THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 02_09

Roberto Patuelli
University of Lugano, Switzerland
The Rimini Centre for Economic Analysis, Rimini, Italy

Daniel A. Griffith
University of Texas at Dallas, USA

Michael Tiefelsdorf
University of Texas at Dallas, USA

Peter Nijkamp

SPATIAL FILTERING AND EIGENVECTOR STABILITY: SPACE-TIME MODELS FOR GERMAN UNEMPLOYMENT DATA

The Rimini Centre for Economic Analysis
Legal address: Via Angherà, 22 – Head office: Via Patara, 3 - 47900 Rimini (RN) – Italy
www.rcfea.org - secretary@rcfea.org

Spatial Filtering and Eigenvector Stability: Space-Time Models for German Unemployment Data

Roberto Patuelli,[1] Daniel A. Griffith,[2] Michael Tiefelsdorf[2] and Peter Nijkamp[3]

[1] Institute for Economic Research (IRE), University of Lugano, Switzerland
The Rimini Centre for Economic Analysis, Italy. Email: roberto.patuelli@usi.ch
[2] School of Economic, Political and Policy Sciences, University of Texas at Dallas, USA
[3] Department of Spatial Economics, VU University Amsterdam, The Netherlands

This version: May 2010

Abstract

Regions, independent of their geographic level of aggregation, are known to be interrelated partly due to their relative locations. Similar economic performance among regions can be attributed to proximity. Consequently, a proper understanding, and accounting, of spatial liaisons is needed in order to effectively forecast regional economic variables. Several spatial econometric techniques are available in the literature, which deal with the spatial autocorrelation in geographically-referenced data. The experiments carried out in this article are concerned with the analysis of the spatial autocorrelation observed for unemployment rates in 439 NUTS-3 German districts. We employ a semi-parametric approach – spatial filtering – in order to uncover spatial patterns that are consistently significant over time. We first provide a brief overview of the spatial filtering method and illustrate the data set. Subsequently, we describe the empirical application carried out: that is, the spatial filtering analysis of regional unemployment rates in Germany. Furthermore, we exploit the resulting spatial filter as an explanatory variable in a panel modelling framework. Additional explanatory variables, such as average daily wages, are used in concurrence with the spatial filter. Our experiments show that the computed spatial filters account for most of the residual spatial autocorrelation in the data.

JEL classification: C33, E24, R12
Keywords: spatial filtering, eigenvectors, Germany, unemployment, GLMM

1. Introduction

Noise and shocks in regional labour markets are not symmetrically distributed in space. Moreover, regions can be seen as small open systems. Consequently, they can be expected to have a great level of reciprocal interaction and to influence each other's economic performance (we can think of, for example, regional spillovers). As a result, spatial matters should be considered of critical importance not only when studying socio-economic phenomena (see, for example, Bockstael 1996, Weinhold 2002), but also because of their implications for policymaking (Lacombe 2004).

Correlation and heterogeneity 'in space' among regions are evident in most countries; a key example is Germany, particularly because of its still-existing East/West (EW) economic divide. This former EW divide is the most relevant spatial structure in defining German regional inequalities. To account for the presence of spatial structures that influence

(positively or negatively) observable economic entities, such as unemployment or trade, implies a call for a rigorous and systematic assessment of their impact and extent. Accordingly, their inclusion in econometric models is necessary in order to correctly assess economic relationships: for example (as observed in this article), the one between unemployment rates and a set of explanatory variables.

Spatial autocorrelation (SAC) (Cliff and Ord 1981) is the correlation, computed among the values of a single georeferenced variable, that is attributable to the geographic proximity of the objects to which the values are attached. The introduction of the SAC concept was a departure from the classical assumption of independence of the observations constituting a single variable. SAC also complements the concept of temporal autocorrelation, which has been extensively studied and dealt with in time-series econometrics. SAC measures, such as Moran's I (the Moran Coefficient) or Geary's c (the Geary Ratio), are used to quantify the nature and degree of the spatial correlation within a variable, or to test the assumption of independence or randomness.

From a statistical analysis viewpoint, spatial correlation patterns are problematic, since they make standard statistics, such as variances and/or ordinary least squares (OLS) estimators, potentially inappropriate. In particular, spatially correlated values of a variable make the estimator of the error variance – in an OLS framework – biased. This is the case when we analyse regional labour market variables such as unemployment. The uneven geographical localization of regional unemployment observed in countries such as Germany (see for example, Bayer and Juessen 2007) may be caused by spatial effects that concern the variable itself. As a result, a linear, non-spatial model, estimated with OLS, has biased regression parameter estimates. In spatial econometrics, 'spatial lag' models are used to accommodate this problem. If spatial effects were to be related to significant unobserved variables – thus causing SAC in a model's error term – the test statistics of the coefficients would be invalid. In this case, a 'spatial error' model is employed in the literature. More general spatial econometric specifications can be attempted, such as the Cliff-Ord-type model (for a taxonomy of spatial econometric models, see, for example, Anselin 1988).

In this article, we investigate the importance of spatial effects in German regional labour markets by developing a single-equation, three-variable regional unemployment model. The focus is not on testing a particular theory or model, but rather on the exploration of spatial patterns, in particular in the case where covariates are included. The aim of the article is twofold: (a) to provide an assessment of how important spatial effects are in explaining German regional unemployment levels and to show that (subsets of) these patterns are consistent over time; and, (b) to develop an econometric model that exploits such consistent spatial patterns in order to improve statistical inference. As an alternative to conventional spatial econometric modelling, we present analyses carried out by means of a semi-parametric 'spatial filtering' technique (described in Griffith 2003), which is based on the decomposition of spatial weights matrices. The additional value added of this estimation approach is that it can be easily replicated for any functional specification (linear and generalized linear models, as well as nonlinear models) and for any number of explanatory variables considered.

The remainder of the article is structured as follows. Section 2 provides a brief overview of the spatial filtering method, while Section 3 illustrates the data set available. Section 4 describes the empirical application carried out: that is, the spatial filtering analysis of regional unemployment rates in Germany, along with the introduction of socio-economic covariates in the spatial filtering framework. Finally, Section 5 offers some summary information and concluding remarks, as well as future research directions.

2. The Spatial Filtering Framework

A wide array of methods, as well as several dedicated 'spatial' econometric procedures (see, for example, Anselin, et al. 2004), for the statistical analysis of georeferenced data are available in the literature. These techniques are useful when analysing regional unemployment data, such as the case study analysed here, and, particularly, when the final aim is to develop forecasting models for some regional scale. Among conventional spatial econometric methods, spatial autoregression (see, among others, Anselin 1988; Griffith 1988) is one method commonly employed. Spatial autoregressive techniques take into account spatial effects using spatial weights matrices (conventionally referred to as **W**). These matrices measure the spatial linkages (dependence) between the values of a georeferenced variable, in terms of geographical contiguity (shared boundaries), distance (between areal unit centroids), or alternative specifications of proximity (for example, social/cultural distance). A general notation for the spatial autoregressive model, which is known as the Cliff-Ord-type model, has been proposed by Anselin (1988):

$$\begin{aligned} y &= \rho \mathbf{W}_1 y + \mathbf{X}\beta + u, \\ u &= \lambda \mathbf{W}_2 u + \varepsilon, \\ \varepsilon &\sim (0, \Omega). \end{aligned} \tag{1}$$

where $\mathbf{W}_1$ and $\mathbf{W}_2$ are two (potentially identical) spatial weights matrices. The first equation posits a spatial lag component, whereas the second equation posits a spatial error term. Models belonging to this generic family can be estimated either by (quasi-)maximum likelihood (ML), as described in Anselin (1988, 2001) and Lee (2004), or by the generalized method of moments (GMM; see Kelejian and Prucha 1998, Kelejian and Prucha 1999, Anselin 2001). These estimators assume that the (spatial) autocorrelation pattern can be combined/concentrated in one or two parameters, and that the spatial weights matrix **W** adequately describes the spatial interdependence.

The estimation approaches mentioned above are based on a number of hypotheses which can sometimes be hard to reconcile with the characteristics of the data. Aside from a few recent cases (see, e.g., Griffith and Paelinck 2009, Lambert, et al. 2010), available spatial econometric techniques are based on the assumption of a linear relation between the dependent and independent variables (i.e., the link function is the identity function), and therefore assume the dependent variable to be normally distributed.

An alternative approach, which allows to overcome this limitation, is the use of spatial filtering techniques, such as the ones described in Griffith (1981) and Haining (1991), Getis and Griffith (2002), and Tiefelsdorf and Griffith (2006). The main advantage of these filtering procedures is that the studied variables (which are – initially – spatially correlated) are split into spatial and non-spatial components, which can then be employed in an OLS modelling framework, as well as in any other generalized linear model (GLM) estimation framework. In addition, filtering out spatially autocorrelated patterns enables one to reduce the stochastic noise normally found in the residuals of standard statistical methods such as OLS. This conversion procedure requires the computation of 'spatial filters.' In this article, we employ the approach developed by Griffith (1996, 2000), which is briefly described here. This approach is preferred, in our case study, to the one by Getis (1990, 1995), which requires variables with a natural origin and a linear model specification. Consequently, rates, percentage changes, and so on, can not be used in the Getis approach.

