Models in R. *J. Stat. Softw.* **2012**, *47*, 1–38. [[CrossRef](#)]
36. Simonovska, R. *Spatial Dynamic Panel Data Modeling*; R Package Version 0.0.1; R Foundation for Statistical Computing: Vienna, Austria, 2022.
37. Rodríguez-Álvarez, M.X.; Boer, M.P.; van Eeuwijk, F.A.; Eilers, P.H. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spat. Stat.* **2017**, *23*, 52–71. [[CrossRef](#)]
38. Eilers, P.; Marx, B.; Li, B.; Gampe, J.; Rodríguez-Alvarez, M.X. *JOPS: Practical Smoothing with P-Splines*; R Package Version 0.1.15; R Foundation for Statistical Computing: Vienna, Austria, 2021.
39. Drukker, D.M.; Prucha, I.R.; Raciborski, R. A Command for Estimating Spatial-Autoregressive Models with Spatial-Autoregressive Disturbances and Additional Endogenous Variables. *Stata J.* **2013**, *13*, 287–301. [[CrossRef](#)]
40. Drukker, D.M.; Prucha, I.R.; Raciborski, R. Maximum Likelihood and Generalized Spatial Two-Stage Least-Squares Estimators for a Spatial-Autoregressive Model with Spatial-Autoregressive Disturbances. *Stata J.* **2013**, *13*, 221–241. [[CrossRef](#)]
41. Belotti, F.; Hughes, G.; Mortari, A.P. Spatial panel-data models using Stata. *Stata J.* **2017**, *17*, 139–180. [[CrossRef](#)]
42. LeSage, J.P. *Spatial Econometrics*; Department of Economics, University of Toledo: Toledo, OH, USA, 1998.
43. Elhorst, J.P. Matlab Software for Spatial Panels. *Int. Reg. Sci. Rev.* **2014**, *37*, 389–405. [[CrossRef](#)]
44. LeSage, J.P. *A Panel Data Toolbox for Matlab*; Department of Economics, University of Toledo: Toledo, OH, USA, 2021.
45. Kuhn, M. *AmesHousing: The Ames Iowa Housing Data*; R Package Version 0.0.4; R Foundation for Statistical Computing: Vienna, Austria, 2020.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.