Spatio-Temporal Autoregressive Semiparametric Model for the analysis of regional economic data

Román Mínguez (University of Castilla-La Mancha)
María L. Durbán (Carlos III University)
Roberto Basile (Seconda Università di Napoli)

July 25, 2016

Preliminary Draft

Abstract

In this paper we propose an extension of the semiparametric P-Spline model to spatio-temporal data including a non-parametric trend, as well as a spatial lag of the dependent variable. This model is able to simultaneously control for *functional form bias*, *spatial dependence bias*, *spatial heterogeneity bias*, and *omitted time-related factors bias*. Specifically, we consider a spatio-temporal ANOVA model disaggregating the trend in spatial and temporal main effects, and second and third order interactions between them. The model can include both linear and non-linear effects of the covariates, and other additional fixed or random effects. Recent algorithms based on spatial anisotropic penalties (SAP) are used to estimate all the parameters in a closed form without the need of multidimensional optimization. An empirical case compares the performance of this model against alternatives models like spatial panel data models.

JEL classification: C33, C14, C63.
Keywords: spatio-temporal trend, mixed models, P-splines, PS-ANOVA, SAR, spatial panel.

1 Introduction

A recent strand of the spatial econometric literature has proposed *Spatial Autoregressive Semiparametric Geoadditive Models* to simultaneously deal with different critical issues typically met when using spatial economic data, that is spatial dependence, spatial heterogeneity and unknown functional form (Montero et al., 2012; Basile et al., 2014). This approach combines penalized regression spline (PS) methods (Eilers et al., 2015) with standard cross-section spatial autoregressive models (such as SAR, SEM, SDM and SLX). An important feature of these models is the possibility 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, that is 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 an extension of the P-Spline spatial auto-regressive model (PS-SAR) to spatio-temporal data when both a large cross-section and a large time series dimensions are available. With this kind of data it is possible to estimate not only spatial trends, but also spatio-temporal trends in a nonparametric way (Lee and Durbán, 2011), so as to capture region-specific nonlinear time trends net of the effect of spatial autocorrelation. In other words, this approach allows to answer questions like: How do unobserved time-related factors (i.e. common factors), such as economic-wide technological or demand shocks, heterogeneously affect long term dynamics of all units in the sample? And how does their inclusion in the model affect the estimation of spatial interaction effects? In this sense, the PS-SAR model with spatio-temporal trend represents an alternative to parametric methods aimed at disentangling *common factors* effects (such as common business cycle effects) and *spatial dependence* effects (local interactions between regions generating spillover effects), where the former is sometimes regarded as 'strong' cross-sectional dependence and the latter as 'weak' cross-sectional dependence (Chudik et al., 2011).

Recently, Bai and Li (2015) and Shi and Lee (2016) have proposed 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 (ε) and not on the observable dependent variable ($\mathbf{y}_n$). Bailey et al. (2016) and Vega and Elhorst (2016) have also proposed a two-step and one-step approach, respectively, to address both forms of cross-sectional dependence, but without including explanatory variables in the model. All these approaches are still parametric and do not properly allow for capturing nonlinearities. On the other hand, Su and Jin (2012) have considered the problem of estimating semiparametric panel data models with cross section 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 effects (CCE) estimator to a semiparametric framework. Nevertheless, they do not take spatial contagion effects (i.e. weak cross-section dependence) into account. By relying on the PS-SAR model with spatio-temporal trends, we handle simultaneously 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 seem complicated and computationally demanding. Nevertheless, we consider a decomposition of the spatio-temporal trend into several components (spatial and temporal main effects, and second and third order interactions between them) that gives further insights into the dynamics of the data. Furthermore, we use a mixed model representation that allows us to use the methods already developed in this area for estimation and inference, and to implementation of 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 case of PS-SAR model that dramatically reduces the computational time needed to estimate the parameters in the model. Also, the use of B-spline nested basis (Lee et al., 2013) for the interaction components contributes to the efficiency of the fitting procedure without compromising the goodness of fit of the model.

We apply the PS-SAR model with a spatio-temporal trend to real data on regional unemployment rates in Italy.[1] As well known, these data are characterized by spatial dependence, unobserved spatial heterogeneity and unobserved common effects. Substantive spatial dependence (weak dependence) occurs due to interregional trade, labor migration and commuting, and knowledge spillovers; it can be captured by including spatial interaction effects in the model (Burridge and Gordon, 1981; Molho, 1995; Henry G. Overman, 2002; Patacchini and Zenou, 2007). Unobserved spatial heterogeneity is mainly due to a strong North-South spatial trend which can be hardly captured by the explanatory variables: regional unemployment rates largely increase moving from the North to the South, reflecting the well-known regional development divide within the Country. A time-invariant smooth spatial trend might be used to filtering out unobserved heterogeneity; thus, the spatial trend assumes the same role as the fixed regional effects. Several unobserved common factors (e.g. aggregate demand shocks, aggregate technological shocks, and global labor market policies) may also heterogeneously affect the level of regional unemployment. The econometric results suggest that the PS-SAR model with a spatio-temporal trend outperforms several parametric and nonparametric competing 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 to the spatial lag term $\mathbf{W}_n\mathbf{y}$ reveals the existence of significant spatial spillovers (weak dependence) net of the effect of the observed and unobserved common factors.

The plan of the paper is as follows. Section 2 sets out the PS-SAR model with a spatio-temporal trend and discusses various technical aspects related to its identification and estimation. Section 3 reports the results of Monte Carlo experiments, while Section 4 discusses the results of the application of the model to regional unemployment data. Section 5 concludes by identifying important areas for extensions and further developments.

[1] We implemented new functions in the R software to estimate PS-SAR models with spatio-temporal trends.

2 Spatio-Temporal Autoregressive Models

2.1 P-splines Mixed Models

Semiparametric models are a flexible tool to incorporate non-linear functional relationships into a regression model. The general form of a semiparametric model is given by:

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\beta}^* + f_1(\mathbf{x}_1) + f_2(\mathbf{x}_2) + f_{3,4}(\mathbf{x}_3, \mathbf{x}_4) + \ldots + \boldsymbol{\epsilon} \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2) \tag{1}$$

where $\mathbf{y}$ is a continuous response variable, $\mathbf{X}^*\boldsymbol{\beta}^*$ is the linear predictor (containing the intercept and categorical and continuous covariates whose functional relationship with the response is linear), and $f_k()$ are unknown smooth functions of single covariates or interaction surfaces. Several approaches have appeared in the literature to fit such models (Green and Yandell, 1985; Hastie and Tibshirani, 1990). We use a penalized regression approach which combines a basis representation of the functions with a penalty to control the wiggliness of the curve/surface. In particular, we use the approach introduced by Eilers and Marx (1996) where each univariate smooth term is represented by:

$$f_k(\mathbf{x}_k) = \sum_{j=1}^{c_k} B_j(\mathbf{x}_k)\theta_j, \; j = 1, ..., c_k$$

with B_j a B-spline basis function and θ_j a vector of regression coefficients of length c_k. The smooth interaction terms are

$$f_{i,k}(\mathbf{x}_i, \mathbf{x}_k) = \sum_{j=1}^{c_i}\sum_{l=1}^{c_k} B_j(\mathbf{x}_i)B_l(\mathbf{x}_k)\theta_{jl}, \text{ with } j = 1, ..., c_i \text{ and } l = 1, ..., c_k,$$

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 $c_i c_k \times 1$. In matrix notation, model (1) becomes:

$$\mathbf{E}[\mathbf{y}|\mathbf{X}^*, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \ldots] = \mathbf{B}\boldsymbol{\theta}, \tag{2}$$

where $\mathbf{B}$ is the full regression matrix, and $\boldsymbol{\theta} = (\boldsymbol{\beta}^*, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_{[3,4]}, \ldots)'$ is a vector of regression coefficients. The model matrix is defined by blocks:

$$\mathbf{B} = [\mathbf{X}^*|\mathbf{B}_1|\mathbf{B}_2|\mathbf{B}_{[3,4]}| \ldots], \tag{3}$$

with marginal bases of the covariates $\mathbf{B}_1 = \mathbf{B}_1(\mathbf{x}_1)$ and $\mathbf{B}_2 = \mathbf{B}_2(\mathbf{x}_2)$ and interaction basis $\mathbf{B}_{[3,4]}$ as the tensor product of the two marginals, i.e.

$$\mathbf{B}_{[3,4]} = \mathbf{B}_3 \Box \mathbf{B}_4 = (\mathbf{B}_3 \otimes \mathbf{1}'_n) * (\mathbf{1}'_n \otimes \mathbf{B}_4), \text{ of dimension } n \times c_3 c_4. \tag{4}$$

where the symbols $\Box$, $\otimes$, and $*$ indicate the box product, the Kronecker product, and the element-by-element product, respectively. The size, c_k, of each individual basis should be large enough (in general, between 4 and 40) to identify nonlinearities, and the smoothness of each term is controlled by a quadratic penalty term, yielding the following penalized regression problem:

$$\|(\mathbf{y} - \mathbf{B}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{B}\boldsymbol{\theta})\|^2 + \boldsymbol{\theta}'\mathbf{P}\boldsymbol{\theta}.$$

Typically, the quadratic penalty 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. The penalty matrix (1) is therefore:

$$\mathbf{P} = \text{blockdiag}(\mathbf{P}_i) \tag{5}$$

with $\mathbf{P}_i = \lambda_i \mathbf{D}_i' \mathbf{D}_i$ or $\mathbf{P}_i = \lambda_{ij} \mathbf{D}_j' \mathbf{D}_j \otimes \mathbf{I} + \lambda_{ik} \mathbf{I} \otimes \mathbf{D}_k' \mathbf{D}_k$ in the case of interaction effects. This last penalty allows for a separate amount of smoothing per covariate (*anisotropy*).