The spatial filtering technique introduced by Griffith is based on the computational formula of Moran's *I* (MI) statistic.[1] This methodology exploits eigenvector decomposition techniques, which extract *orthogonal* and *uncorrelated* numerical components from a *N* x *N* matrix (Tiefelsdorf and Boots 1995).[2] These components can be seen as independent map patterns, and represent the latent SAC of a georeferenced variable, according to a given spatial weights matrix. They also can be interpreted as redundant information due to spatial interdependencies, in the framework of standard regression analysis.

Formally, these orthogonal components are the computed eigenvectors of the modified spatial weights matrix:

$$(\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N)\mathbf{W}(\mathbf{I} - \mathbf{1}\mathbf{1}^{\mathrm{T}}/N), \tag{2}$$

where **I** is an identity matrix of dimension *N* x *N*, and **1** is an *N* x 1 vector containing 1's. The eigenvectors of the modified matrix are computed, in sequence, to maximize the sequential residual MI values. The first computed eigenvector, E_1, is, therefore, the one whose numerical values generate the largest MI value among all eigenvectors of the modified matrix. Similarly, the second eigenvector, E_2, is the set of numerical values that, again, maximize the MI value, while being orthogonal and uncorrelated with E_1. The process continues until *N* eigenvectors have been computed. This is the complete set of all possible (mutually) orthogonal and uncorrelated map patterns (Getis and Griffith 2002). When employed as regressors, these eigenvectors may function as proxies for missing explanatory variables.

However, employing all *N* eigenvectors in a regression framework is not desirable for reasons of model parsimony and statistical significance, and is altogether impossible in a cross-sectional framework, since the number of explanatory variables would be equal to or greater than the number of observations. A smaller set of 'candidate' eigenvectors can be selected from the *N* eigenvectors, on the basis of their MI values. A pre-specified threshold value can be used for selection screening purposes. Since the eigenvectors are both orthogonal and uncorrelated, a stepwise procedure for linear regression can be used to achieve this end. In this framework, the advantage given by the orthogonality of the eigenvectors is the absence of partial correlations and, therefore, of multicollinearity issues.

The residuals obtained with a stepwise regression constitute the *spatially filtered* component of the georeferenced variable examined. Each eigenvector selected for inclusion is considered to be part of a 'spatial filter' for the dependent variable. The top two eigenvectors computed (E_1 and E_2) often identify map patterns relating to the underlying geocoding reference axes (for example, major North-South and East-West patterns). Eigenvectors with intermediate values of MI display regional map patterns, whereas eigenvectors with smaller values of MI

---

[1] Moran's *I* is the preferred, and oldest, indicator of SAC. It is calculated as:

$$I = \frac{N\sum_i \sum_j w_{i,j}(x_i - \overline{x})(x_j - \overline{x})}{(\sum_i \sum_j w_{i,j})\sum_i (x_i - \overline{x})^2},$$

where: *n* is the number of cases; x_i is the value of variable *X* at location *i*; and $w_{i,j}$ is the cell (*i*, *j*) of the spatial weights matrix **W** considered. Positive autocorrelation $(I > -(N-1)^{-1})$ implies that geographical proximity tends to produce similar values of the variable examined. This is a phenomenon that often is observed in reality, especially in economics. Negative SAC $(I < -(N-1)^{-1})$ is a much rarer phenomenon.

[2] Griffith's spatial filtering techniques may be compared to principal components analysis (PCA), as in fact both methodologies generate orthogonal and uncorrelated new 'variables' that can be employed in a regression analysis framework. Nevertheless, the components derived in PCA have an economic interpretation, because eigenvectors are used to construct linear combinations of attribute variables, whereas spatial filters are linear combinations of the eigenvectors themselves. As such, the latter should be regarded mostly as patterns of independent spatial dimensions.

display local map patterns. A linear combination of the above eigenvectors can be defined as *the* spatial filter for the variable examined.

The preceding spatial filters are computed on the basis of a modified spatial weights matrix. Formally, a spatial weights matrix is a (squared) N x N matrix containing, most often, binary values (0 and 1). A value of 1 for the generic cell (i, j) implies that the two georeferenced objects (for example, regions) i and j are neighbours. The opposite applies for the value 0. The choice of the matrix to be used is relevant with regard to: (a) the definition of proximity; (b) the variable chosen (if necessary) to indicate proximity; and, (c) the coding scheme employed in the calculation of the matrix. These aspects are critical in defining the spatial weights matrix and the related spatial filter. Patuelli et al. (2010) address such issues more extensively, with reference to the German regional labour markets case. They test five different spatial weights matrix specifications.

On the basis of the aforementioned study, we select, for the present article, only one spatial weights matrix, which is based on the rook's contiguity definition (that is, on shared boundaries), and coded according to the so-called C-coding scheme. This scheme is commonly used in spatial statistics and yields a symmetric maxtrix **W.** It is referred to as a 'globally standardizing' scheme, and it tends to emphasize spatial objects with a greater linkage degree (see Tiefelsdorf and Griffith 2007). As a result of the coding scheme choice, our matrix can be expected to present stronger patterns in the inner study area when compared with alternative schemes, such as the common row-standardized W-coding scheme (which will show more 'extreme' values along the edges of a study area).

3. The Data

This article presents results based on the analysis of German unemployment data (unemployment rates). The dataset employed in our experiments consists of a panel of 439 German districts (*kreise*), for which the years from 1996 to 2002 are available, while the level of aggregation of the dataset is NUTS-3. This aggregation level enables a more detailed examination of 'local' unemployment patterns. Data at the NUTS-2 level would have only 41 regions (*Regierungsbezirke*). Alternatively, an intermediate approach is proposed by Kosfeld and Dreger (2006), who carry out a spatial filtering analysis of German regional labour market data using 180 regional labour market areas (defined in Eckey 2001). Such 'functional' areas might be deemed more suitable for regional labour market analyses, if a more specific research question is to be addressed, because of the potential interactions arising between *kreise* and within these areas. In this article, we choose to preserve the richness of information of *kreise*-level data for showcasing our econometric approach.

In addition to our dependent variable (unemployment rates), we employ information at the same aggregation level about: (a) regional daily wages of full-time workers; (b) number of full-time employees; and, (c) working-age population. For the analysis presented in Section 4.3, we employ these three variables over the period from 1994 to and including 2001.

The data are provided by the German Institute for Employment Research (*Institut für Arbeitsmarkt- und Berufsforschung*, IAB), and cover the entire German territory, consisting of 326 districts in the former West Germany, and 113 in the former East Germany. While longer data series are available – for all the aforementioned variables – for West Germany, East German data only became available after reunification. Therefore, if a comprehensive analysis for all of Germany is to be carried out, the time span of our data is limited to the aforementioned periods.

4. The Empirical Application: Computation and Choice of Spatial Filters for German Unemployment Modelling

4.1. Spatial Analysis of Regional Unemployment

Analysing unemployment rates and their geographic distribution is of great interest because, in particular, of their relevance in the determination of welfare policies, and they serve as indicators of socio-economic performance (López-Bazo, et al. 2002, Cracolici, et al. 2007, Patuelli 2007). The importance of spatial matters in regional labour markets, specifically with regard to the case of Germany, has been highlighted recently in various articles (see, for example, Niebuhr 2003, Elhorst, et al. 2007, Longhi and Nijkamp 2007). Kosfeld and Dreger (2006) show that consideration of spatial dependencies is needed in order to correctly estimate unemployment and employment thresholds in their relationship with changes in production levels. Patuelli et al. (2006) report improved performance in the forecast of regional employment variations by means of neural networks upon employing explanatory variables related to geographical contiguity (spatial shift-share). Numerous further examples apply.

Moreover, a number of economic variables measured at the regional scale exhibit strong spatial heterogeneity, mostly because of the coexistence – in the case of Germany – of highly performing areas such as Bavaria, and lowly performing areas such as the former East Germany. If this spatial heterogeneity can be (in part) traced back to the relationships of proximity between regions, it can be modelled according to proximity itself. In particular, a description of the spatial patterns underlying yearly aggregate observations of the same variable may help forecast future values and provide a visual description of time-invariant spatial patterns (see Griffith and Paelinck 2009).

4.2. Spatial Filtering of Unemployment

4.2.1. Computation and Selection of Spatial Filters over Time

The first step in the construction of a spatial filter to be applied to the variable of study is the computation of the eigenvectors of the spatial weights matrix, followed by the choice of a subset of 'candidate' eigenvectors from which selection is made. Candidate eigenvectors are selected on the basis of their MI values and their correlations with the georeferenced regional unemployment data. A minimum threshold value of 0.25 for the statistic MI/max(MI), which roughly corresponds to 5 per cent of the variance being accounted for in the regression of a generic georeferenced variable **Z** on **WZ** (Griffith 2003), has been used in our case to identify the candidate set. When carried out for a (C-coding scheme, see Section 2) rook's definition of contiguity spatial weights matrix, the computation and selection process described here results in the identification of a set of 98 candidate eigenvectors. The highest MI shown by the eigenvectors (that is, the MI of the first selected eigenvector) is 1.24. To date, only the spatial weights matrix **W** (in its modified version, see Equation (2)), *and not the data*, is used for the selection of the eigenvectors.

