Examining the Structure of Spatial Health Effects using Hierarchical Bayes Models

Peter Eibich[1] and Nicolas R. Ziebarth[2]

This Version: February 19, 2013

Abstract

This paper makes use of Hierarchical Bayes Models to model and estimate spatial health effects. We focus on Germany, combining rich individual-level household panel data with administrative county–level information to estimate spatial county-level health dependencies. As dependent variable, we use the generic, continuous, and quasi-objective SF12 health measure. Our findings reveal strong and highly significant spatial dependencies and clusters. The strong and systematic county-level impact is comparable to an age effect on health of up to 30 years. Even 20 years after the peaceful German reunification, we detect a clear spatial East-West health pattern that equals an age impact on health of up to 10 life years.

Keywords: spatial health effects, Hierarchical Bayes Models, Germany, SOEP, SF12

JEL codes: C21, C11, I12, I14, I18

---

[1] Corresponding author: DIW Berlin, Mohrenstrasse 58, 10117 Berlin, Germany, e-mail: peibich@diw.de, Phone: +49-(30) 897 89-223, Fax: +49-(30) 897 89-200.

[2] Cornell University, Policy Analysis and Management (PAM), 106 Martha van Rensselaer Hall, Ithaca, NY 14853, DIW Berlin, and IZA Bonn, e-mail: nrz2@cornell.edu, Phone: +1-(607) 255-1180, Fax: +1-(607) 255-4071.

1 Introduction

It is long recognized by applied empirical researchers that regional and neighborhood effects may play a crucial role in the analysis of a wide array of relevant outcomes. For example, in the economics literature, Burchardi and Hassan (2011) exploit the natural experiment of the German reunification to demonstrate that social ties have a long-lasting impact on individual and regional economic prosperity. Studies in other fields, e.g. sociology, analyze the relevance of relative deprivation theories by looking at a respondents' relative income position in the neighborhood (Durlauf, 2003; Knies et al., 2008).

When it comes to regional or neighborhood impacts on individuals' health, the literature is fragmented, both within and across fields. A relatively large set of studies in the epidemiological and public health literature focuses on geographical variation in health or health care (Frakes, 2013; Sundmacher et al., 2012, Voigtländer et al., 2010). This rich strand of the literature basically emerged from the seminal "Small area variations in health care delivery" article by Wennberg and Gittelsohn (1973). The Small Area Variation Literature is extremely influential, particularly in the US. The DARTMOUTH ATLAS OF HEALTH CARE provides detailed descriptive information on how the health care infrastructure, as well as health care utilization, vary on a disaggregated geographical level, e.g., on the US county level (Dartmouth Atlas of Health Care, 2012). The high policy relevance comes from the fact that utilization and spending measures are not systematically correlated with better health outcomes. Understanding the driving forces of regional differences might, therefore, offer opportunities to limit health care spending while improving health outcomes (Wennberg et al., 2002).

A new subfield in the health economics literature is emerging. This subfield acknowledges the importance of regional factors in the analysis of health and health care. More precisely, this subfield econometrically models spatial interdependencies explicitly. In their synthesis of multilevel studies, Riva et al. (2007) point out that the vast majority of studies account only for within-area correlation and disregard between-area dependency. Not taking spatial patterns in the empirical models into account means that one implicitly assumes that the geographical units are statistically independent. This might be a strong and misleading assumption, since administrative or statistical boundaries might not reflect appropriately underlying ecological, social, and economic processes. Spillover effects are likely to occur.

Within the field of health economics, the statistical and econometric spatial modeling approaches, as well as the field of inquiry, may differ. For example, Baltagi and Moscone (2010) and Moscone and Tosetti (2010) analyze the relationship between income and health expenditures using the Common Correlated Effects (CCE) Approach and impose a spatial autoregressive (SAR) structure on the error terms (Pesaran, 2006). Moscone et al. (2007) use a similar approach, but analyze cross-municipality variation in mental health expenditures. Felder and Tauchmann (2013) combine a nonparametric efficiency analysis with parametric regressions and model spatial dependence in healthcare provision at the county level in Germany.