Since the intercept of the model is contained in $\mathbf{X}^*\boldsymbol{\beta}^*$ and a column of 1's is also spanned by each basis $\mathbf{B}_i$, there is a problem of identifiability (common to any additive model). There are several ways to solve it, but we choose to re-parametrize model (1) as a mixed model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \quad \boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{G}) \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \tag{6}$$

by transforming the bases and the penalty. Several transformations are possible, but the most popular one is based on the singular value decomposition of the penalty matrix (see Lee, 2010b, for details). Matrix $\mathbf{X}$ will include parametric components such as the intercept, continuous covariates and categorical covariates, while $\mathbf{Z}$ includes all the nonlinear components of the smooth effects. The covariance matrix of random effects, $\mathbf{G}$, is a diagonal matrix which depends on the eigenvalues of the singular value decomposition and variance components τ_i^2, and the smoothing parameters become $\lambda_i = \sigma^2/\tau_i^2$.

The estimates of the coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ follow from the standard mixed model theory (see Searle et al., 1992):

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{7}$$

$$\widehat{\boldsymbol{\alpha}} = \mathbf{G}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}), \tag{8}$$

where $\mathbf{V} = \sigma^2\mathbf{I} + \mathbf{Z}\mathbf{G}\mathbf{Z}'$.

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

$$\ell(\tau_i^2, \sigma^2) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{y}'(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y}. \tag{9}$$

The estimated values of the observed variable are obtained as:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\boldsymbol{\alpha}} = \mathbf{H}\mathbf{y}$$

where $\mathbf{H}$ is the hat matrix of the model given by:

$$\mathbf{H} = [\mathbf{X} : \mathbf{Z}] \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} [\mathbf{X} : \mathbf{Z}]'. \tag{10}$$

The trace of this matrix is defined as the *effective dimension*, which is a measure of the complexity of the model. Also, confidence bands for the estimated values can be calculated using an approximation of the variance–covariance matrix of the estimation error given by $V(\mathbf{y} - \hat{\mathbf{y}}) = \sigma_\varepsilon^2\mathbf{H}$.

2.2 Spatio-Temporal Smooth Models

When data are collected over space and time, models deal with these types of effects in different ways, but their formulation is constrained by the size of the data set, and the level of complexity used to estimate these models. Spatio-temporal smoothing is an approach computationally efficient, and, at the same time, it allows for the estimation of complex trends. The simplest example of this type of models is the additive model:

$$f(space) + f(time),$$

proposed by Kneib and Fahrmeir (2006) which ignores the space-time interaction, and cannot reflect important features in the data. This model also implies a spatio-temporal correlation structure given by separable covariance matrices for the spatial and temporal components. In most real applications, this approach is too simplistic, and its natural extension includes space-time interaction terms. Our proposal is based on the PS-ANOVA model introduced by Lee and Durbán (2011), including second- and third-order interactions between spatial and temporal components.

Let us assume that the data are collected at n spatial locations at t time points (the model could be easily generalized to the case in which the response variable is measured at different time points in each location), and no further covariates are available (in the next section we will combine this model with the semiparametric model introduced in the previous section). The smooth space-time model has the form:

$$\begin{aligned} \mathbf{y} &= \gamma + f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_t(\mathbf{x}_t) + \\ & f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_t) + f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_t) + \\ & f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t) + \boldsymbol{\epsilon} \end{aligned} \tag{11}$$

where $\mathbf{x}_{s_1}$ and $\mathbf{x}_{s_2}$ are the geographical coordinates of the spatial location, and $\mathbf{x}_t$ the vector of time points. The B-spline basis for model (11) would be:

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

and the penalty matrix is block-diagonal with blocks corresponding to the different terms in the model. In this case, several constrained need to be imposed, since the space spanned by $\mathbf{B}_i \otimes \mathbf{B}_j$, already spans the space spanned by the marginal bases $\mathbf{B}_i$ and $\mathbf{B}_j$ (see Lee, 2010a, for more details). Again, the transformation of model (11) into a mixed model, using the singular value decomposition of marginal penalties, will solve the identifiability problem yielding matrices:

- $\mathbf{X} = [(\mathbf{X}_{s_1}\Box\mathbf{X}_{s_2})\otimes\mathbf{X}_t]$
- $\mathbf{Z} = [(\mathbf{Z}_{s_1}\Box\mathbf{X}_{s_2})\otimes\mathbf{X}_t|(\mathbf{X}_{s_1}\Box\mathbf{Z}_{s_2})\otimes\mathbf{X}_t|(\mathbf{X}_{s_1}\Box\mathbf{X}_{s_2})\otimes\mathbf{Z}_t|(\mathbf{Z}_{s_1}\Box\mathbf{Z}_{s_2})\otimes\mathbf{X}_t$ $(\mathbf{Z}_{s_1}\Box\mathbf{X}_{s_2})\otimes\mathbf{Z}_t|(\mathbf{X}_{s_1}\Box\mathbf{Z}_{s_2})\otimes\mathbf{Z}_t|(\mathbf{Z}_{s_1}\Box\mathbf{Z}_{s_2})\otimes\mathbf{Z}_t]$,

where $\mathbf{X}_{s_k}$, $\mathbf{Z}_{s_k}$ $(k = 1, 2)$, $\mathbf{X}_t$, and $\mathbf{Z}_t$ are the mixed model matrices obtained for the reparametrization of the marginal basis and penalties (Lee and Durbán, 2011). The covariance matrix of random effects,$\mathbf{G}$, is given by

$$\begin{aligned} \mathbf{G}^{-1} = \text{blockdiag}\Bigg(&\mathbf{0}, \frac{1}{\tau_1^2}\widetilde{\mathbf{\Lambda}}_1, \frac{1}{\tau_2^2}\widetilde{\mathbf{\Lambda}}_2, \frac{1}{\tau_3^2}\widetilde{\mathbf{\Lambda}}_3, \frac{1}{\tau_4^2}\widetilde{\mathbf{\Lambda}}_4 + \frac{1}{\tau_5^2}\widetilde{\mathbf{\Lambda}}_5, \frac{1}{\tau_6^2}\widetilde{\mathbf{\Lambda}}_6 + \frac{1}{\tau_7^2}\widetilde{\mathbf{\Lambda}}_7, \frac{1}{\tau_8^2}\widetilde{\mathbf{\Lambda}}_8 + \frac{1}{\tau_9^2}\widetilde{\mathbf{\Lambda}}_9, \\ &\frac{1}{\tau_{10}^2}\widetilde{\mathbf{\Lambda}}_{10} + \frac{1}{\tau_{11}^2}\widetilde{\mathbf{\Lambda}}_{11} + \frac{1}{\tau_{12}^2}\widetilde{\mathbf{\Lambda}}_{12}\Bigg) \end{aligned}$$

where

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

and $\widetilde{\mathbf{\Sigma}}$ matrices correspond to the eigenvectors of the singular value decomposition of penalty matrices, and c_k and q_k are the dimensions of the bases and the order of the penalty used for the marginal smooth. It is worth noticing 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 intractable. We follow Lee et al. (2013) in using *nested* B-spline bases for the interactions terms, in order to reduce the computational burden without compromising the fit of the model. 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, in general, the information about the interaction usually is sparse. In the ANOVA context, the main effects are more important than the interactions, so in most situations this would be reasonable. The way to ensure that the new basis is nested relative to the original basis is to assume that the number of knots (`ndx*`) in $\breve{\mathbf{B}}$ is a divisor of the number of knots used to construct the original basis (`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. Matrices $\mathbf{Z}_{s_k}$ and $\mathbf{Z}_t$ would be modified by the corresponding $\breve{\mathbf{Z}}_{s_k}$ and $\breve{\mathbf{Z}}_t$ in $f_{(1,2)}$, $f_{(1,t)}$, $f_{(2,t)}$ and $f_{(1,2,t)}$. Estimation of fixed and random effects and variance components would proceed as in the case of a semiparametric model.

Before concluding this subsection, it is important to give some insights about the relevance and the meaning of the spatio-temporal components of the PS-ANOVA model 11 in applied econometric studies. First of all, as already pointed out in Basile et al. (2014), the geoadditive terms ($f_1(\mathbf{x}_{s_1})$, $f_2(\mathbf{x}_{s_2})$, and $f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$) work as control functions to filter the spatial trend out of the residuals, and transfer it to the mean response in a model specification. Thus, they allow to capture the shape of the spatial distribution of $\mathbf{y}$, eventually conditional on the determinants included in the model. This control function also isolates stochastic spatial dependence in the residuals, that is spatially autocorrelated unobserved heterogeneity. Thus, it can be regarded as an alternative to individual regional dummies to capture unobserved spatial heterogeneity as long as the latter is smoothly distributed over space. Regional dummies peak significantly higher and lower levels of the mean response variable. If these peaks are smoothly distributed over a two-dimensional surface (i.e. if unobserved spatial heterogeneity is spatially auto-correlated), the smooth spatial trend is able to capture them.

The smooth time trend, $f_t(\mathbf{x}_t)$, and the smooth interactions between space and time - $f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_t)$, $f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_t)$, and $f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t)$ - work as control functions to capture the heterogeneous effect of common shocks, thus allowing for strong cross-sectional dependence in the data. In this sense, the PS-ANOVA model 11 works as an alternative to the Common Correlated Effects (CCE) method proposed by Pesaran (2006) based on the use of cross-sectional averages of the observations. Nevertheless, like the CCE model, also the spatio-temporal ANOVA model 11 (even if combined with the semiparametric model 1) neglects the presence of weak cross-dependence (i.e. spatial spillovers) in the data. While Bai and Li (2015), Shi and Lee (2016), Pesaran and Tosetti (2011), Bailey et al. (2016) and Vega and Elhorst (2016) have proposed parametric spatial panel extensions of common factors models to accommodate both strong and weak cross-sectional dependence, here we rely on a combination of the ANOVA spatio-temporal trend model described above and the PS-SAR model developed by Montero et al. (2012); Basile et al. (2014) to handle simultaneously spatial spillovers and strong cross-sectional dependence (see the next subsection).