Once a set of 'candidate' eigenvectors has been selected, its statistical significance, as an explanatory variable for German regional unemployment rates, has to be established. This process is carried out, for each cross-section, by means of a stepwise logistic regression analysis (as in Griffith 2004), estimated in a generalized linear model (GLM) framework using a binomial link function. Because of the GLM estimation, obtained by maximum likelihood (ML), the stepwise regression employed is based on a likelihood measure, namely

the Akaike information criterion (AIC).[3] Similarly, other measures, such as the corrected AIC (McQuarrie and Tsai 1998), which corrects for small samples, could be employed. Because AIC-based stepwise regressions tend to overfit, and consequently to overselect, we carry out a further manual backward elimination of regressors (eigenvectors) on the basis of χ^2 tests based on a 5 per cent significance level. This process is expected to finally result in spatially uncorrelated residuals.[4] Consequently, following the notations of Griffith's (2004) and McCullogh et al. (2008), the resulting logistic model for a particular time t is:

$$E(y_i) = \pi(\mathbf{E}_{i,k}) = \frac{1}{1+e^{-(\alpha+\beta\mathbf{E}_{i,k})}}, \tag{3}$$

where π is the mean function, y_i is the ith element of the dependent variable y, $\mathbf{E}_{i,k}$ is the ith row of the matrix containing the final set of k eigenvectors selected, and α is a constant. For simplicity, year-specific subscripts are omitted here.

We also cope with overdispersion in the data, which is a frequently experienced phenomenon when analysing economic variables. We adjust for so-called extra-binomial variation by employing Williams's (1982) quasi-likelihood overdispersion adjustment. The method postulates 'a source of extra-binomial random variation between observations' (Williams 1982, p. 144), and iteratively estimates the dispersion parameter associated with the data, adjusting the GLM weights accordingly.[5]

The same process is repeated for all years of available data – from 1996 to 2002 for the aforementioned C-coding scheme with the rook's definition of contiguity spatial weights matrix. Consequently, seven sets of 'significant' eigenvectors (one set for each year) are selected. These are used to construct the 'spatial filters' for each year.

Next, we pinpoint a subset of eigenvectors that is common to the years 1996 to 2002; that is, the subset of eigenvectors that has been selected for all years of data, according to the stepwise procedure described. Detailed results about the eigenvectors selected in each year are given in Table 1. Our results show that we found a set of 22 eigenvectors that are significant, as explanatory variables of regional unemployment, over the entire time period considered.

TABLE 1 ABOUT HERE

In terms of statistical relevance, the amount of variance explained by the spatial filtering regressors is fairly consistent over the years (reasonably, unemployment patterns do not change much from year to year). The pseudo-R^2 values found for these analyses are in the 0.734–0.797 range. Plots of the observed and estimated unemployment values are shown in Figure 1, for the years 1996 to 2002, and display a fairly good fit, though a tendency toward underestimation can be observed.

FIGURE 1 ABOUT HERE

[3] The Akaike information criterion (AIC) was proposed by Akaike (1974) and is a goodness-of-fit measure based on the concept of entropy. The AIC takes into account the trade-off between model complexity and model fit. It is calculated as: $AIC = 2k - 2\ln(L)$, where k is the number of estimated parameters and L is the likelihood function of the estimated model.

[4] Spatial uncorrelatedness of the residuals can be assured in particular by minimizing MI instead of a likelihood measure or with a mixed strategy (see Griffith 2004).

[5] Because of the GLM weighting, pure orthogonality of the eigenvectors is lost (which may in fact create computational problems). We therefore consider the eigenvectors as quasi-orthogonal when employed in a GLM estimation framework.

As mentioned in Section 2, the constructed spatial filters, which are the linear combinations of the selected eigenvectors using their logistic regression estimated coefficients, can be interpreted not only as potential explanatory variables substituting for missing ones, but also as map patterns. A graphical visualization of the spatial filters uncovered by our analysis provides an example of the map features embedded in the eigenvectors' values. Figure 2 shows the four eigenvectors with the largest MI values computed for the employed contiguity matrix, and that are common to all the years examined (E_2, E_4, E_6 and E_7). As noted previously, the first two eigenvectors for contiguity matrices usually show underlying geocoding reference axis patterns. Spatial filter (a) (E_2) in Figure 2 seems, in fact, to be characterized by a North-South pattern (a 'global' pattern). As we observe the subsequent spatial filter components (b, c, and d), the geographic patterns mapped relate to characteristics of smaller geographical scale, showing patterns that can be categorized first as 'regional', and then as 'local'. Although they may contain some common map patterns (for example, North-South and East-West patterns), spatial filters computed with different spatial weights matrices will vary to some degree.[6] In addition, results from the application of a 'queen' contiguity definition are not considered here, since the two specifications of adjacency differ by only 25 neighbour links.

FIGURE 2 ABOUT HERE

4.2.2. A Spatial Structured Random Effects Panel Model for German Unemployment

The preceding section focuses on computing and selecting sets of eigenvectors that are commonly significant for all the years examined (1996–2002). In this section, we exploit these findings by estimating a spatially structured random effects panel model in order to evaluate the explanatory power of a time-invariant spatial filter. We employ a generalized linear mixed model (GLMM), which we develop for the case of the C-coding scheme with the rook's definition of contiguity weights matrix illustrated above. The 22 common selected eigenvectors (see preceding section) are entered as regressors in a generalized linear model (GLM) with a binomial link function, together with a normally-distributed random-effects intercept variable, in order to handle temporal correlation. Conditionally on the random effects, a standard GLM is indeed estimated (Venables and Ripley 2002). A similar analysis is presented in Griffith (2008) for space-time agricultural production.

In a GLMM, the intercept (of the linear predictor) is specified as a geographically-varying random variable, which accounts for the serial correlation in short time series such as that employed in our case study. This random effects intercept also supports inferences beyond the employed surface partitioning and set of points in time.

Technically, a GLMM can be viewed as a non-linear model whose non-linearity is given by the link function chosen (the logit in our binomial case), and its variance is a function of the mean (Venables and Ripley 2002). We fit our GLMM by means of a penalized quasi-likelihood approach (Breslow and Clayton 1993), which makes use of quadratic Taylor expansions (Wolfinger and O'Connell 1993, Evans and Swartz 2000). The chosen estimation method also accounts for extra-binomial variation – similar to Williams's (1982) approach for the year-by-year analyses – by estimating dispersion and adjusting the significance levels

[6] An in-depth analysis of the issues related to the choice of a coding scheme, particularly in view of the type of data patterns that a spatial analyst wants to emphasize (different coding schemes accentuate different kinds of patterns) goes beyond the scope of this article; however, an interesting treatment can be found in Tiefelsdorf et al. (1999).

accordingly. Table 2 presents summary results regarding the spatial autocorrelation accounted for by this model.

TABLE 2 ABOUT HERE

The statistical results presented in Table 2 show that the spatial filter accounts for a large share of SAC, though not all of it.[7] A graphical visualization of the spatial filter appears in Figure 3. In terms of goodness-of-fit, the model has a pseudo-R^2 of 0.922 (year-by-year pseudo-R^2s are given in Table 2), and all the eigenvectors employed are significant.

FIGURE 3 ABOUT HERE

While the estimation described above provides comforting results, a further level of analysis is necessary in order to carry out more detailed experiments about the dynamics of unemployment patterns. In this regard, the limitation of the experiments presented above is that they refer to an unemployment autoregression. Therefore, we propose the utilization of additional explanatory variables in the model. The joint employment of spatial filters and socio-economic explanatory variables involves further attention to the mechanics of spatial filtering. Eigenvectors that are significant both to the explained and to the explanatory variable(s) also imply filtering of the latter.

4.3. Inclusion of Explanatory Variables in Spatial Filtering

4.3.1. Selection of the Spatial Filters for the Unemployment Models

The next step in our analysis is to further the preceding spatial statistical treatment by including covariates with socio-economic meaning. By doing so, we fulfil two main objectives: (a) we go beyond the limit of the previous analyses, which account only for the purely geographical distribution of the variable concerned (German unemployment rates); and, (b) we fully exploit the potential of spatial filtering, as we compute new spatial filters. This procedure allows us to obtain spatially adjusted estimates of the regression parameters relating to the real covariates employed.

To include all the factors that may determine regional unemployment differentials as well as the observed spatial patterns in an econometric model is a demanding task. These factors may be socio-economic or locational: spillover effects, as well as rigidities in labour markets (highly unionized workers) or in mobility (high real estate prices). Consequently, an analyst may choose to focus on a few main explanatory variables relating to labour demand and supply, such as employment, population, or wages, in order to explain – as in our case study – unemployment variations. The effects of the remaining (excluded) factors – in particular if related to location – might identify a set of spatial structures. With this objective in mind, we include in our analysis three explanatory variables: (a) the number of full-time employed individuals; (b) average daily wages of full-time employees; and, (c) working age population (age 15-65). All data are available for all German regions and at the same level of disaggregation as the dependent variable (that is, NUTS-3).

[7] It should be noted that there is no spatial autocorrelation indicator available for the residuals of a logistic regression. Some work has recently been carried out with regard to Poisson models or more generally non-normal data (Jacqmin-Gadda, et al. 1997, Lin and Zhang 2007, Griffith 2010), but further investigation of the new estimators proposed has been called for (Bivand, et al. 2008). In this article, we stick to standard randomization-based MI tests, though acknowledging that their adequacy in this context may be questioned. Also, the stand-alone MI test is applied rather than the testing procedure based on linear regression residuals, because of the GLMM estimation.