This paper models and estimates spatial patterns of health in Germany. In terms of our empirical modeling techniques, we apply Hierarchical Bayes Models and advance the literature by combining three methodological approaches: First, hierarchical models are used to account for correlation within regions and to disentangle the effect of individual and regional predictors on individual outcomes (Bryk and Raudenbush, 1992; Hox, 2002). We use a three-stage hierarchical model that also accounts for the temporal dimension of the data. Second, we employ methods from spatial econometrics to analyze and model correlation between regions (i.e., spatial correlation), such as MORAN'S I or Conditional Autoregressive (CAR) models (Arbia, 2006; Cressie, 1993). We are aware of very few studies that use such methods in a health-related context (e.g., Costa-Font and Pons-Novell, 2007; Ocaña-Ricola and Mayoral-Cortés, 2010). In addition, spatial models are almost exclusively used with aggregated time-series data. Moreover, most studies investigate regional differences in health care utilization and health care spending, but not in health outcomes (Fuchs et al., 2004). We are not aware of any study in the field of health economics that models individual-level health using spatial econometrics and panel data.[3] Third, we carry out our analysis using Bayesian methods (Banerjee et al., 2004; Cressie and Wikle, 2011), which facilitate modeling and estimation of the regional effects.

In terms of content, we focus on a continuous and generic health measure, the SF12. The SF12 was developed by public health scientists and has advantages over the conventional categorical subjective health measures (Frick and Ziebarth, 2013). It can be interpreted as a quasi-objective single measure of an individual's health status (RAND, 1995). This paper intentionally focuses on a single continuous measure of individual quasi-objective health rather than measures of the health care infrastructure, which do not vary at the individual level. By taking into account a rich set of socio-demographic individual-level information, we net out important impact factors on health such as age, gender, employment, and marital status. Importantly, we incorporate health behavior measures such as smoking, alcohol consumption, and BMI. In addition, we wash out individual variation in health that can be explained by differences in health care utilization. Moreover, we incorporate ten county-level indicators that vary across the 401 German counties, such as the unemployment rate and population density. Hence, in our three level hierarchical models, we strive to unravel the conditional spatial structure of population health after having corrected for the rich set of individual and county-level factors just listed.

Our results show that large regional differences in health exist within Germany. Spatial interdependences and health clusters are of great importance and are not fully explained by observable

---

[3] Notable exceptions in the fields of Biology and Sociology include Mueller et al. (2001), who examine spatial patterns of child growth in Papua New Guinea, and Browning et al. (2003), who analyze variation in health status across Chicago neighborhoods during the 1990s.

individual and regional characteristics. Surprisingly, the systematic county-level impact on individual health is the most important driving force in our variable list above. County of residence may have an impact on individual health that equals the age effect of up to 31 life years.[4] Particularly stunning is the clear East-West health gradient that seems to be persistent even 20 years after the fall of the Berlin Wall. It may reflect the four decade long impact of Communism versus Capitalism on health and results in an age impact of up to 10 life years.

The rest of the paper is organized as follows: In Section 2 we provide a short description of our underlying dataset, the German Socio-Economic Panel Study (SOEP). Sections 3.1 and 3.2 explain the econometric models and review the basics of Bayesian inference. Section 4 presents and discusses the results. Section 5 concludes.

2 Data

2.1 Dataset

For the empirical analysis we make use of individual health data and further socioeconomic characteristics provided by the German Socio-Economic Panel Study (SOEP). The SOEP is a representative panel study of private households. Starting in 1984, SOEP interviews subjects annually, and expanded to include residents in former East German lands, starting in 1990. All respondents answer one main individual questionnaire covering about 150 questions on a range of topics, such as the labor market and family situation, attitudes and perceptions, as well as health. Additionally, a household questionnaire is completed by the head of the household. Since 2000 the survey has reached more than 20,000 individuals across 10,000 households. For further details, see Wagner, Frick and Schupp (2007).