2.3 Spatio-Temporal PS-SAR Models

As mentioned above, spatio-temporal models do not permit to distinguish between strong and weak cross-sectional dependence (Chudik et al., 2011), because all the correlation between spatial units is collected by the spatio-temporal trend and, perhaps, by the effect of covariates. In order to assess the presence of residual spatial spillovers net of the effect of common factors, we combine the PS-ANOVA model 11 described in the previous section with the spatial lag model. By including also linear ($\mathbf{X}^*$), and nonlinear ($\mathbf{z}_j$) additive covariates, the full model becomes:

$$
\begin{aligned}
(\mathbf{A} \otimes \mathbf{I}_T)\mathbf{y} = {} & f_1(\mathbf{x}_{s_1}) + f_2(\mathbf{x}_{s_2}) + f_t(\mathbf{x}_t) + f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + \\
& + f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_t) + f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_t) + f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t) + \\
& + \mathbf{X}^*\boldsymbol{\beta}^* + \sum_{j=1}^{l} f(\mathbf{z}_j) + \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{NT}) \qquad (12) \\
\mathbf{A} = {} & \mathbf{I}_N - \rho \mathbf{W}_N
\end{aligned}
$$

where $\mathbf{W}_N$ is a neighborhood spatial matrix, and ρ is the spatial parameter associated to $\mathbf{W}_N$ matrix. It measures the degree of the spatial weak dependence net of the strong dependence. This model is flexible enough to gather both types of cross-sectional dependence. Specifically, the inclusion of the ANOVA decomposition of the spatio-temporal trend helps interpret the evidence of significance spatial spillovers as weak cross-dependence net of the effect of common effects (strong dependence). In this sense, model 12 can be regarded as a valid alternative to Bai and Li (2015), Shi and Lee (2016), or Bailey et al. (2016) which consider a jointly modeling of both spatial interaction effects and common-shocks effects. Our framework is also flexible enough to control for residual spatial heterogeneity (due to a spatial trend) and for the linear and non-linear functional relationships between the dependent variable and the covariates.

To estimate all the parameters of the model it is possible to maximize REML function as in (9) slightly modified by the kronecker matrix product $(\mathbf{A} \otimes \mathbf{I}_T)$:

$$
\begin{aligned}
\ell(\tau_i^2, \sigma^2, \rho) = & -\frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} \log |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \\
& - \frac{1}{2} \left[(\mathbf{A} \otimes \mathbf{I}_T)\mathbf{y} \right]' \left(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \right) \left[(\mathbf{A} \otimes \mathbf{I}_T)\mathbf{y} \right] + \\
& + \log |\mathbf{A} \otimes \mathbf{I}_T|
\end{aligned} \tag{13}
$$

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

To get estimates for all the parameters, we need to maximize the REML function which is a very complex numerical problem. Recently, Rodriguez-Alvarez et al. (2015) have 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:

$$
\mathbf{G}^{-1} = \sum_{i=1}^{12} \frac{1}{\tau_i^2} \boldsymbol{\Lambda}_i, \quad \boldsymbol{\Lambda}_i = \text{blockdiag}(\mathbf{0}, \ldots, \widetilde{\boldsymbol{\Lambda}}_i, \ldots, \mathbf{0}) \tag{14}
$$

This expression allows to get closed estimates for all the variance component parameters τ_i^2 and σ^2 very efficiently. We have adapted this algorithm to include also the estimation of ρ parameter. The steps to apply SAP algorithm to optimize (13) can be summarized as follows:

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

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

3. Compute the estimates:

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

where $\mathbf{\Lambda}_i \quad i = 1, \cdots, 12$ is defined in (14) and $\mathbf{\Lambda}_{z_j} = \text{blockdiag}\left(\mathbf{0}, \cdots, \widetilde{\Sigma}_{z_j}, \cdots, \mathbf{0}\right)$.

4. Estimate the variance components:

$$\hat{\tau}_i^{2,(k+1)} = \frac{\hat{\boldsymbol{\alpha}}^{(k)'}\mathbf{\Lambda}_i\hat{\boldsymbol{\alpha}}^{(k)}}{ed_i^{(k)}} \quad i = 1 \cdots, 12, \cdots, z_j \cdots$$

Estimate the variance of the noise as:

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

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

$$\frac{\partial \ell(.)}{\partial \rho} = -\frac{1}{2}\left[2\hat{\mathbf{P}}^{(k)}\left((\mathbf{A} \otimes \mathbf{I}_T)\mathbf{y}\right)\right]'\left(\frac{\partial(\mathbf{A} \otimes \mathbf{I}_T)}{\partial \rho}\mathbf{y}\right) + \text{trace}\left((\mathbf{A} \otimes \mathbf{I}_T)^{-1}\frac{\partial(\mathbf{A} \otimes \mathbf{I}_T)}{\partial \rho}\right) =$$
$$= \mathbf{y}'(\mathbf{A} \otimes \mathbf{I}_T')\hat{\mathbf{P}}^{(k)}(\mathbf{W}_N \otimes \mathbf{I}_T)\mathbf{y} - T\text{trace}(\mathbf{A}^{-1}\mathbf{W}_N) = 0$$

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:

$$\text{edf} = \sum_i ed_i^{(k)} + \text{rank}(\mathbf{X}) + 1$$

This quantity is increased in one unit with respect to spatio-temporal smooth models because of the need to estimate the ρ parameter.

The previous algorithm allows to get estimates of all the parameters of model (12) without the necessity to use numerical optimization and, as a consequence, a huge reduction of computational burden. Moreover, the inference techniques usually applied in REML estimation of mixed models can also be used for this case.

3 Empirical Case

Starting from Partridge and Rickman (1997) and Taylor and Bradley (1997), regional unemployment differentials have been 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 (static and dynamic) spatial panel framework (Lottmann, 2012; Basile et al., 2012; Rios, 2014). However, all these studies neglect the role of spatio-temporal trends. In particular, they do not consider the possibility that the observed spatial dependence in regional unemployment rates is partially driven by the existence of a spatial trend or a spatio-temporal trend in the data.

Moving from these considerations, we analyze the performance of the PS-SAR model with a spatio-temporal trend against different competing parametric and semiparametric models using data on regional unemployment in Italy. We first describe these data and their features in terms of spatial and temporal trend (section 3.1). Then, we briefly discuss the theoretical background and the set of variables used to explain regional unemployment differentials (section 3.2). Finally, we report the results of the econometric analysis (section 6).

3.1 Regional unemployment data

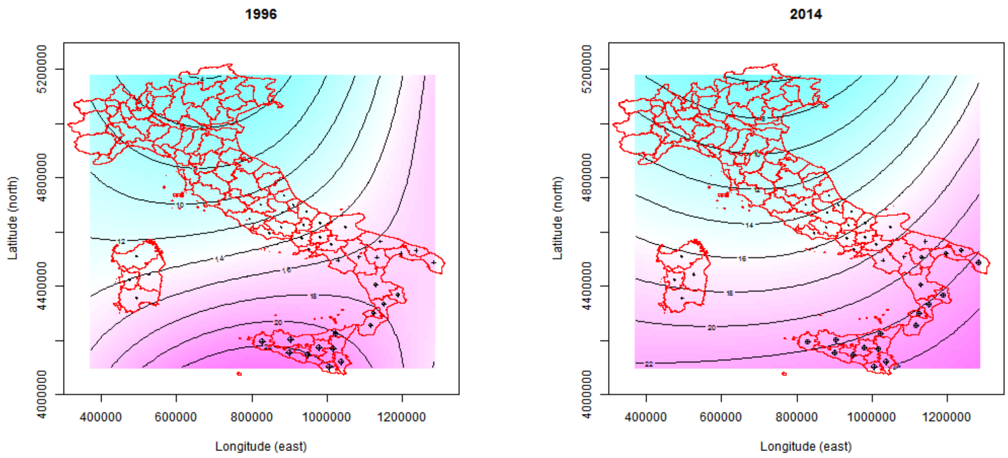
The data on regional unemployment rates ($unrate_{i,t}$) for each Italian province $i = 1, ..., N$ (N=103) which corresponds to an Italian NUTS-3 region, and for each time period $t = 1996, ..., 2014$ (T=19) used in this analysis are provided online by the Italian National Institute of Statistics (ISTAT). 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 and $LF_{i,t}$ is the labor force.

Regional unemployment rates differ strongly in Italy. In 2014, Southern provinces showed an average unemployment rate of 20% with a standard deviation of 4%, while Northern provinces registered an average rate of 10% with a standard deviation of 3%. In the same year, the province with the lowest unemployment exhibited a rate of 4.4% (Bolzano in Trentino-Alto-Adige), while the highest regional unemployment rate amounted to 27.8% (Cosenza in Calabria). In 1996, the North-South gap was already very high: Southern provinces registered an average unemployment rate of 17% (the standard deviation was 6%), while Northern provinces had an average rate of 7% (the standard deviation was 3%). The province with the lowest unemployment rate (2.2%) in 1996 was Lecco (in Lombardia), while the province with the highest rate (32%) was Enna (in Sicily).

3.1.1 Spatial and time trends

The North-South divide can also be depicted by mapping the predicted values of a simple regression of provincial unemployment rates on the smooth interaction between longitude and latitude (figure 1). A clear (albeit nonlinear) spatial trend emerges and is persistent over time. These findings might 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 and, 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 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 the demand or on the supply side).