We develop a simple three-variable unemployment model, as the focus is not on testing a particular theory or model, but rather on exploring the impact and potential of the spatial filtering technique proposed in the case when covariates are included. The revised model estimated is therefore:

$$unempl_{it} = \Delta wage_{i,t-1} + \Delta empl_{i,t-1} + \Delta pop_{i,t-1} + \varepsilon_{it}, \tag{4}$$

where $unempl_{it}$ is the unemployment rate of region i at time t, $\Delta wage_{i,t-1}$ is the variation of wages in the same region in the period ($t-2$, $t-1$), $\Delta empl_{i,t-1}$ and $\Delta pop_{i,t-1}$ respectively are the corresponding variations in full-time employment and working-age population for the same period, and ε_{it} is the error term. Longer lags, in particular with regard to population variations, could be used (see, for example, Carlino and Mills 1987), but are not considered in our experiments because of the limited period of data availability.

In our model, the wages and employment variables refer to the labour demand factors that influence unemployment. Meanwhile, the population variable can be seen as an indicator of both labour supply and demand factors, because it accounts for several demographic aspects. With regard to labour supply, natural growth and immigration may lead to changes in the age structure of the workers' pool, where a younger working population has been found to experience more persistent unemployment (Elhorst 1995). However, the dataset analysed in this article is too short to expect such a significant effect to be detectable. Migration, instead, may have a neutral effect, if migrants fill vacancies left unfilled, or if they do not join the labour force. With regard to the labour demand effect of changes in population levels, a positive net immigration may induce higher productivity or higher investments if new/higher skills are introduced in a labour market, or simply higher levels of production to satisfy the increased population. As a result of the conflicting effects described, the expected sign for the effect of population change on unemployment is ambiguous.

The expected signs for changes in wages and employment are more straightforward. The negative effect of wage increases on labour demand is expected to lead to increases in the unemployment rates, implying a positive expected sign. The inverse relationship between employment (an indicator of labour demand) and unemployment implies a negative expected sign.

Clearly, the model could be estimated in terms of unemployment rate *variations*. Although this solution would be more suitable in economic reasoning terms (relating variations in the explanatory variables to variations in the dependent variable), we choose to proceed, as in Section 4.2.1, with the analysis of unemployment rates. As a result, the spatial filters obtained for this model specification are comparable to the ones found for the previous specification presented in the article. The differences between the new and the old spatial filters may result from inclusion in the model of substantive covariates, for which the spatial filters previously selected were, in part, a surrogate. With the inclusion of spatial filter components (eigenvectors of the modified spatial weights matrix), Equation (4) becomes:

$$unempl_{it} = \Delta wage_{i,t-1} + \Delta empl_{i,t-1} + \Delta pop_{i,t-1} + sf_i + \varepsilon_{it}, \tag{5}$$

where sf_i is the linear combination – for region i – of the selected spatial filter components.

The first step in estimating Equation (5) is to find the appropriate spatial filters for this empirical case. Again, we employ the C-coding scheme with the rook's definition of contiguity spatial weights matrix **W** used in Section 4.2.1. We start from the set of 98 candidate eigenvectors, and follow a spatial filter selection procedure similar to the previously employed one: a stepwise logistic regression of Equation (5), where the socio-economic

covariates are the initial regressors included (and therefore cannot be dropped in the stepwise selection), and the subsequent inclusion of single eigenvectors as additional regressors is decided on the basis of the model's Akaike information criterion (AIC) during the stepwise procedure, and on the basis of χ^2 tests in the manual backward elimination subsequently carried out.

For each year (1996–2002), we compute the spatial filter concerning jointly the dependent and the independent variables. As shown in Table 3, we find spatial filters comprising between 32 and 38 eigenvectors each. The pseudo-R^2 values of the models are significantly higher than those found in Section 4.2.1: they range from 0.820 to 0.885. The improved statistical power of the analysis (with respect to the preceding range: 0.734–0.797) is a reasonable finding, since we introduced 'real' explanatory variables. With regard to the spatial filters, the set of eigenvectors common to all years that we find is slightly smaller (21 components) than the previously found set (22 components), as the inclusion of the covariates 'eats up' a share of the variance to be accounted for in the data.

TABLE 3 ABOUT HERE

With regard to the explanatory variables employed (wages/employment/population), we observe, in Table 4, that:

- The related regression parameters are mostly significant. While a comparison model comprising *only* wages, employment and population variations (not shown here) gives just three non-significant parameters, the significance levels of the spatial filter model are still satisfactory, as they generally confirm the relevance of the variables.
- The signs of the explanatory variables are as expected, and constant over the years (aside from the case of wages in 2002). However, the stable result of a negative parameter for the population growth variable appears to suggest a dominance of demand factors with regard to demographic change, and surely deserves further investigation in order to be fully interpreted in this context.

TABLE 4 ABOUT HERE

The results presented in Table 3 and Table 4 summarize the statistical power of our spatial filter-enhanced models. We next present the results of the models with regard to SAC. Table 5 summarizes our empirical findings with respect to model residuals. According to the tabulated results, if our naïve unemployment model is carried out *without* including the spatial filter components, the regression residuals' SAC obtained by the logistic regression for each year range between 0.363 and 0.722, implying rather strong SAC. The re-computation of the models *with* the inclusion of the spatial filters decreases SAC, in the range from –0.027 to 0.017. Further, if we re-run our logistic regression models by including, together with the covariates, only the set of common eigenvectors for 1996–2002, we find residual SAC varying between 0.170 and 0.240, implying a loss in the SAC abatement power of about 0.20 between the full yearly spatial filters and the time-invariant spatial filter. This is the compromise we accept by selecting a common spatial filter for the entire dataset.

TABLE 5 ABOUT HERE

Given the above results, the next necessary step is to exploit the time-invariant spatial filter found in Table 3 in a wider framework.

4.3.2. A Spatial Filtering Panel Model for German Unemployment

The analyses carried out on the joint inclusion, in a logistic regression framework, of our socio-economic explanatory variables and spatial filter components show that acceptably low levels of SAC can be reached by replacing the spatial filters separately computed for each year with one spatial filter common to all years. The advantage of employing this reduced set of eigenvectors (see Table 3) is that it can be employed in the GLMM framework previously outlined in Section 4.2.2.

As in our first GLMM approach, the German regional unemployment rates are the dependent variable, while our three economic covariates (wages, employment and population), as well as the spatial filter selected in Section 4.3.1, serve as explanatory variables. The results of our new GLMM estimation are presented in Table 6, while a graphical visualization of the emerging spatial filter can be seen in Figure 4.

TABLE 6 ABOUT HERE

FIGURE 4 ABOUT HERE

Not surprisingly, the map visualization of the spatial filter emerging from our GLMM estimation outlines a clear contrast between the former West and East Germany, as well as a clear evidence of higher (unexplained) unemployment along the German borders, which may be due to cross-border commuting. This finding was to be expected, since our analysis is concerned with the levels of regional unemployment, (rather than with variations in it). As a result, the spatial filter takes into account the stock of unemployment that is not explained by recent labour market trends (that is, the stock acquired prior to the time period examined). Consequently, it is not surprising that the spatial filter shown in Figure 4 resembles closely the one of Figure 3 (the autoregression description of the average regional unemployment rates obtained without the use of explanatory variables).

With regard to estimation of the model parameters, Table 6 shows that the employed covariates are statistically significant, with regard to the two economic variables (employment and wages) and the spatial filter components. The signs of the former are as expected and consistent with the findings of the separate year-by-year analyses. Meanwhile, the non-significant coefficient for population growth suggests that further investigation is needed in order to correctly include demographic aspects in the model specification, and that labour demand factors to some extent may counterbalance the expected labour supply effect. This finding, for example, appears to be consistent with the one by Oud et al. (2008), who carry out a continuous-time spatial-dependence panel analysis of German regional unemployment and population development. The share of variance explained by the GLMM, in terms of pseudo-R^2, is reported in Table 7.

TABLE 7 ABOUT HERE

Results for the GLMM estimation now can be compared with those of selected benchmark models. For purposes of comparison, we estimate three alternative models, each employing, as explanatory variables, the growth rates of wages, employment and population:

- a pooled OLS regression;
- a spatial lag panel model;
- a spatial lag panel model with time fixed effects.

A spatial lag model is computed as follows:

$$\begin{aligned} y &= \rho \mathbf{W} y + \mathbf{X}\beta + u, \\ u &\sim (0, \Omega), \end{aligned} \tag{6}$$

where the values assumed by the dependent variable y are explained by spatial autoregressive values defined according to a row-standardized spatial weights matrix **W**, and by the values of the explanatory variables. We choose to compute a spatial lag panel model on the basis of a set of specification search LM tests (Anselin 1988, Anselin 2002), carried out year-by-year,[8] a summary of which is presented in Table 8. The spatial-lag time-fixed-effects specification is an expansion of the spatial lag model illustrated above, in that it also employs year dummies to take into account temporal shocks. Results for the three models appear in Table 9.

TABLE 8 ABOUT HERE

Results reported in Table 9 indicate that the fittings of the three benchmark model specifications are poorer than that for the GLMM specification (which has an average pseudo-R^2 of 0.945), mostly because its random effects term is a surrogate for various model deficiencies. The signs of the covariates were found to be consistent with those previously observed (Table 4 and Table 6).