The SOEP provides a variety of health measures. Both the standard 5-categorical Self-Assessed Health (SAH) measure and the 11-categorical health satisfaction measure are included every survey wave. In addition, since 2002, in every second year, the continuous quasi-objective SF12 measure and the objective grip strength measure are included in the survey. Furthermore, information on health-related behavior (e.g. alcohol or tobacco consumption) is available since 2006. Therefore, we restrict our analysis to the three years 2006, 2008 and 2010. We use only observations without item-non-response. In total we obtain 54,734 person-year observations from 23,414 different individuals.

2.2 Dependent Variable: SF12

Our dependent variable is the generic health measure SF12. A specific algorithm generates the continuous SF12 measure on the basis of 12 different health-related questions. More precisely, the

---

[4] Note that this does not refer to a change in life expectancy but to the marginal effects of the age polynomial in our regression models.

algorithm generates eight subscales and two superordinate dimensions, namely physical (*pcs*) and mental health (*mcs*). We then average over both components to obtain our dependent variable *SF12*. In the standard SOEP version, the SF12 takes on continuous values between 0 and 100, has mean 50, and a standard deviation of 7 (see Table 1).[5]

2.3 Individual Covariates

By controlling for demographic factors, educational characteristics, labor market participation, health-related behavior as well as health care utilization, we net out all differences in health as measured by the SF12 that can be traced back to these factors.

The ***demographic factors*** are age, age squared, gender, marital status, and the number of children under 14 in the household. Table 1 shows that average age is 49.3 and that about half of our sample is female. The majority of respondents is married and live in the same household as their spouse. There are no children in the majority of SOEP households.

In terms of ***education and labor market participation,*** we control for whether an individual completed a vocational training or holds a university degree, the working status of the respondent, and the monthly equivalent household income.[6] Less than a quarter holds a university degree, but about 60 percent completed vocational training. Vocational training is equivalent to a college degree in the US. Slightly less than half of our sample works and the average equivalent gross household income per capita is €1,800 per month.

Concerning ***health-related behavior***, we control for alcohol and tobacco consumption and Body Mass Index (BMI). Alcohol consumption is captured by dummy variables. In the SOEP questionnaire, the participants are asked to state how often they drink wine and sparkling wine, beer, spirits and mixed drinks. If an individual states that they do not drink any alcohol at all, *no alcohol consumption* is assigned the value “1”, otherwise it is “0”. In contrast, if an individual states to drink any kind of alcohol on a regular basis, *regular alcohol consumption* takes on the value “1”, otherwise “0” (Ziebarth and Grabka, 2009). 17 percent drink alcohol on a regular basis and 13 percent never drink any alcohol. The dummy variable *Smoking* captures whether the respondent consumes tobacco (cigarettes, cigars or pipes). It takes on the value “1” for about 30 percent of the respondents in our sample. The *BMI* is measured as body weight in kilogram divided by the squared height in meters (Burkhauser and Cawley, 2008). The average BMI lies slightly below 26.

---

[5] A detailed description of the algorithm and an overview over the differences between the original “SF12v2™ Health Survey” and the SOEP version can be found in Andersen et al. (2007).

[6] The monthly equivalent household income uses the OECD-modified scale and assigns households a value of 1 to the first adult, 0.5 to each additional adult and 0.3 to each additional child. Further details are provided in *OECD Project on Income Distribution and Poverty (2009).*

Health care utilization is measured by the number of hospital stays during the previous calendar year and the number of doctor visits in the 12 months prior to the interview.[7] Summary statistics for all variables are given in Table 1.