A nonlinear time trend also characterizes unemployment data. The national unemployment rate (red line in Figure 2) shows a fall from 1996 (11.2%) to 2007 (6.1%); with the outbreak of the financial crisis and its extension to the productive economy in the subsequent years, it picked up reaching 12.7% in 2014. Both Northern and Southern provinces followed a similar time path, thus suggesting that common business cycles factors affect all the regions. However, it is also evident from Figure 2 that there are relevant differences across provinces, thus indicating that regions may differ in their elasticity to common shocks. This feature is rather usual in regional unemployment studies. Thus,



(a) Year: 1996

(b) Year: 2014

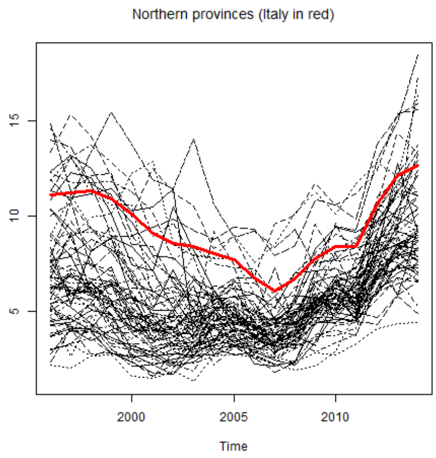
FIGURE 1
Spatial trend of provincial unemployment rates

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 one 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, while the usual finding in many applied settings is that some regions are more sensitive 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 has also 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.

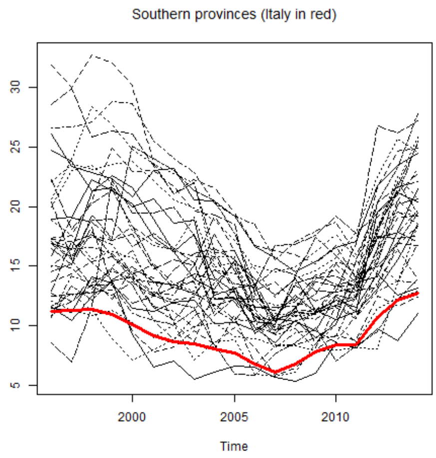
3.1.2 Spatial dependence and cross-sectional dependence

As mentioned above, recent studies have applied spatial econometric models to analyze the determinants of regional unemployment. Spatial autocorrelation is justified on the basis of a theoretical 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 will not only affect the respective labor market, but spills over to neighboring regions. Given this interdependence, the induced changes of 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.

Many empirical studies confirm the existence of positive spatial autocorrelation (or 'weak cross-section' dependence) in the data. Nevertheless, they neglect the possibility that the observed spatial dependence in regional unemployment rates may be (at least



(a) North



(b) South

FIGURE 2
Time trend of provincial unemployment rates: 1996-2014

in part) the result of the existence of a strong spatial trend in the data. To clarify the issue, we compute year-by-year Moran I statistics for the regional unemployment rates with and without filtering out the spatial trend in the data (Figure 3).[2]

A distance based spatial weights matrix (W) has been used to compute Moran's I statistics as well as to estimate spatial lag models throughout the paper. 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}:

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

where d_{ij} is the great-circle distance between the centroids of provinces i and j.[3] The selected cut-off distance (d^*) corresponds to the minimum distance that allows all provinces to have at least one neighbor.

Using the raw data, the standardized Moran's I ranges between 8 and 11, thus suggesting a very strong spatial autocorrelation. Then, to capture the North-South trend emerging from Figure 1, we simply regress for each year the unemployment rates on the latitude and test for spatial autocorrelation in the residuals. Removing such a linear trend, the Moran I values decrease a lot, ranging between 3 and 5, but still they always provide evidence of a significant positive spatial autocorrelation. Finally, once we filter out a non-linear smooth trend (that is once we regress for each year the unemployment

[2]As well known, the Moran's I test test does not correct for serial dependence among observations. Therefore, the results reported in Figure 3 must only be considered as a way to assess the effect of the spatial trend on the spatial dependence.

[3]Geographic distance has frictional effects on labor market activity. Workers prefer to find a job in their closer environment because commuting and moving entail monetary and psychological costs. Therefore, we use great circle distances between centroids of provinces to define the entries of the spatial weights matrix.

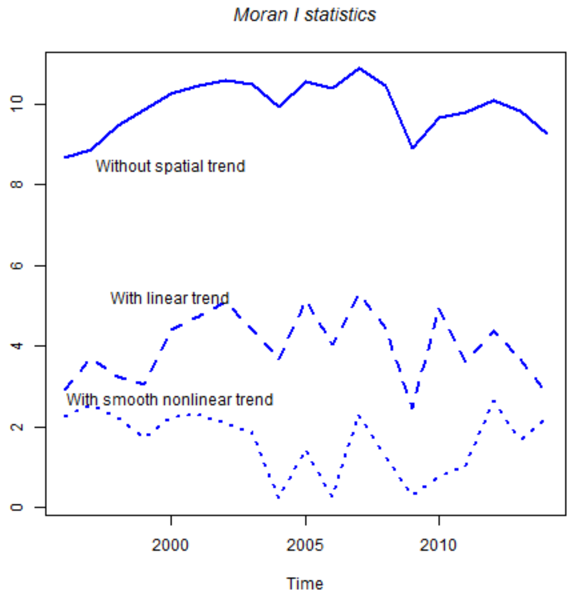


FIGURE 3
Moran I statistics of provincial unemployment rates: 1996-2014

rate on the smooth interaction between latitude and longitude and test for spatial autocorrelation in the residuals), the Moran's I statistic is always lower than 2.7 and in 6 cases out 19 (during the period between 2004 and 2011) it turns out to be not statistically significant.

The evidence of positive spatial autocorrelation (or 'weak cross-section' dependence) in the data may also mask the existence of 'strong' cross-sectional dependence due to common cyclical factors. Strong cross-sectional dependence can be tested using th CD test developed in Pesaran (2004) and Pesaran (2015). Differently from the the Moran's I test, the CD 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 (CD=180.7) is highly significant, confirming the existence of cross-sectional dependence (see Table 1). Applying the same test on the residuals of an AR(2) model (to accommodate for serial correlation), we obtain a CD value of 87.4 still highly significant. As also suggested by Bailey et al. (2016), however, this result does not exclude the possibility that both forms of dependence (weak and strong) are present in the data. Therefore, we may conclude that potential sources of interaction between regions are both weak due to for example commuting flows, and strong due to common factors.

	Test statistic	p-value
Without control for serial correlation	180.7	0.00
With control for serial correlation	87.4	0.000

TABLE 1
Cross-sectional dependence test (Pesaran, 2004, 2015)

An important issue is to assess the stationarity of regional unemployment rates series. To this end, we use panel unit root tests. The results of both the standard and the cross-sectionally demeaned Im et al. (2003) (IPS) tests do not allow us to reject the null hypothesis of a unit root in regional unemployment rates. However, the robust cross-sectional dependence test proposed by Pesaran (2007) clearly rejects the hypothesis of a unit root at all reasonable significance levels. Hence, these test results give a strong indication regarding stationarity of the data once cross-sectional dependence is taken into account.

Deterministic component	Standard IPS	Cross-sectionally demeaned IPS	Robust against cross-sectional dependence IPS
None	0.625	-0.626	-1.462*
Drift	-0.930	-1.603	-2.387**
Drift and trend	-1.603	-1.603	-10.109**

TABLE 2
Panel unit root tests for regional unemployment rates

3.2 Background and explanatory variables

3.2.1 Equilibrium and disequilibrium view

Economic theory provides two different explanations on the nature of regional unemployment disparities: the *equilibrium view* and the *disequilibrium view*. The first approach is based on the hypothesis of a stable equilibrium of spatial labor markets, where equilibrium is defined as "a situation of uniform utility across areas for (each) homogeneous labour group, such that there are no incentives for further labour migration (a further condition would be uniform profits such that capital movements are eliminated)" (Molho, 1995, p. 642). Accordingly, long-run differentials represent an equilibrium where factors such as favorable climatic conditions or an attractive social or institutional environment encourage people to stay in regions where unemployment rates are high (Marston, 1985). Each region tends to its own equilibrium unemployment rate which is determined by regional demand and supply factors, amenities and institutions. Therefore, a high unemployment rate in a given area needs to be compensated by some other positive factors which act as a disincentive to migration.

According to the disequilibrium view, in the long run all regions tend to a competitive equilibrium unemployment rate and the unemployment rate will level off across areas (Blanchard et al., 1992). In the short run, however, regional disparities may reflect labor market rigidities that restrict mobility or slow adjustment processes to asymmetric shocks (e.g., a shortage of labor demand). The adjustment process may be faster or slower, and depending on its speed, local unemployment disparities could persist for a long time. The speed of adjustment may depend on determining factors connected to both labor demand and supply. This view stems from neoclassical theory where with increased economic integration and the removal of impediments to the free flow of production factors, unemployment rates should converge given the convergence in factor returns.

Thus, the unemployment rate is a reduced form function of a variety of factors affecting labor demand, supply and wages. According to the pioneering work of Partridge and Rickman (1997), these factors can be broadly categorized as disequilibrium factors (e.g., employment growth rates), and 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, the set of our variables is limited by data availability.

3.2.2 Selection of explanatory variables

Seven explanatory variables have been selected for the econometric analysis: *i*) Employment growth rate (*empgrowth*), *ii*) Employment share in agriculture (*agri*), *iii*) Employment share in industry (*ind*), *iv*) Employment share in construction (*cons*), *v*) Employment share in services (*serv*), *vi*) Labor force participation rate (*partrate*), *vii*) Population density in logs (*lpopdens*). Table 3 reports simple descriptive statistics of these variables.