TABLE 9 ABOUT HERE

Given these results, we can conclude that the GLMM estimation provides a satisfactory statistical description, showing higher fitting than the benchmark models and providing parameter estimates consistent with the expectations. However, implementing more suitable comparison models, which mirror the serial correlation captured by the GLMM, as well as the geographically-varying effect of the GLMM intercept, is more desirable. This need for further computations is reflected in the conclusions of this article.

5. Conclusions

In this article we presented an analysis of German regional unemployment data by means of 'spatial filtering' techniques. The analysis enabled us to uncover underlying spatial structures by selecting sets of 'spatial filters' that significantly explain geographic variations in a given dataset. In addition, we observed subsets of spatial filters that (partially) define spatial structure over time. The spatial filters selected in this case are the ones that were common to the analyses carried out for each year in the 1996–2002 period.

If shown as graphical visualizations, the spatial filters found in our analyses provide certain indications of the geographical distribution of unemployment trends. Using Figure 2 as an example, map (a) can be interpreted as the visualization of a North-South divide, while maps (b), (c), and (d) seem to distinguish particular areas from the rest of the country. Additional eigenvectors (not shown here) show smaller scale patterns of the regional/local spatial dependence structure.

The initial analysis then was repeated, in Section 4.3, by introducing into the initial spatial statistical framework three explanatory variables with socio-economic meaning: wages,

[8] Ideally, single specification tests could be carried out for the entire time range (see, for example, Bouayad-Agha and Védrine 2010, p. 217, Table 2). Because of software restrictions, we resorted to cross-sectional diagnostics.

employment and population. We constructed new sets of spatial filters, which, in this case, are the result not only of the analysis of the dependent variable, but also of the covariates. We show, in this case as well, the possibility to select a time-invariant spatial filter subset that accounts for spatial structures in all the years of data analysed. Subsequently, a GLMM, estimated via a penalized quasi-likelihood procedure, was used in order to model unemployment rates by means of the covariates and the spatial filter components *jointly*. We show that the GLMM estimation provides a high level of statistical reliability, as well as parameter estimates consistent with the literature.

The results obtained in this article illustrate spatial structure underlying georeferenced unemployment data. Nevertheless, future research along these lines is needed. On the empirical side, a proxy of spatial economic linkages could be employed as an alternative to a spatial weights matrix based on a contiguity rule. Also, the analysis of unemployment levels has its counterpart in that of employment growth rates. Future investigations should address this objective. Furthermore, the analysis of unemployment levels should be more formally concerned with the joint analysis of factors pertaining to labour supply and demand. While the introduction in this analysis of three covariates is a first step, future investigations need to address this issue, for example adopting a full regional labour markets model, such as the one of Blanchard and Katz (1992). On the methodological side, a comparison of the performance of the spatial statistical approach with other conventional spatial econometrics methods, as well as with non-linear approaches, such as neural networks, is desirable. Mixed neural networks/spatial filtering approaches also should be tested. Policy-wise, more in-depth examination of the spatially-filtered GLMM residuals resulting from the analysis should be carried out, in order to fully grasp the benefits of the methodology applied.

The analyses presented in this article have highlighted the relevance – and most importantly the persistence – of spatial structures in German regional unemployment rates (and, we could generalize, in the corresponding labour markets). Our finding of common spatial filters for different years is a reflection of this general stability.[9] Consequently, the spatial filtering technique employed here is one of several useful tools that can be deployed in the analysis of regional disparities.

Finally, a detailed spatial filter analysis of the individual covariates used here also is desirable, for comparison purposes and to attain a better understanding of the role played by spatial structure.

Acknowledgements

We are grateful to two anonymous referees, as well as to session participants at the 53rd Annual North American Meetings of the Regional Science Association International (Toronto) and at the 47th Congress of the European Regional Science Association (Paris), for useful comments. We wish to thank Uwe Blien and the German Institute for Employment Research (IAB) for providing the data. The code written for this article in the statistical software `R` is available on request from the first author.

References

Akaike, Hirotsugu. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* 19, no. 6 (1974): 716-23.

[9] Alternatively, the stability of the proposed spatial filter could be investigated by carrying out poolability tests, for example in the fashion of the Chow test (Chow 1960), which however applies to linear models.

Anselin, Luc. "Spatial Econometrics." In *A Companion to Theoretical Econometrics*, edited by Badi H. Baltagi, 310-30. Malden: Blackwell, 2001.

———. *Spatial Econometrics: Methods and Models*. Dordrecht Boston: Kluwer Academic Publishers, 1988.

———. "Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models." *Agricultural Economics* 27, no. 3 (2002): 247-67.

Anselin, Luc, Raymond J. G. M. Florax, and S. J. Rey, eds. *Advances in Spatial Econometrics*, *Advances in Spatial Science*. Berlin Heidelberg New York: Springer, 2004.

Bayer, C, and F Juessen. "Convergence in West German Regional Unemployment Rates." *The German Economic Review* 8, no. 4 (2007): 510-35.

Bivand, Roger S., Edzer J. Pebesma, and Virgilio Gómez-Rubio. *Applied Spatial Data Analysis with R*. New York London: Springer, 2008.

Blanchard, Olivier J., and Lawrence F. Katz. "Regional Evolutions." *Brookings Papers on Economic Activity* 1 (1992): 1-75.

Bockstael, Nancy Elaine. "Economics and Ecological Modeling: The Importance of a Spatial Perspective." *American Journal of Agricultural Economics* 78, no. 5 (1996): 1168-80.

Bouayad-Agha, Salima, and Lionel Védrine. "Estimation Strategies for a Spatial Dynamic Panel Using Gmm. A New Approach to the Convergence Issue of European Regions." *Spatial Economic Analysis* 5, no. 2 (2010): 205-27.

Breslow, N. E., and D. G. Clayton. "Approximate Inference in Generalized Linear Mixed Models." *Journal of the American Statistical Association* 88, no. 421 (1993): 9-25.

Carlino, G.A., and E.S. Mills. "The Determinants of County Growth." *Journal of Regional Science* 27, no. 1 (1987): 39-54.

Chow, Gregory C. "Tests of Equality between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28, no. 3 (1960): 591-605.

Cliff, A.D., and J.K. Ord. *Spatial Processes: Models & Applications*. London: Pion, 1981.

Cracolici, Maria Francesca, Miranda Cuffaro, and Peter Nijkamp. "Geographical Distribution of Unemployment: An Analysis of Provincial Differences in Italy." *Growth and Change* 38, no. 4 (2007): 649-70.

Eckey, H.-F. "Der Wirtschafliche Entwicklungsstand in Den Regionen Des Vereinigten Deuschland." Kassel: University of Kassel, Institute of Economics, 2001.

Elhorst, J. Paul. "Unemployment Disparities between Regions in the European Union." In *Convergence and Divergence among European Unions*, edited by Harvey W. Armstrong and Roger W. Vickerman, 209-21. London: Pion, 1995.

Elhorst, J. Paul, Uwe Blien, and Katja Wolf. "New Evidence on the Wage Curve: A Spatial Panel Approach." *International Regional Science Review* 30, no. 2 (2007): 173-91.

Evans, M., and T. Swartz. *Approximating Integrals Via Monte Carlo and Deterministic Methods*. Oxford: Oxford University Press, 2000.

Getis, Arthur. "Screening for Spatial Dependence in Regression Analysis." *Papers of the Regional Science Association* 69 (1990): 69-81.

———. "Spatial Filtering in a Regression Framework: Examples Using Data on Urban Crime, Regional Inequality, and Government Expenditures." In *New Directions in Spatial Econometrics*, edited by Luc Anselin and Raymond J. G. M. Florax, 172-85. Heidelberg: Springer, 1995.

Getis, Authur, and Daniel A. Griffith. "Comparative Spatial Filtering in Regression Analysis." *Geographical Analysis* 34, no. 2 (2002): 130-40.

Griffith, Daniel A. "A Comparison of Four Model Specifications for Describing Small Heterogeneous Space-Time Datasets: Sugar Cane Production in Puerto Rico, 1958/59-1973/74." *Papers in Regional Science* 87, no. 3 (2008): 341-55.