[Insert Table 1 about here]

2.4 Regional Information

The SOEP provides information on the place of residence at multiple levels, ranging from the federal states to specific postal codes (Knies and Spiess, 2007). The SOEP also provides frequency and probability weights to ensure representativeness for most federal states. The state-level might be too aggregate to detect any significant regional effects – especially if these are caused by small scale phenomena (e.g. structural environmental and economic factors). Therefore, in this paper, we focus on counties ("*Landkreise*" and "*Kreisfreie Städte*"). This offers several advantages: First, they are the smallest administrative units in Germany. Furthermore, administrative data exists at the county-level, which can be used as independent covariates to explain variation between counties. Lastly, the SOEP contains observations from almost all counties. Nevertheless there are some disadvantages. For instance, the data is not representative and there are no probability weights available to correct for over- and undersampling in certain regions. The number of observations per county varies between 2 (Sömmerda) and 2,025 (Berlin) with a mean of 136 observations and a standard deviation of 148. Thus the findings of this analysis should be interpreted carefully. For some counties the number of observations is too low to draw definite conclusions, especially when it comes to the health development of the county over time.

The number of counties, and their respective borders, has evolved since 2006. Therefore, the SOEP Group constantly updates county codes. To ensure consistency of the data with the shapefiles provided by the FEDERAL AGENCY FOR CARTOGRAPHY AND GEODESY (2012) ("*Bundesamt für Kartographie und Geodäsie*"), we recoded the county codes to reflect the boundaries as of January 1, 2012.

As of January 1, 2012, Germany consisted of 402 counties. Our dataset contains observations from 401 counties with Memmingen, a county in Bavaria, missing. We ignore this county in our analysis and remove it from the neighborhood matrix for the spatial models.

In order to explain systematic county-level differences in health, we incorporate county-level information into our analysis. For this purpose, we exploit information provided by the FEDERAL INSTITUTE FOR RESEARCH ON BUILDING, URBAN AFFAIRS, AND SPATIAL DEVELOPMENT (2012)

---

[7] It should be noted that the respondents are asked to state the number of doctor visits during the last three months prior to the interview. The answer is then multiplied by four to generate the annual number of doctor visits. This may overestimate the actual annual number because approximately two-thirds of all interviews are carried out in the first quarter of a year and there is a clear seasonal pattern in doctor visits.

("*Bundesinistitut für Bau-, Stadt- und Raumforschung*") in their INKAR ("INDICATORS AND MAPS ON SPATIAL DEVELOPMENT") database.[8] As can be inferred from Table 1, we use county-level data on the *area size,* the *population density,* the *proportion of the area used for settlement and transport,* and the *proportion of the area suitable for recreation* in order to ***distinguish urban from rural counties***. Note that there is significant variation in population density across counties, ranging from 39 to 4,355 inhabitants per square kilometer. There is also wide variation in terms of the size of the area that is used for urban purposes, i.e., the degree of urbanization. The values for this variable vary from 5 to 76 percent. The size of the area for recreational purposes varies between 0.2 and 15 percent. Obviously, these indicators capture a great deal of cross-county heterogeneity that is not captured by individual-level indicators and might affect health. For example, in general, population density is highly correlated with the density of health care providers and the health care infrastructure. On the other hand, low urbanization rates or large recreational facilities may represent low air pollution and low noise (e.g. Coneus and Spieß, 2011).

We use average *available monthly income per capita* and the annual *unemployment rate* to characterize the ***degree of regional deprivation***. The *population share of immigrants* represents the diversity of the community. Again, we observe huge variation across counties for all these indicators. Average per capita income varies from €1,108 to €2,701. Between 1 and 26 percent of the population of a county are immigrants and unemployment rates vary from 1.9 to 26.20 percent. The *ratio of overnight stays to inhabitants* in the tourism sector may both indicate that a county is located in a landscape that is beneficial to health, e.g., seaside beaches or in the mountains. Alternatively, it may measure the degree of the traffic volume (caused by tourism) in the county. The *car density* may serve as an indicator of air and noise pollution. It varies between 316 and 881 cars per 1,000 inhabitants.

The *physician density* (measured as inhabitants per physician) represents the ***health care infrastructure*** of the county. In counties with fewer physicians, residents may experience longer travel and waiting times if they want to consult a physician, i.e., a lower degree of access to and utilization of the health care infrastructure and, consequently, health. On the other hand, following the conclusions of Wennberg and Gittelssohn (1973), physician density might be uncorrelated or negatively correlated with individual health.