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. dev.
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
lpopdens	-3.49	-2.24	-1.75	-1.76	-1.32	0.98	0.77

TABLE 3
Summary statistics

In order to account for regional disequilibrium labor market dynamics, the employment growth rate ($empgrowth_{i,t} = 100 \times \frac{Emp_{i,t} - Emp_{i,t-1}}{Emp_{i,t-1}}$), i.e. the annual growth rate in province employment, is included in the set of explanatory variables. It is expected to have a negative effect on unemployment. This is not surprising because the change in employment directly affects unemployment. Another variable capturing disequilibrium effects are wages or unit labor costs. Unfortunately, this data is only available at NUTS-2 level and not at the NUTS-3 (province) level. So, we decided to exclude it from the analysis.

The other variables 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 and industry, might show higher structural unemployment rates than provinces specializing in services and construction. The share of employment in agriculture ($agri_{i,t} = 100 \times \frac{Agri_{i,t}}{Emp_{i,t}}$), in the industry sector ($ind_{i,t} = 100 \times \frac{Ind_{i,t}}{Emp_{i,t}}$), in the service sector ($serv_{i,t} = 100 \times \frac{Serv_{i,t}}{Emp_{i,t}}$) and in the construction sector ($cons_{i,t} = 100 \times \frac{Cons_{i,t}}{Emp_{i,t}}$) over total provincial employment are proxies of the provincial economic structure.

The labor force participation rate ($partrate_{i,t} = 100 \times \frac{LF_{i,t}}{Workpop_{i,t}}$) is the ratio between total labor force and the working population (population aged between 16 and 65 years). It is an indicator of labor supply. The expected effect is negative since factors determining low participation rates in a particular region also reflect relatively low investments in human capital and low commitment to working life, resulting in higher risks for people with these characteristics to become unemployed. However, a positive effect may also occur if a faster growth of the labour force (i.e., young people) is not compensated by an as mush faster growth of new jobs (or vacancies).

As stated above, amenities are considered as 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 and we only included the log of population density, that is the ratio between total population and the area surface of the province ($lpopdens_{i,t} = \log \frac{Pop_{i,t}}{Area_i}$), as a proxy for urbanization following López-Bazo et al. (2005) and Cracolici et al. (2007). However, the sign of its coefficient is not unambiguous. On the one hand, a high urban density may increase the efficiency of matching workers to jobs (unemployed persons have more employment opportunities and the matching process is expected to be more efficient in urban areas), but on the other hand, it may increase the time spent by workers to collect information about vacancies on the job market.

Obviously, population density cannot capture all kinds of regional amenities explaining regional differences in unemployment rates. In addition, there are many other equilibrium and disequilibrium variables affecting regional unemployment differentials. These include workers migration and commuting, which are relevant in this type of spatial context, the age structure of the population and human capital variables. This means that there is a huge amount of spatial *unobserved heterogeneity* in modeling regional unemployment rates. The inclusion of a *spatial trend* in the model is a way to clean up the residuals. In other words, the spatial trend captures all time-invariant region-specific unobservable factors, simultaneously allowing these factors to be freely correlated with observable determinants of regional unemployment rates. The alternative approach used in the literature consists of introducing spatial fixed effects in the model to measure time-invariant unobservable equilibrium effects.

Moreover, as stated above, unobserved common business cycle factors may influence labor market dynamics with a heterogeneous effect across regions and the CCE approach (or a combination of the CCE with a spatial model) has been recently proposed as a possible solution (Bai and Li, 2015; Shi and Lee, 2016; Pesaran and Tosetti, 2011; Bailey et al., 2016; Vega and Elhorst, 2016). The PS-ANOVA model with spatio-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* ($\sum_{jt} w_{ij} unrate_{jt}$) on the right-hand side of the model. It is important to remark again that this endogenous variable should only capture substantive spatial dependent (i.e. externalities in regional labor markets) – which 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 experienced by the remaining regions – rather than spatially correlated unobserved heterogeneity or common time effects.

3.3 Econometric results

3.3.1 Model selection and diagnostrics

We use the data described above to compare the performance of the spatio-temporal PS-SAR model against different competing parametric and semiparametric models in terms of model fitting and residual diagnostics, focusing on the test for cross-sectional dependence in the residuals (see Table 4 for the list of models considered and the Appendix for their formal representation).

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	Spatio-Temporal SAR model with linear terms (PS-SAR-ANOVA-Linear)
Model 9	Spatio-Temporal model with nonlinear terms (PS-ANOVA-Nonlinear)
Model 10	Spatio-Temporal SAR model with nonlinear terms (PS-SAR-ANOVA-Nonlinear)

Table 4
List of models

The most restricted specifications are the *parametric* a-spatial linear models with spatial fixed effects (**Model 1**) and with spatial and time fixed effects (**Model 2**), estimated using the standard fixed effects 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. Including interactions between individual fixed effects and cross-sectional averages of the data (**Model 5**) (that is using the pooled common correlated effects - CCEP - estimator proposed by Pesaran, 2006), the evidence of cross-dependence disappears (see Table 5).[4] These results strongly confirm the existing literature. However, using 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. Moreover, using the CCEP estimator, the unobserved common factors are treated as unknown parameters but the factor loadings are interactive individual effects that induce an incidental parameter problem since their number grows with N. In other words, the CCEP estimator requires a huge number of degrees of freedom (*edf*) which is reflected in a high BIC value. [5]

[4]Pesaran (2006) noted that linear combinations of the unobserved factors can be well approximated by cross-sectional averages of the dependent variable and the observed regressors. This leads to a new set of estimators, referred to as the Common Correlated Effects (CCE) estimators, that can be computed by running standard panel regressions augmented with the cross-sectional averages of the dependent and independent variables.

[5]The EDF's 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.

TABLE 5
Model selection and diagnostics

Model	CD test	p-value	rho	LR test	p-value	total EDF	σ^2	BIC
Parametric models								
Model 1	86.64	(0.00)				110.00	4.78	3791.07
Model 2	2.48	(0.01)				128.00	3.12	3069.46
Model 3	-0.13	(0.90)	0.46	23.76	(0.00)	111.00	3.43	3146.04
Model 4	0.26	(0.80)	0.25	11.07	(0.00)	129.00	2.85	2898.96
Model 5	-1.54	(0.12)				934.00	0.90	5601.85
Model 6	0.16	(0.87)	0.04	1.48	(0.14)	953.00	0.91	5742.51
Spatio-temporal models								
Model 7	-0.09	(0.93)				124.58	2.64	2720.39
Model 8	-0.37	(0.71)	0.13	72.16	(0.00)	124.60	2.54	2646.94
Model 9	0.14	(0.89)				163.24	2.24	2648.92
Model 10	-0.19	(0.85)	0.10	45.17	(0.00)	162.96	2.19	2607.62

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, 2014b), 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 likely the spatial lag term has 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 fixed effects and cross-sectional averaged of dependent and independent variables, in line with recent contributions (Bai and Li, 2015; Shi and Lee, 2016; Bailey et al., 2016; Vega and Elhorst, 2016), we should allow for both strong and weak cross-dependence. Nevertheless, the evidence is only in favor of strong dependence as the ρ parameter is not statistically significant (according to the results of a likelihood ratio test) and, obviously, Model 6 shares with model 5 the problem of the large number of *edf* (and, thus, high BIC).

The results of the CD test on the residuals confirm that the smooth spatio-temporal trend (**Models 7-10**) is able to capture the unobserved cross-sectional dependence and, thus, it represents a valid alternative to the inclusion of cross-sectional averages in the model (i.e. the CCE method). **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 section 2.** Indeed, the BIC values of PS-ANOVA models are much lower than those computed for any other model. Moreover, the estimated ρ parameter for the two SAR versions (Models 8 and 10) is statistically significant, indicating the existence of weak dependence net of the effect of common business cycle effects. This parameter, however, is much lower than the one estimated with the SAR-FE (0.46) and the SAR-FE/TE models (0.25). In absolute terms, Model 10 (i.e. the spatio-temporal ANOVA SAR model with nonlinear terms) shows the best performance with a BIC value of 2,607, lower than its linear counterpart (model 8), suggesting that the functional form of the relationship between the response variable (regional unemployment) and the covariates cannot be assumed to be linear. It is also worth noticing that, by allowing for nonlinearities, the value of the estimated ρ parameter decreases from 0.13 to 0.10, confirming the trade-off between spatial autocorrelation and nonlinearities.

Finally, we have more deeply investigated the role of the ANOVA spatio-temporal trend in capturing the unobserved cross-sectional dependence. Specifically, we have tested the assumption that the smooth time trend, $f_t(\mathbf{x}_t)$, along with the smooth interaction between space and time - $f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_t)$, $f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_t)$, and $f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t)$ - work as control functions to capture the heterogeneous effect of common shocks. To this end, we have re-estimated model 9 with only the spatial trend - that is with $f_1(\mathbf{x}_{s_1})$, $f_2(\mathbf{x}_{s_2})$, and $f_{1,2}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$, and without $f_t(\mathbf{x}_t)$, $f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_t)$, $f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_t)$, and $f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t)$. Pesaran's CD statistics computed on the residuals of this model turned out to be equal to 81.2, clearly indicating that the null hypothesis of no cross-sectional dependence is rejected. We have also estimated this spatial ANOVA model including the main time trend, $f_t(\mathbf{x}_t)$, but not the interactions between space and time. Although the CD statistics decreases a lot (from 81 .5 to 9.4), it remains still significant. All in all, we may conclude that the heterogeneous effect of common shocks is only partially gathered by $f_t(\mathbf{x}_t)$ (consistently with the result of the CD test for the residuals of the parametric FE/TE model), while the interaction terms between space and time trends - $f_{1,t}(\mathbf{x}_{s_1}, \mathbf{x}_t)$, $f_{2,t}(\mathbf{x}_{s_2}, \mathbf{x}_t)$, and $f_{1,2,t}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}, \mathbf{x}_t)$ - are necessary control functions to fully capture strong cross-sectional dependence.