———. "A Linear Regression Solution to the Spatial Autocorrelation Problem." *Journal of Geographical Systems* 2, no. 2 (2000): 141-56.
———. "The Moran Coefficient for Non-Normal Data." *Journal of Statistical Planning and Inference* (forthcoming) (2010).
———. "Spatial Autocorrelation and Eigenfunctions of the Geographic Weights Matrix Accompanying Geo-Referenced Data." *The Canadian Geographer* 40 (1996): 351-67.
———. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*, *Advances in Spatial Science.* Berlin New York: Springer, 2003.
———. "A Spatial Filtering Specification for the Autologistic Model." *Environment and Planning A* 36, no. 10 (2004): 1791-811.
Griffith, Daniel A., and Jean H. Paelinck. "Specifying a Joint Space- and Time-Lag Using a Bivariate Poisson Distribution." *Journal of Geographical Systems* 11, no. 1 (2009): 23-36.
Haining, R. "Bivariate Correlation and Spatial Data." *Geographical Analysis* 23 (1991): 210-27.
Jacqmin-Gadda, Hélène, Daniel Commenges, Chakib Nejjari, and Jean-François Dartigues. "Tests of Geographical Correlation with Adjustment for Explanatory Variables: An Application to Dyspnoea in the Elderly." *Statistics in Medicine* 16, no. 11 (1997): 1283-97.
Kelejian, Harry H, and Ingmar R Prucha. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40 (1999): 509-33.
———. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." *Journal of Real Estate Finance and Economics* 17, no. 1 (1998): 99-121.
Kosfeld, Reinhold, and Christian Dreger. "Thresholds for Employment and Unemployment. A Spatial Analysis of German Regional Labour Markets 1999-2000." *Papers in Regional Science* 85, no. 4 (2006): 523-42.
Lacombe, D.J. "Does Econometric Methodology Matter? An Analysis of Public Policy Using Spatial Econometric Techniques." *Geographical Analysis* 36, no. 2 (2004): 105-18.
Lambert, Dayton M., Jason P. Brown, and Raymond J. G. M. Florax. "A Two-Step Estimator for a Spatial Lag Model of Counts: Theory, Small Sample Performance and an Application." *Regional Science and Urban Economics* (forthcoming) (2010).
Lee, Lung-Fei. "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models." *Econometrica* 72 (2004): 1899-925.
Lin, Ge, and Tonglin Zhang. "Loglinear Residual Tests of Moran's *I* Autocorrelation and Their Applications to Kentucky Breast Cancer Data." *Geographical Analysis* 39, no. 3 (2007): 293-310.
Longhi, Simonetta, and Peter Nijkamp. "Forecasting Regional Market Developments under Spatial Autocorrelation." *International Regional Science Review* 30, no. 2 (2007): 100-51.
López-Bazo, Enrique, Tomás del Barrio, and Manuel Artis. "The Regional Distribution of Spanish Unemployment: A Spatial Analysis." *Papers in Regional Science* 81, no. 3 (2002): 365-89.
McCullogh, Charles E., Shayle R. Searle, and John M. Neuhaus. *Generalized, Linear, and Mixed Models*. Hoboken: John Wiley & Sons, 2008.
McQuarrie, A.D.R., and C.-L. Tsai. *Regression and Time Series Model Selection*. Singapore River Edge: World Scientific, 1998.
Niebuhr, Annekatrin. "Spatial Interaction and Regional Unemployment in Europe." In *European Journal of Spatial Development*, 2003.

Oud, J., H. Folmer, Roberto Patuelli, and Peter Nijkamp. "A Spatial-Dependence Continuous-Time Model for Regional Unemployment Analysis in Germany." Lugano: University of Lugano, 2008.

Patuelli, Roberto. "Regional Labour Markets in Germany: Statistical Analysis of Spatio-Temporal Disparities and Network Structures." Ph.D. Thesis, VU University Amsterdam, 2007.

Patuelli, Roberto, Daniel A. Griffith, Michael Tiefelsdorf, and Peter Nijkamp. "Spatial Filtering Methods for Tracing Space-Time Developments in an Open Regional System: Experiments with German Unemployment Data." In *Societies in Motion: Regional Development, Industrial Innovation and Spatial Mobility*, edited by Amnon Frenkel, Peter Nijkamp and Philip McCann. Cheltenham Northampton: Edward Elgar, 2010.

Patuelli, Roberto, Aura Reggiani, Peter Nijkamp, and Uwe Blien. "New Neural Network Methods for Forecasting Regional Employment: An Analysis of German Labour Markets." *Spatial Economic Analysis* 1, no. 1 (2006): 7-30.

Tiefelsdorf, Michael, and B. Boots. "The Exact Distribution of Moran's *I*." *Environment and Planning A* 27 (1995): 985-99.

Tiefelsdorf, Michael, and Daniel A. Griffith. "Semiparametric Filtering of Spatial Autocorrelation: The Eigenvector Approach." *Environment and Planning A* 39, no. 5 (2007): 1193-221.

Tiefelsdorf, Michael, Daniel A. Griffith, and B. N. Boots. "A Variance Stabilizing Coding Scheme for Spatial Link Matrices." *Environment and Planning A* 31 (1999): 165-80.

Venables, W. N., and B. D. Ripley. *Modern Applied Statistics with S*. 4th ed. New York: Springer, 2002.

Weinhold, Diana. "The Importance of Trade and Geography in the Pattern of Spatial Dependence of Growth Rates." *Review of Development Economics* 6, no. 3 (2002): 369-82.

Williams, D.A. "Extra-Binomial Variation in Logistic Linear Models." *Applied Statistics* 31, no. 2 (1982): 144-48.

Wolfinger, R., and M. O'Connell. "Generalized Linear Mixed Models: A Pseudo-Likelihood Approach." *Journal of Statistical Computation and Simulation* 48, no. 3-4 (1993): 233-43.

Table 1 – Selected eigenvectors, 1996–2002: year-specific and common

Year	# of eigenvecs	Year-specific eigenvectors			Common eigenvectors			Pseudo-R^2	AIC
		Global	Regional	Local	Global	Regional	Local		
*Eigenvectors extracted from matrix **W** (C-coding scheme with the rook's definition of contiguity) (98 candidate eigenvectors)*									
1996	34	E_1, E_3, E_5	E_{19}, E_{21}, E_{26}, E_{27}, E_{33}, E_{41}, E_{43}, E_{44}	E_{81}	E_2, E_4	E_6, E_7, E_9, E_{11}, E_{15}, E_{16}, E_{17}, E_{18}, E_{20}, E_{24}, E_{25}, E_{28}, E_{30}, E_{34}, E_{38}, E_{39}, E_{51}, E_{52}, E_{60}	E_{74}	0.752	1032.9
1997	34	E_3	E_{12}, E_{19}, E_{21}, E_{26}, E_{27}, E_{41}, E_{42}, E_{44}, E_{48}, E_{65}	E_{81}				0.777	997.5
1998	33	E_3	E_{13}, E_{19}, E_{21}, E_{26}, E_{27}, E_{42}, E_{48}, E_{65}	E_{81}				0.754	913.4
1999	29	E_3	E_8, E_{21}, E_{26}, E_{42}, E_{55}, E_{65}					0.734	868.6
2000	33	E_3	E_8, E_{13}, E_{19}, E_{21}, E_{26}, E_{42}, E_{44}, E_{55}, E_{65}, E_{66}					0.738	843.8
2001	35	E_3	E_8, E_{12}, E_{13}, E_{19}, E_{21}, E_{42}, E_{43}, E_{55}, E_{56}, E_{65}, E_{66}	E_{81}				0.784	825.2
2002	37		E_8, E_{12}, E_{13}, E_{19}, E_{23}, E_{31}, E_{36}, E_{41}, E_{42}, E_{43}, E_{55}, E_{56}, E_{65}	E_{81}				0.797	862.0

Table 2 – Spatial autocorrelation measures for German unemployment, based upon the rook's definition of contiguity (C-coding scheme) spatial weights matrix

Year	Observed values	Spatial filter residuals	Fitted values
	MI	MI	*Pseudo*-R^2
1996	0.836	0.279	0.943
1997	0.873	0.245	0.962
1998	0.859	0.207	0.965
1999	0.860	0.194	0.956
2000	0.891	0.264	0.949
2001	0.897	0.306	0.934
2002	0.903	0.292	0.922
Spatial filter	1.120	–	–

Table 3 – Selected eigenvectors, 1996–2002: year-specific and common

Year	# of eigenvecs	Year-specific eigenvectors			Common eigenvectors			Pseudo-R^2	AIC
		Global	Regional	Local	Global	Regional	Local		
Eigenvectors extracted from the rook matrix (C-coding) (98 candidate eigenvectors)									
1996	38	E_3, E_4, E_5	E_{19}, E_{21}, E_{27}, E_{33}, E_{34}, E_{41}, E_{43}, E_{44}, E_{51}, E_{54}	E_{68}, E_{72}, E_{81}, E_{83}	E_2	E_6, E_7, E_8, E_9, E_{11}, E_{15}, E_{16}, E_{17}, E_{18}, E_{20}, E_{24}, E_{25}, E_{26}, E_{28}, E_{30}, E_{38}, E_{39}, E_{42}, E_{60}	E_{74}	0.820	1232.9
1997	37	E_3, E_4, E_5	E_{13}, E_{19}, E_{21}, E_{33}, E_{34}, E_{43}, E_{44}, E_{51}, E_{52}, E_{54}, E_{65}	E_{79}				0.846	1244.5
1998	36	E_3, E_4, E_5	E_{13}, E_{19}, E_{21}, E_{32}, E_{34}, E_{51}, E_{52}, E_{65}, E_{66}	E_{68}, E_{72}, E_{81}				0.831	1123.5
1999	35	E_3, E_4	E_{10}, E_{14}, E_{21}, E_{34}, E_{40}, E_{41}, E_{43}, E_{51}, E_{52}, E_{55}, E_{65}	E_{81}				0.838	1131.1
2000	37	E_3, E_4	E_{10}, E_{13}, E_{14}, E_{19}, E_{21}, E_{33}, E_{34}, E_{43}, E_{44}, E_{52}, E_{65}	E_{68}, E_{76}, E_{81}				0.854	1081.0
2001	34	E_3	E_{10}, E_{13}, E_{19}, E_{23}, E_{33}, E_{34}, E_{40}, E_{44}, E_{55}, E_{65}	E_{76}, E_{81}				0.885	1132.1
2002	32	E_4	E_{10}, E_{23}, E_{31}, E_{44}, E_{54}, E_{55}, E_{64}, E_{65}	E_{72}, E_{87}				0.882	1099.8