3 Hierarchical Bayes Models and Methods

---

[8] As for the individual level variables, we recode the county codes to the status as of January 1, 2012. However, the data taken from the INKAR database reflects the territorial status as of January 1, 2010. To take the 2011 county mergers into account, we calculate the values for the new counties as a weighted average of the values of the merging counties. The variables are weighted either by population number or by area size, depending on the respective variable (e.g. the *unemployment rate* refers to the population and is thus weighted by population count. In contrast, the *area used for settlement and transport* refers to the area and is therefore weighted by area size).

3.1 Econometric Models

The aim of our analysis is to detect and model spatial patterns in the distribution of regional effects on health. With a few notable exceptions (discussed in Section 1), most studies using methods from spatial econometrics or spatial statistics rely on aggregate measures of morbidity (e.g., disease counts) as a dependent variable. In contrast, we model individual health. At the same time, we control for systematic differences between counties as outlined above. In addition, we control for common time shocks across counties. However, the main objective of the paper is to model spatial heterogeneity in health at the individual level. The resulting models are inherently hierarchical with observations nested within individuals nested within regions. The following model serves as our baseline model throughout the analysis:

$$SF12_{ist} \sim N(\mu_{ist}, \sigma^2)$$

$$\mu_{ist} = \theta + X_{it}\beta + Z_{st}\gamma + c_i + b_s + \delta_t \qquad (3.1.1)$$

We assume that the *SF12* indicators for individuals $i = 1, \dots, 23{,}414$ in counties $s = 1, \dots, 401$ and years $t = 2006, 2008, 2010$ are conditionally independent and normally distributed with mean μ_{ist} and variance σ^2. The mean μ_{ist} is a linear function of an intercept θ, the (1 x 15) vector of individual regressors X_{it} with a (15 x 1) parameter vector β, the (1 x 10) vector of county-level covariates Z_{st} with a (10 x 1) parameter vector γ, an individual effect c_i a regional effect b_s and aggregate time effects δ_t.

Our interest lies mainly in the regional effects b_s and their spatial distribution. In the following section, we specify four different assumptions on the distribution of these effects and their spatial and temporal dependency, which will lead us to our candidate models

3.1.1 Model 1: Unstructured Regional Effects (URE)

In our first model, we assume that the county-level effects follow a normal distribution with mean zero and variance σ_b^2:

$$\mathrm{b_s} \sim \mathrm{N}(0, \sigma_\mathrm{b}^2)\ .$$

This implies that, on average, the regional effects are zero, which is reasonable since we include an intercept in our model. This model would be equivalent to a model with dummy variables for each county, i.e., county-fixed effects. We do not specify the type of spatial dependency in this model, which is why we refer to this model as the **Unstructured Model (Model 1).** In this initial stage of our analysis, we seek to obtain estimates of the regional effects whose spatial distribution we would like to analyze. These estimates are “raw” in the sense that they are not smoothed by assumptions on the spatial dependency; they are used to examine the spatial dependency patterns of the regional effects by maps and test statistics.

Then, in the next stage, we impose a structure on the spatial patterns and use models that smooth the regional effects. This approach is proposed by Banerjee et al. (2004).

3.1.2 Model 2: Intrinsic Conditional Autoregressive Regional Effects (ICAR)

In order to account for spatial dependency between counties, we model regional effects using an **Intrinsic Conditional Autoregressive (ICAR) Model (Model 2)** as proposed by Besag et al. (1991). This model is autoregressive in a sense that the value for each area b_s depends on its spatial lags $\mathrm{b_r}, \mathrm{r} \neq \mathrm{s}$. It is characterized by the conditional distributions:

$$b_s |\, b_{r \neq s} \sim N(\bar{b}_s, \frac{\sigma_b^2}{m_s}). \tag{3.1.2}$$

The regional effect b_s is normally distributed around the mean, where $\mathrm{\bar{b}_s = m_s^{-1} \sum_{r \in \eta_s} b_r}$. m_s denotes the number of neighbors and η_s is the set of neighbors of county s (Banerjee et al., 2004). The mean equals the average effect of the neighboring counties of county s. σ_b^2 is a variance parameter. The variance depends on this parameter as well as the number of neighbors. Thus isolated counties with fewer neighbors exhibit greater variance than counties with more neighbors.