3.3.2 Estimation results

Table 6 reports the estimated β parameters of the linear terms included in models 1-8, along with the associated standard errors. Obviously, these parameters can be interpreted as marginal effects only in non-spatial models (models 1, 2, 5 and 7), while the interpretation of the various SAR specifications (models 3, 4, 6 and 8) requires the computation of direct and indirect marginal effects, reported in Table 7.[6]

[6] Any SAR model 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, $\mathbf{A}_n \equiv (\mathbf{I}_n - \rho \mathbf{W}_n)^{-1} \equiv \mathbf{I}_n + \rho \mathbf{W}_n + \rho^2 \mathbf{W}_n^2 + ...$, 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* spatial effects. *Direct* effects measure the impact of a change in regressor x_k in region i on the outcome of the same region: $\frac{\partial y_i}{\partial x_{ki}}$, while *indirect* effects measure the impact of a change in regressor 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 summmarize 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 $\overline{M}_{ind}^k = \overline{M}_{ind}^k = \overline{M}_{tot}^k - \overline{M}_{dir}^k$. 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 ML estimates. Efficient simulation approaches can be used to produce an empirical distribution of the parameters $\boldsymbol{\beta}, \boldsymbol{\theta}, \rho, \sigma^2$ that are needed to calculate the scalar summary measures. This distribution can be constructed using a large number of simulated parameters drawn from the multivariate distribution

Table 6
Estimation results

Model	*empgrowth*	$\ln$ *popdens*	*partrate*	*agri*	*ind*	*cons*	*serv*
$\hat{\beta}$ and sd($\hat{\beta}$) (between parenthesis) for linear terms							
Model 1	-0.15	6.88	0.46	0.71	0.47	-0.42	0.38
	(0.02)	(1.86)	(0.03)	(0.06)	(0.02)	(0.06)	(0.04)
Model 2	-0.16	16.88	0.52	0.21	0.20	-0.05	0.30
	(0.01)	(1.71)	(0.03)	(0.07)	(0.06)	(0.06)	(0.07)
Model 3	-0.13	3.95	0.38	0.32	0.23	-0.40	0.15
	(0.01)	(1.53)	(0.03)	(0.05)	(0.02)	(0.05)	(0.03)
Model 4	-0.15	13.49	0.47	0.18	0.20	-0.07	0.27
	(0.01)	(1.61)	(0.03)	(0.07)	(0.05)	(0.06)	(0.06)
Model 5	-0.18	35.47	0.59	-0.01	-0.07	0.05	-0.10
	(0.01)	(12.27)	(0.04)	(0.16)	(0.14)	(0.12)	(0.15)
Model 6	-0.18	35.23	0.59	-0.01	-0.07	0.04	-0.10
	(0.01)	(8.39)	(0.03)	(0.11)	(0.10)	(0.09)	(0.10)
Model 7	-0.10	0.36	0.15	0.18	0.12	0.01	0.17
	(0.01)	(0.13)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)
Model 8	-0.11	0.27	0.15	0.16	0.10	-0.01	0.15
	(0.01)	(0.13)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)
EDF for non-linear terms							
Model 9	4.84	4.92	6.16	5.18	5.55	5.18	5.74
Model 10	4.84	4.70	6.16	5.12	5.39	5.22	5.72

The results suggest that there is a clear explanation of unemployment differentials in terms of spatial equilibrium and disequilibrium factors and a significant degree of spatial dependence among labor markets at the provincial level in Italy. In all non-spatial models higher employment growth rates lowers provincial unemployment rates, as suggested by the disequilibrium approach. The magnitude of the estimated β parameter associated to the variable *empgrowth* is rather stable across the non-spatial models in spite of the strong differences among the various specifications. Both average direct and indirect marginal effects of this variable computed for the fixed effects SAR models (3 and 4) have a negative sign and are strongly significant indicating that an increase in the employment growth rate in one region reduces not only the unemployment rate of that region, but also the unemployment rate of the other provinces with a distance decay effect. However, spatial spillovers (indirect effects) are not significant in the case of model 6 (SAR-CCEP) and they are much lower in the case of the spatio-temporal ANOVA PS-SAR model 8.

Regional unemployment rates turn out to be positively affected by labor force participation rates in any model. The positive effect 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 faster growth of the labor force (i.e., young people) in contrast to a lower growth of new jobs (or vacancies). Increasing population density exerts detrimental effects on local labor market performances; the parameters associated to the variable $\ln$ *popdens* vary greatly among the different non-

of the parameters implied by the ML estimates.

spatial model specifications. Both direct and indirect marginal effects of these variables computed for the four SAR models have a positive sign.

The coefficients of the regressors related to the structure of the economy also turn out to be very different among the various specifications. Furthermore, the shares of employed persons working in the construction industry have a negative or a null impact on regional unemployment. Hence, provinces that are specialized in these industries exhibit lower unemployment than provinces with a different industrial structure.

TABLE 7
Direct, indirect and total marginal effects in SAR models. ** (*) indicates significance at 1% (5%)

Model		*empgr.*	ln *popd.*	*partr.*	*agri*	*ind*	*cons*	*serv*
Model 3	Direct	-0.14**	4.25**	0.41**	0.34**	0.25**	-0.43**	0.16**
	Indirect	-0.10**	3.04**	0.29**	0.25**	0.18**	-0.31**	0.12**
	Total	-0.25**	7.29**	0.70**	0.59**	0.42**	-0.73**	0.28**
Model 4	Direct	-0.16**	13.75**	0.48**	0.18**	0.20**	-0.07	0.28**
	Indirect	-0.05**	4.22**	0.15**	0.06**	0.06**	-0.02	0.08**
	Total	-0.21**	17.98**	0.63**	0.24**	0.26**	-0.09	0.36**
Model 6	Direct	-0.18**	35.25**	0.59**	-0.01	-0.07	0.04	-0.10
	Indirect	0.00	1.29	0.02	0.00	0.00	0.00	0.00
	Total	-0.18**	36.54**	0.61**	-0.01	-0.07	0.04	-0.10
Model 8	Direct	-0.11**	0.27*	0.15**	0.16**	0.10*	-0.01	0.15**
	Indirect	-0.02**	0.04*	0.02**	0.02**	0.01	0.00	0.02**
	Total	-0.13**	0.31*	0.17**	0.18**	0.11*	-0.01	0.17**

Table 6 also reports the *edf* for the nonlinear terms included in models 9 and 10, a broad measure of nonlinearity (an *edf* equal to 1 indicates linearity, while a value higher than 1 indicates nonlinearity). Focusing on model 10, that is the one with the best performance, we report the plots of direct and indirect effects of the smooth terms in figure 4. Starting from direct effects, nonlinearities in the relationship between regional unemployment rates and the covariates are clearly detected, although most of the figures display monotonic relationships. Specifically, an increase in the employment growth rate within a province is negatively associated with a reduction in the unemployment rate in the same province, but the direct effect vanishes for employment growth rates higher than 5%. The positive direct effect of the participation rate is particularly strong and highly significant at low levels of the variable, as indicated by the narrow confidence band. This means that, starting from very low levels of the participation rate (a status which characterizes Southern provinces), a faster growth of the labor force (due to the enter of young or previously discouraged people) is not compensated by an as much faster growth of new jobs. After a certain threshold, the detrimental effect of the participation rate decreases, probably because after such a threshold people entering the labor force have a lower risk to become unemployed. The level of uncertainty in the relationship between population density and unemployment rates is rather high (the confidence band contains the zero horizontal line for a large range of values of this explanatory variable) and does not allow us to make any ultimate statement. The direct effect of *agri*, *ind* and *serv* appears monotonically positive for most of the range of these independent variables, while the direct effect of *cons* turns out to be negative for low and high levels of the variable, confirming that provinces specialized in construction exhibit lower unemployment than provinces with a different sectoral structure. As expected, indirect effects are always much lower than direct effects; nevertheless, these effects remain statistically significant since the point-wise confidence interval crosses the zero horizontal line.

Finally, figures 6 and 5 report the yearly estimated spatial trend maps, and the regional specific time trends, respectively, from model 10. The map plots clearly show that, even after controlling for the role of equilibria and disequilibria factors, as well as for common time effects, the spatial distribution of expected regional unemployment rates remains persistently characterized by a strong North-South spatial trend. The estimated regional specific temporal trends also confirm the presence of common business cycles factors heterogeneously affecting all the regions.

FIGURE 4

Direct and Indirect functions for each covariate in Spatio-Temporal ANOVA SAR with nonlinear terms (Model 10). The intervals correspond to 95% of confidence