Table 4 – Sign and statistical significance of the socio-economic covariates, 1996–2002

	1996	1997	1998	1999	2000	2001	2002
Wages	$+^{**}$	$+^{***}$	$+^{***}$	+	+	+	$-^{*}$
Employment	–	$-^{***}$	–	$-^{***}$	–	$-^{***}$	$-^{***}$
Population	$-^{***}$	$-^{***}$	$-^{***}$	$-^{***}$	$-^{***}$	$-^{***}$	$-^{***}$

*** Significant at the 1 per cent level.
** Significant at the 5 per cent level.
* Significant at the 10 per cent level.

Table 5 – Spatial autocorrelation of model residuals, 1996–2002

	1996		1997		1998		1999		2000		2001		2002	
	MI	Pr	MI	Pr	MI	Pr	MI	Pr	MI	Pr	MI	Pr	MI	Pr
GLM	0.603	0.000	0.577	0.000	0.578	0.000	0.652	0.000	0.722	0.000	0.451	0.000	0.363	0.000
GLM-SF	0.006	0.780	0.001	0.903	–0.012	0.755	0.003	0.862	–0.013	0.710	–0.027	0.398	0.017	0.527
GLM-SF RD	0.193	0.000	0.207	0.000	0.185	0.000	0.170	0.000	0.211	0.000	0.206	0.000	0.240	0.000
GLM-SF RD ALL	0.190	0.000	0.180	0.000	0.162	0.000	0.188	0.000	0.279	0.000	0.231	0.000	0.204	0.000
GLMM-SF RD	0.266	0.000	0.267	0.000	0.229	0.000	0.206	0.000	0.275	0.000	0.281	0.000	0.275	0.000

Notes: GLM uses only the three covariates; GLM-SF uses the covariates and the selected eigenvectors (year by year); GLM-SF RD uses the covariates and the reduced set of eigenvectors common to the seven years; GLM-SF RD ALL uses the entire panel and the common eigenvectors, but ignoring the repeated measurements correlation; GLMM-SF RD uses the entire panel and the common eigenvectors, and random effects (see next section).

Table 6 – GLMM parameter estimates, 1996–2002

Parameter	Value	Std.Error	*t*-value	*p*-value
Intercept	–0.658	0.317	–2.079	0.038**
Wages	0.375	0.119	3.137	0.002***
Employment	–2.037	0.082	–24.867	0.000***
Population	0.119	0.340	0.350	0.726
E_2	7.409	0.246	30.146	0.000***
E_6	–2.279	0.246	–9.268	0.000***
E_7	0.894	0.246	3.641	0.000***
E_8	0.888	0.245	3.617	0.000***
E_9	1.669	0.245	6.825	0.000***
E_{11}	–0.873	0.245	–3.556	0.000***
E_{15}	–1.882	0.245	–7.681	0.000***
E_{16}	0.828	0.245	3.385	0.001***
E_{17}	–0.888	0.246	–3.611	0.000***
E_{18}	1.006	0.245	4.103	0.000***
E_{20}	–0.805	0.246	–3.273	0.001***
E_{24}	–0.988	0.245	–4.033	0.000***
E_{25}	0.652	0.244	2.667	0.008***
E_{26}	–0.765	0.246	–3.110	0.002***
E_{28}	0.877	0.244	3.590	0.000***
E_{30}	–1.041	0.245	–4.248	0.000***
E_{38}	–0.625	0.245	–2.556	0.011**
E_{39}	0.825	0.245	3.370	0.001***
E_{42}	0.586	0.244	2.401	0.017**
E_{60}	0.562	0.245	2.293	0.022**
E_{74}	–0.528	0.245	–2.158	0.032**

*** Significant at the 1 per cent level.
** Significant at the 5 per cent level.

Table 7 – GLMM fitting, 1996–2002

Year	1996	1997	1998	1999	2000	2001	2002
Pseudo-R^2	0.937	0.959	0.960	0.952	0.944	0.939	0.921

Table 8 – Year-by-year specification search LM test results, 1996–2002

	1996	1997	1998	1999	2000	2001	2002
LM-lag	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LM-error	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Robust LM-lag	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Robust LM-error	No	No	No	No	Yes (5%)	No	No
Suggested model	Spatial lag	Spatial lag	Spatial lag	Spatial lag	??	Spatial lag	Spatial lag

Yes: H_0 rejected (significant at the 1 per cent level).
No: H_0 not rejected.

Table 9 – Fit statistics for the benchmark model specifications, 1996–2002

Model	(Pseudo-)R^2	Lag coefficient
OLS	0.3276	–
Spatial lag	0.7528	0.57^{***}
Spatial lag with time fixed effects	0.7934	0.57^{***}

*** 1 per cent significant.

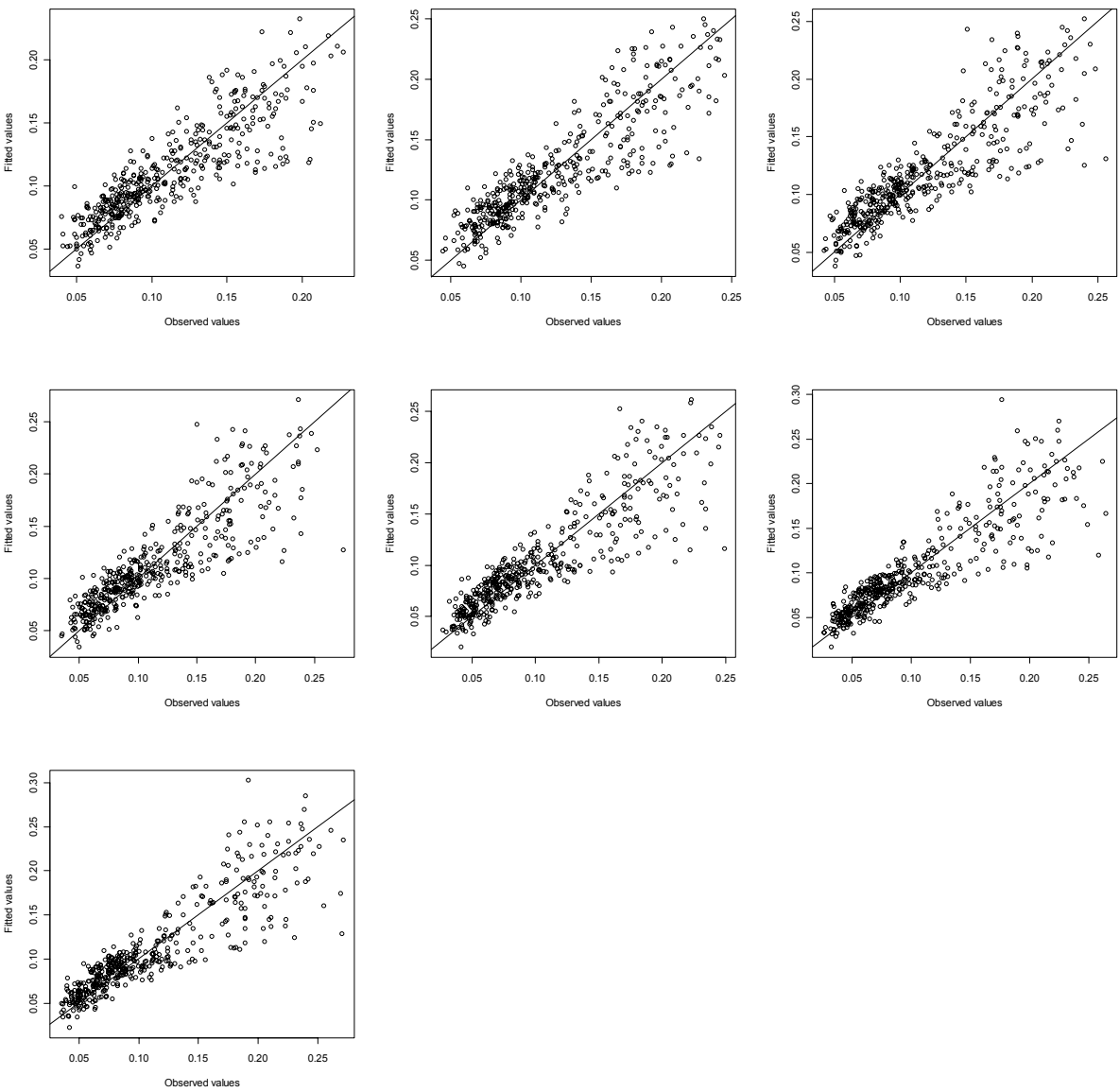
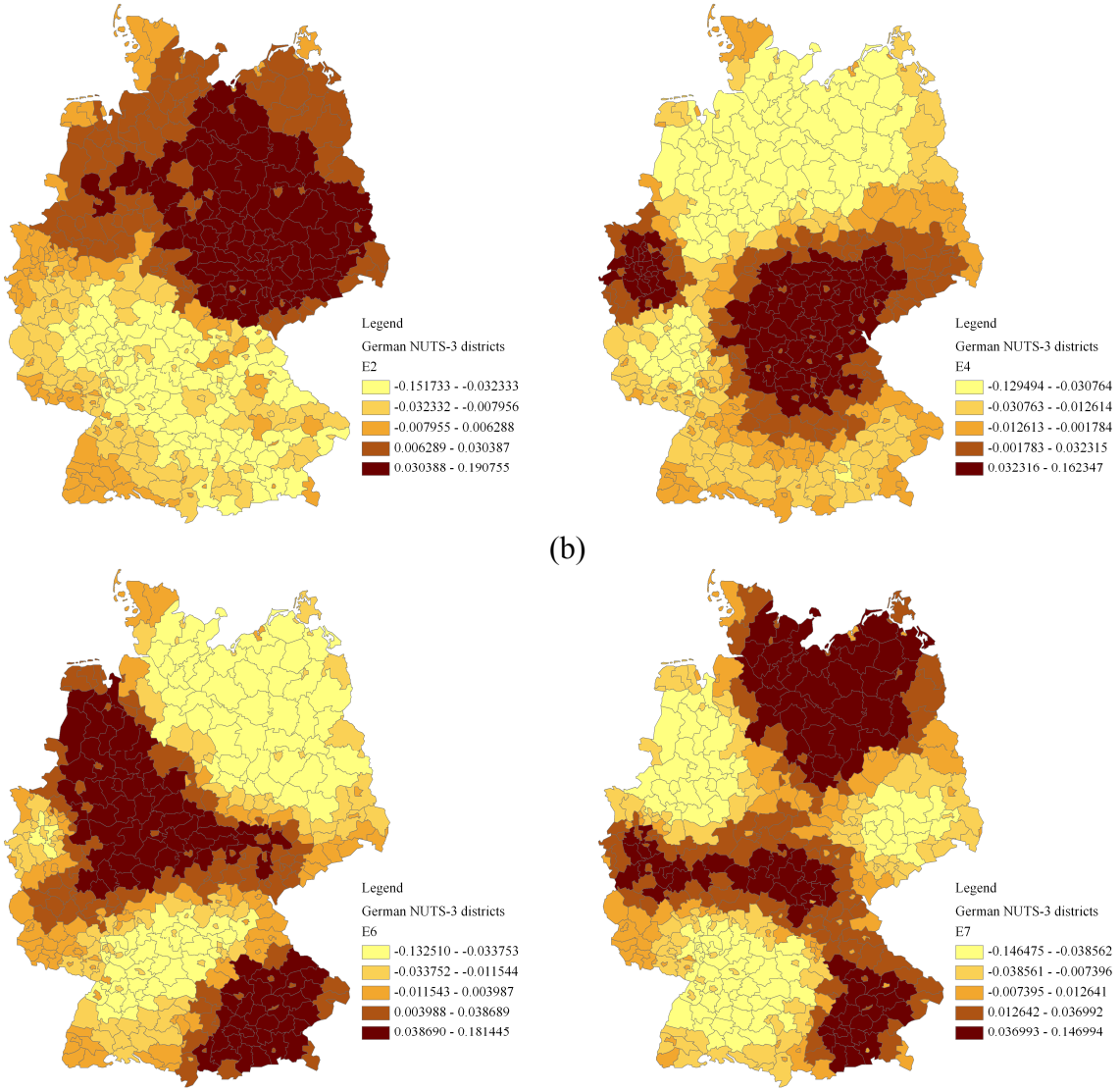


Figure 1 – Plots of observed (*x*-axis) and fitted (*y*-axis) values (left to right, top row: 1996, 1997 and 1998; middle row: 1999, 2000 and 2001, bottom row: 2002)



(a) (b)

(c) (d)

Figure 2 – Maps of main common eigenvectors (E_2, E_4, E_6, E_7) according to MI values (C-coding scheme with the rook's definition of contiguity spatial weights matrix)

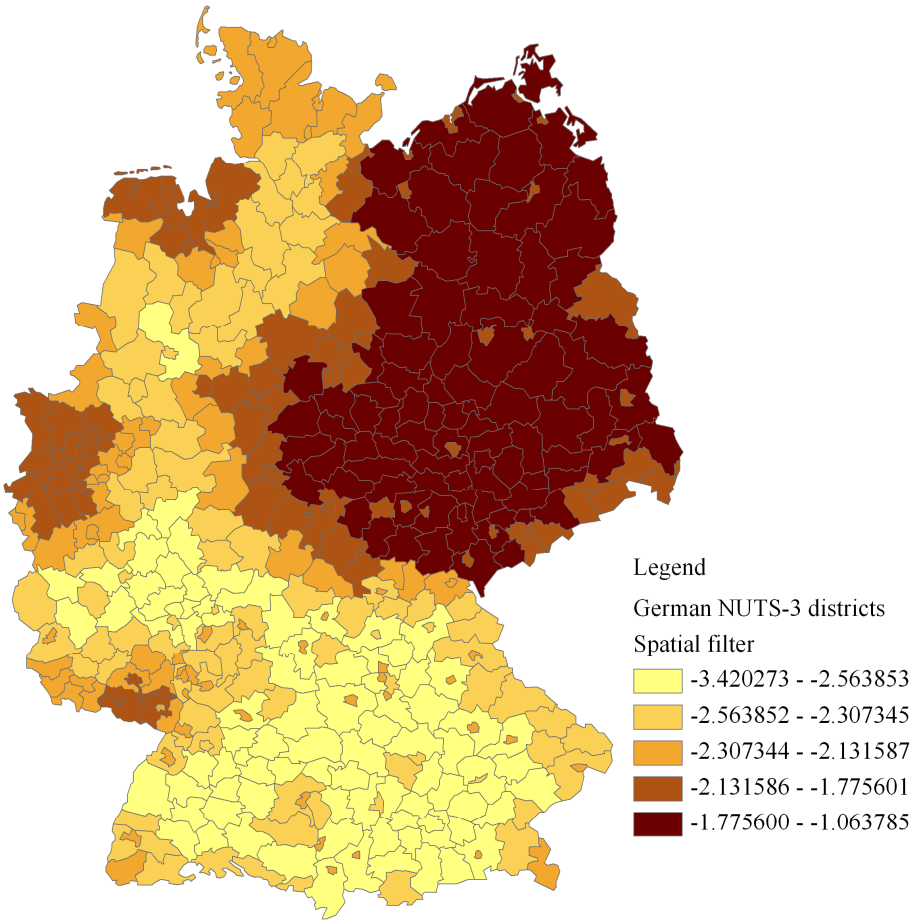


Figure 3 – Graphical visualization of the spatial filter obtained in the case of the rook's definition of contiguity (C-coding scheme) spatial weights matrix, GLMM estimation

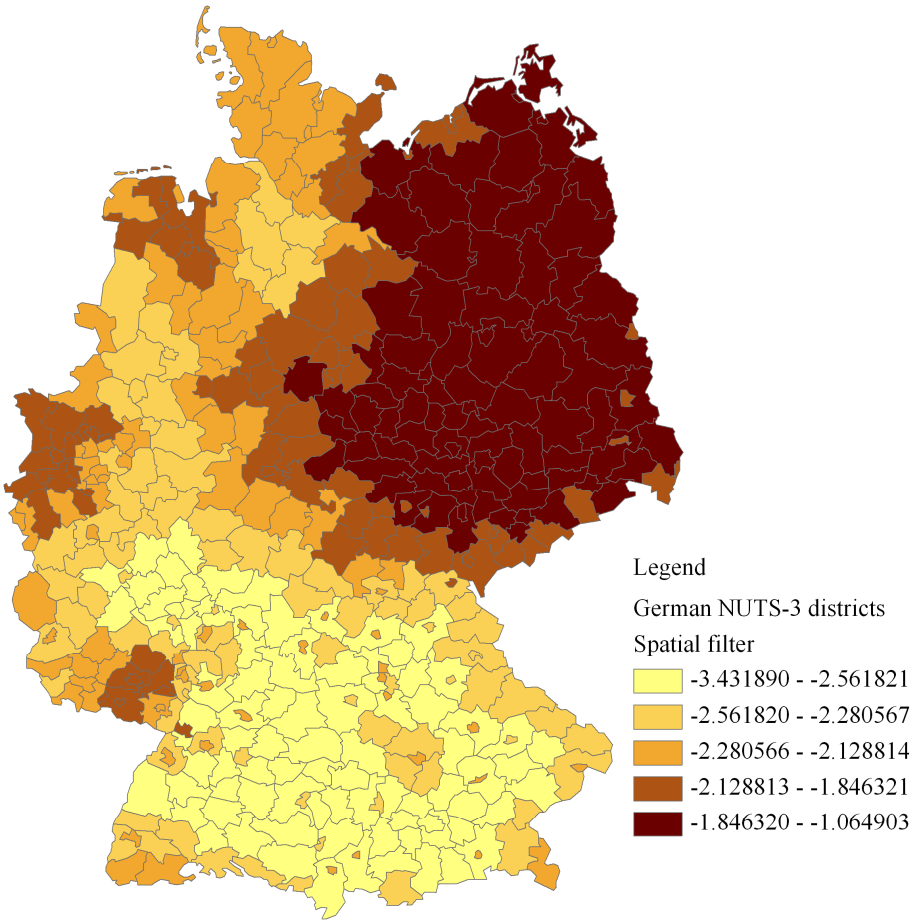


Figure 4 – Graphical visualization of the spatial filter obtained in the case of the rook's definition of contiguity (C-coding scheme) spatial weights matrix, GLMM estimation