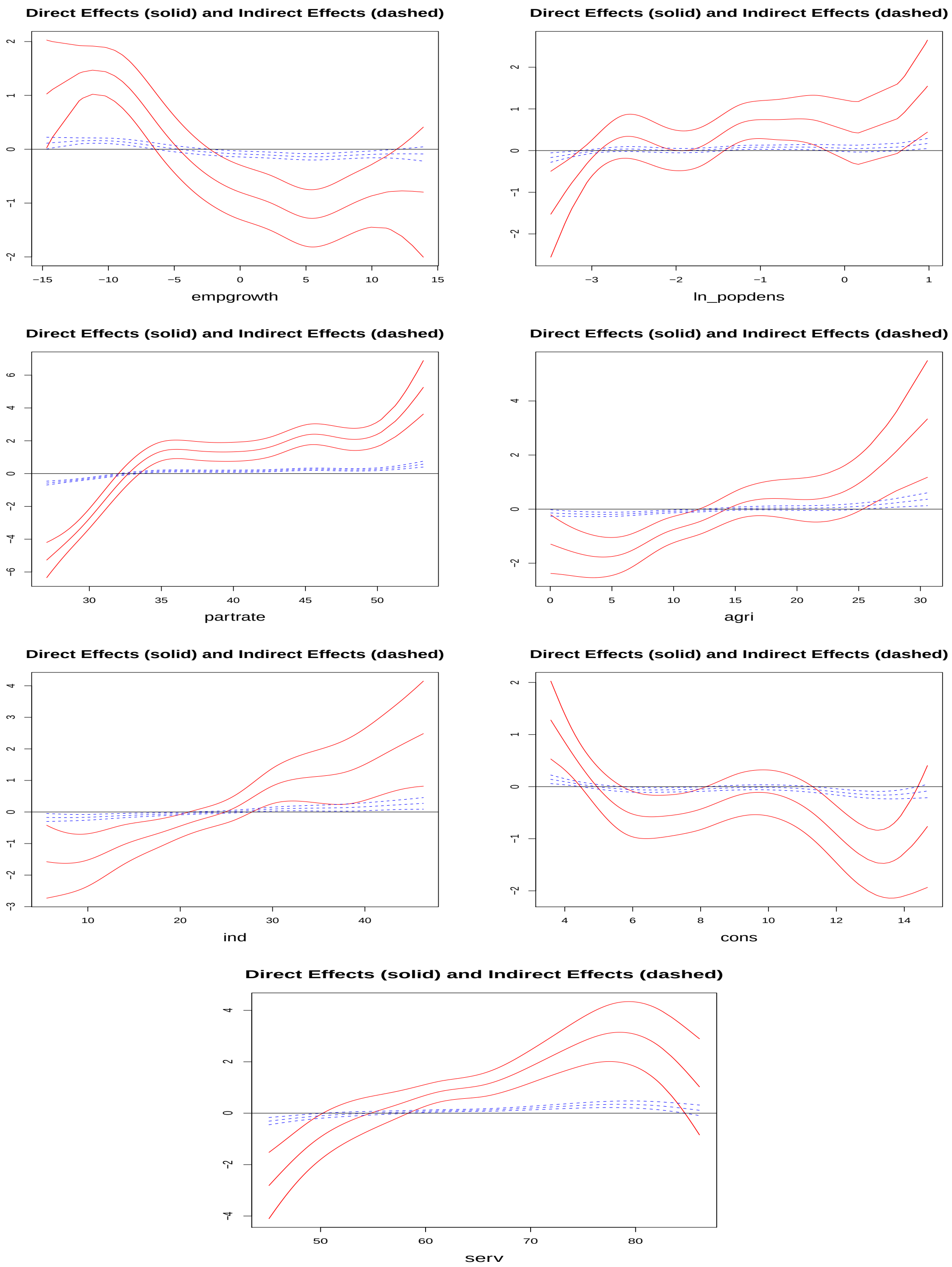
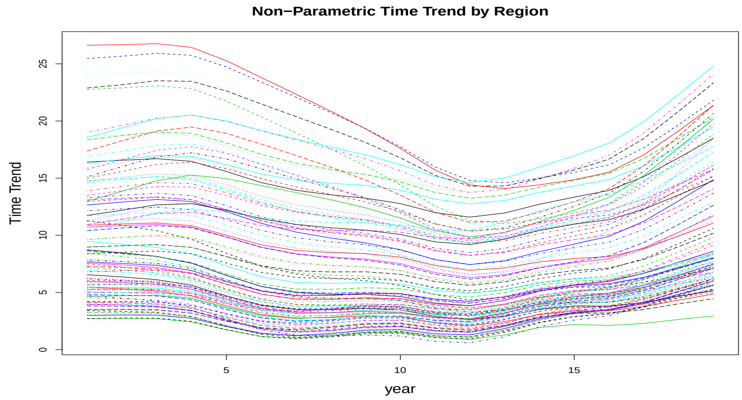


FIGURE 5
Regional time trends estimated by the Spatio-Temporal ANOVA SAR with nonlinear terms (Model 10)



4 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 interactions (spatial spillovers) or common factors (aggregate shocks). Spatial spillovers are the results of local interactions and, thus, are classified as weak dependence effects; while common factors represent latent economic-wide technological and/or demand shocks, heterogeneously affecting all regions' dynamics and, thus, are classified as strong dependence effects.

Traditionally, each type of effect has been analyzed separately in the literature by the so-called 'factor' approach and 'spatial econometric' approach, respectively. Recently, however, some authors have proposed a joint modeling of both types to determine whether one or both of these effects are present (Bailey et al., 2016; Vega and Elhorst, 2016; Bai and Li, 2015; Shi and Lee, 2016). Specifically, Bailey et al. (2016) and Vega and Elhorst (2016) follow Pesaran (2006) in using cross-sectional averages of the observed variables as proxies for common factors.

In the present paper, we have shown that, the spatio-temporal trend can be interpreted as an alternative to cross-sectional averages of the observations to capture the heterogeneous effect of unobserved common factors. Specifically, the ANOVA decomposition of the spatio-temporal trend in 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 interpret the evidence of significance spatial spillovers as weak cross-dependence net of the effect of common effects (strong dependence). In this sense, our proposed framework can be regarded as a valid alternative to parametric approaches which considers a jointly modeling of both spatial interaction effects and common-shocks effects. We have implemented this new framework using real data on unemployment rates in Italy. The results clearly suggest that the PS-SAR model with the ANOVA spatio-temporal trend outperforms parametric panel data SAR models with common effects.

As a concluding remark, it is worth noticing that regional unemployment rates, like many other regional and national economic variables, are typically characterized by strong persistence over time. Thus, a control for serial correlation is definitely needed. Our future research agenda will address this challenge by extending the PS-SAR model with the ANOVA spatio-temporal trend to a dynamic specification.

Appendix: Estimated models

The analysis of regional unemployment rates in Italy is based on the comparison of ten different econometric models. The first four are fully parametric standard fixed effects models with and without a spatial lag; the fifth and the sixth are still parametric models but with unobserved common effects; while the last four are semiparametric models with spatio-temporal trends.

Models 1-4 - Fixed effects models with and without spatial lag terms (FE, FE/TE, SAR-FE and SAR-FE/TE)

Let y_{it} be the observation on the i-th cross section unit at time t for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ and suppose that it is generated according to the following linear panel data model

$$y_{it} = \alpha_i + \tau_t + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \qquad [Model\, 2 \quad (FE/TE)]$$

where $\mathbf{x}_{it}$ 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 τ_t capture unobserved temporal heterogeneity, that is the effect of the omitted variables that are peculiar to each time period. α_i and τ_t are allowed to be correlated with $\mathbf{x}_{it}$, while the idiosyncratic errors, ε_{it}, are assumed to be independently distributed over $\mathbf{x}_{it}$. Consistent $\boldsymbol{\beta}$ parameters are estimated using the standard within-group estimator.

Model 1 is obtained from model 2 by imposing the restrictive assumption $\tau_t = 0$:

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \qquad [Model\, 1\, (FE)]$$

Fixed effects models can be extended to a spatially lagged dependent variable:

$$y_{it} \;\; = \;\; \alpha_i + \tau_t + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \qquad [Model \quad 4 \quad (SAR - FE/TE)]$$

where $W_N = (w_{ij,N})_{N \times N}$ is a specified spatial weights matrix whose diagonal elements $w_{ii,N}$ are 0; and ε_{it} are the idiosyncratic errors. $\rho \sum_{j=1}^{N} w_{ij,N} y_{jt}$ captures the spatial spillover effects. This model is estimated using a quasi-maximum likelihood estimator (QMLE) (Elhorst, 2014a; Lee and Yu, 2010).

Model 3 is obtained from model 4 by imposing the restrictive assumption $\tau_t = 0$:

$$y_{it} \;\; = \;\; \alpha_i + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \qquad [Model\, 3\, (SAR - FE)]$$

Model 5-6 - Models with unobserved common effects (CCEP and SAR-CCEP)

Now, suppose that y_{it} is generated according to the following linear heterogeneous panel data model

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \gamma_i^{'}\mathbf{f}_t + \varepsilon_{it} \qquad [Model\, 5\, (CCEP)]$$

where $\mathbf{f}_t$ is the $m \times 1$ vector of unobserved common effects (introduced to allow for cross-sectional dependence) with γ_i the corresponding heterogeneous response. The common factors are allowed to be correlated with $\mathbf{x}_{it}$, while the idiosyncratic errors, ε_{it}, are assumed to be independently distributed over $\mathbf{x}_{it}$. Pesaran (2006) has shown 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 Common Correlated Effects Pooled (CCEP) estimator, that can be viewed as a generalized fixed effects estimator.

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

$$y_{it} = \alpha_i + \rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \mathbf{x}'_{it}\boldsymbol{\beta} + \gamma'_i \mathbf{f}_t + \varepsilon_{it} \qquad [Model\ 6\ (SAR-CCEP)]$$

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

Models 7-10 - Spatio-Temporal models (PS-ANOVA-Linear, PS-SAR-ANOVA-Linear, PS-ANOVA-Nonlinear, PS-SAR-ANOVA-Nonlinear)

The last 4 specifications are semiparametric models always including an ANOVA spatio-temporal trend. Model 7 is based on the restrictive linearity assumption for the main r.h.s. terms:

$$\begin{aligned} y_{it} &= f_1(s_{1,i}) + f_2(s_{2,i}) + f_t(t) + f_{1,2}(s_{1,i}, s_{2,i}) + \\ &\quad f_{1,t}(s_{1,i}, t) + f_{2,t}(s_{2,i}, t) + f_{1,2,t}(s_{1,i}, s_{2,i}, t) + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \qquad [Model\ 7] \end{aligned}$$

Its spatial lag extension (model 8) is:

$$\begin{aligned} y_{it} &= f_1(s_{1,i}) + f_2(s_{2,i}) + f_t(t) + f_{1,2}(s_{1,i}, s_{2,i}) + \\ &\quad f_{1,t}(s_{1,i}, t) + f_{2,t}(s_{2,i}, t) + f_{1,2,t}(s_{1,i}, s_{2,i}, t) + \\ &\quad +\rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \qquad [Model\ 8] \end{aligned}$$

Model 9 relaxes the linearity assumption for the main r.h.s. terms

$$\begin{aligned} y_{it} &= f_1(s_{1,i}) + f_2(s_{2,i}) + f_t(t) + f_{1,2}(s_{1,i}, s_{2,i}) + \\ & f_{1,t}(s_{1,i}, t) + f_{2,t}(s_{2,i}, t) + f_{1,2,t}(s_{1,i}, s_{2,i}, t) + \\ & +\rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \sum_{k=1}^{K} f_k(x_{k,it}) + \varepsilon_{it} \qquad [Model\ 9] \end{aligned}$$

and its spatial lag extension (model 10) is:

$$\begin{aligned} y_{it} &= f_1(s_{1,i}) + f_2(s_{2,i}) + f_t(t) + f_{1,2}(s_{1,i}, s_{2,i}) + \\ & f_{1,t}(s_{1,i}, t) + f_{2,t}(s_{2,i}, t) + f_{1,2,t}(s_{1,i}, s_{2,i}, t) + \\ & +\rho \sum_{j=1}^{N} w_{ij,N} y_{jt} + \sum_{k=1}^{K} f_k(x_{k,it}) + \varepsilon_{it} \qquad [Model\ 10] \end{aligned}$$

As discussed in the main text, the geoadditive terms ($f_1(s_{1,i})$, $f_2(s_{2,i})$, and $f_{1,2}(s_{1,i}, s_{2,i})$), work as control functions to filter spatial trend out of the residuals and transfer it to the mean response in a model specification, while the smooth time trend, $f_t(t)$, and the smooth interaction between space and time - $f_{1,t}(s_{1,i}, t)$, $f_{2,t}(s_{2,i}, t)$, and $f_{1,2,t}(s_{1,i}, s_{2,i}, t)$ - work as control functions to capture the heterogeneous effect of common shocks, thus allowing for cross-section dependence in alternative to cross-sectional averages of the observations. Thus, the inclusion of the ANOVA decomposition of the spatio-temporal trend helps interpret the evidence of significance spatial spillovers as weak cross-dependence net of the effect of common effects (strong dependence).

References

ANSELIN, L. (1988): *Spatial econometrics: methods and models*, vol. 4, Kluwer Academic Pub.

ARAGON, Y., D. HAUGHTON, J. HAUGHTON, E. LECONTE, E. MALIN, A. RUIZ-GAZEN, AND C. THOMAS-AGNAN (2003): "Explaining the pattern of regional unemployment: The case of the Midi-Pyrénées region," *Papers in Regional Science*, 82, 155–174.

BAI, J. AND K. LI (2015): "Dynamic spatial panel data models with common shocks," .

BAILEY, N., S. HOLLY, AND M. H. PESARAN (2016): "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence," *Journal of Applied Econometrics*, 31, 249–280.

BASILE, R., M. DURBÁN, R. MÍNGUEZ, J. M. MONTERO, AND J. MUR (2014): "Modeling regional economic dynamics: spatial dependence, spatial heterogeneity and nonlinearities," *Journal of Economic Dynamics and Control*, 48, 229–245.

BASILE, R., A. GIRARDI, AND M. MANTUANO (2012): "Migration and regional unemployment in Italy," *The Open Urban Studies Journal*, 5, 1–13.

BLANCHARD, O. J., L. F. KATZ, R. E. HALL, AND B. EICHENGREEN (1992): "Regional evolutions," *Brookings papers on economic activity*, 1992, 1–75.

BURRIDGE, P. AND I. GORDON (1981): "Unemployment in the British metropolitan labour areas," *Oxford Economic Papers*, 33, 274–297.

CHUDIK, A., M. H. PESARAN, AND E. TOSETTI (2011): "Weak and strong cross-section dependence and estimation of large panels," *The Econometrics Journal*, 14, C45–C90.

CRACOLICI, M. F., M. CUFFARO, AND P. NIJKAMP (2007): "Geographical distribution of unemployment: An analysis of provincial differences in Italy," *Growth and Change*, 38, 649–670.

DECRESSIN, J. AND A. FATAS (1995): "Regional labor market dynamics in Europe," *European Economic Review*, 39, 1627–1655.

EILERS, P. AND B. MARX (1996): "Flexible Smoothing with B-Splines and Penalties," *Statistical Science*, 11, 89–121.

EILERS, P. H., B. D. MARX, AND M. DURBÁN (2015): "Twenty years of P-splines," *SORT-Statistics and Operations Research Transactions*, 39, 149–186.

ELHORST, J. (2014a): *Spatial Econometrics. From Cross-Sectional Data to Spatial Panels*, SpringerBriefs in Regional Science, Berlin-Heidelberg-New York: Springer.

ELHORST, J. P. (1995): "Unemployment disparities between regions in the European Union," *Convergence and Divergence among European Unions*, 190–200.

——— (2014b): "Spatial panel data models," in *Spatial Econometrics*, Springer, 37–93.

Green, P. and B. Yandell (1985): "Semi-parametric generalized Linear Models," in *Generalized Linear Models*, Berlin: Springer, vol. 32 of *Lecture Notes in Statist.*, 44–55.

Hastie, T. and R. Tibshirani (1990): *Generalized Additive Models*, Monographs on Statistics and Applied Probability, Chapman and Hall.

Henry G. Overman, Diego Puga, H. V. (2002): "Unemployment Clusters across Europe's Regions and Countries," *Economic Policy*, 17, 117–147.

Im, K. S., M. H. Pesaran, and Y. Shin (2003): "Testing for unit roots in heterogeneous panels," *Journal of econometrics*, 115, 53–74.

Kneib, T. and L. Fahrmeir (2006): "Structured additive regression for categorical space-time data: A mixed model approach," *Biometrics*, 62, 109–118.

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

Lee, D.-J. (2010b): "Smoothing mixed models for spatial and spatio-temporal data," Ph.D. thesis, Universidad Carlos III, Leganés.

Lee, D.-J. and M. Durbán (2011): "P-spline ANOVA-type interaction models for spatio-temporal smoothing," *Statistical Modelling*, 11, 49–69.

Lee, D.-J., M. Durban, and P. Eilers (2013): "Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases," *Computational Statistics & Data Analysis*, 61, 22–37.

Lee, L.-f. and J. Yu (2010): "Estimation of spatial autoregressive panel data models with fixed effects," *Journal of Econometrics*, 154, 165–185.

LeSage, J. and K. Pace (2009): *Introduction to Spatial Econometrics*, Boca Raton: CRC Press.

López-Bazo, E., T. D. Barrio, and M. Artís (2005): "Geographical distribution of unemployment in Spain," *Regional Studies*, 39, 305–318.

Lottmann, F. (2012): "Spatial dependencies in German matching functions," *Regional Science and Urban Economics*, 42, 27–41.

Marston, S. T. (1985): "Two views of the geographic distribution of unemployment," *The Quarterly Journal of Economics*, 57–79.

Millo, G. (2014): "Maximum likelihood estimation of spatially and serially correlated panels with random effects," *Computational Statistics and Data Analysis*, 71, 914 – 933.

Molho, I. (1995): "Spatial autocorrelation in British unemployment," *Journal of Regional Science*, 35, 641–658.

Montero, J., R. Mínguez, and M. Durbán (2012): "SAR models with nonparametric spatial trends. A P-spline approach," *Estadística Española*, 54, 89–111.

Partridge, M. D. and D. S. Rickman (1997): "The dispersion of US state unemployment rates: the role of market and non-market equilibrium factors," *Regional Studies*, 31, 593–606.

Patacchini, E. and Y. Zenou (2007): "Spatial dependence in local unemployment rates," *Journal of Economic Geography*, 7, 169–191.

Patterson, H. and R. Thompson (1971): "Recovery of inter-block information when block sizes are unequal," *Biometrika*, 58, 545–554.

Pesaran, M. H. (2004): "General diagnostic tests for cross section dependence in panels," .

——— (2006): "Estimation and inference in large heterogeneous panels with a multifactor error structure," *Econometrica*, 74, 967–1012.

——— (2007): "A simple panel unit root test in the presence of cross-section dependence," *Journal of Applied Econometrics*, 22, 265–312.

——— (2015): "Testing weak cross-sectional dependence in large panels," *Econometric Reviews*, 34, 1089–1117.

Pesaran, M. H. and E. Tosetti (2011): "Large panels with common factors and spatial correlation," *Journal of Econometrics*, 161, 182–202.

Rios, V. (2014): "What Drives Regional Unemployment Disparities in European Regions? A Space-Time Econometric Modeling Approach," .

Rodriguez-Alvarez, M. X., T. Kneib, M. Durban, D.-J. Lee, and P. Eilers (2015): "Fast smoothing parameter separation in multidimensional generalized P-splines: the SAP algorithm," *Statistics and Computing*, 25, 941–957.

Searle, S., G. Casella, and C. McCulloch (1992): *Variance Components*, New York: John Wiley and Sons.

Shi, W. and L.-f. Lee (2016): "A Spatial Panel Data Model with Time Varying Endogenous Weights Matrices and Common Factors," .

Su, L. and S. Jin (2012): "Sieve estimation of panel data models with cross section dependence," *Journal of Econometrics*, 169, 34–47.

Taylor, J. and S. Bradley (1997): "Unemployment in Europe: A comparative analysis of regional disparities in Germany, Italy and the UK," *Kyklos*, 50, 221–245.

Thirlwall, A. P. (1966): "Regional Unemployment As A Cyclical Phenomenon1," *Scottish Journal of Political Economy*, 13, 205–219.

Vega, S. H. and J. P. Elhorst (2016): "A regional unemployment model simultaneously accounting for serial dynamics, weak and strong cross-sectional dependence," Mimeo.

Zeilstra, A. S. and J. P. Elhorst (2014): "Integrated analysis of regional and national unemployment differentials in the European Union," *Regional Studies*, 48, 1739–1755.