The conditional distribution in equation (3.1.2) specifies that the regional effects b_s are Markovian, i.e., they depend only on their neighbors.[9] From equation (3.1.2), the joint density $p(b_1, \dots, b_n)$ can be obtained using Brook’s Lemma (Besag, 1974) and the Hammersley-Clifford Theorem (Hammersley and Clifford, 1971). However, Besag (1995) notes that the resulting distributions would not exhibit appreciable correlations unless the parameters are close to the boundaries of the parameter space. Therefore, he suggests the following pair-wise specification:

$$p(b_1, \dots, b_n) \propto \exp\{-\frac{1}{2\sigma_b^2} \sum_{s \neq r} w_{sr} (b_s - b_r)^2\}, \tag{3.1.3}$$

i.e., the joint distribution of $b_1, \dots, b_n$ is proportional to the pairwise difference density, where w_{sr} are the non-diagonal entries of the proximity matrix W. The non-diagonal entries of the proximity matrix W specify which areas are considered to be associated (and therefore spatially dependent). The choice of W will be explained in more detail in Section 4.

---

[9] The model specified through eq. (3.1.2) is also called the Autonormal or Autogaussian model, i.e., a model for normally distributed data on a Markov Random Field (Arbia, 2006)

Equation (3.1.3) is a limiting form[10] of standard conditional autoregressive normal distributions and it is usually referred to as an Intrinsic Autoregressive (IAR) or Intrinsic Conditional Autoregressive model (ICAR) (Besag, 1995). However, equation (3.1.3) does not yield a proper distribution, since the b_s are not centered and need an additional “sum to zero” constraint (Banerjee et al., 2004). Nevertheless, if used as prior information about the unknown parameter (i.e., as a prior distribution, see section 3.2.1) in Bayesian Analysis, the resulting inference (through the posterior distribution) is proper.

For Model 2, we assume an ICAR distribution for the regional effects and choose a neighborhood matrix W based on first-order adjacency (see below).

3.1.3 Models 3 and 4: (Spatiotemporal) Convolution Prior (SCP)

In Model 2 described above, we assume that the regional effects can be described by an ICAR-model, which implies that the regional effect of county *s* depends only on the values of *all* neighboring counties. This assumption may be too strict, since the effect might be a combination of a spatially dependent part and a randomly distributed part (noise).

Therefore, now we decompose the effects

$b_s = \varphi_s + \omega_s$,

where

$\omega_s \sim N(0,\ \sigma_\omega^2)$

and

$$\varphi_s |\, \varphi_{r\neq s} \sim N(\bar{\varphi}_s, \frac{\sigma_\varphi^2}{m_s}), \qquad (3.1.4)$$

that is, we discriminate between regional heterogeneity captured by ω_s and regional clustering captured by φ_s (Banerjee et al., 2004). We call this a ***Convolution Prior*** since the distribution of b_s is a convolution of the distributions described above (Besag et al. 1991; Mollié, 1996). Using this structural assumption and the neighborhood matrix based on adjacency, we obtain our **Model 3: the Convolution Prior Model (CP)**.

Models 2 and 3 incorporate spatial dependence and aggregate time effects. However, the spatial effects b_s, φ_s and ω_s do not depend on time *t*, i.e., they are assumed to be time-invariant. This may be a plausible assumption if the main driving factors of the regional differences are time-invariant, e.g., structural environmental and economic factors. Nevertheless, we check this assumption by extending the CP-Model to include space-time interaction effects. As such, we obtain our fourth and final model: the **Spatiotemporal Convolution Prior (SCP)**.

---

[10] I.e. the parameter that determines the strength of the spatial association is fixed to the boundaries of the parameter space.

Consider the model

$$\mu_{ist} = \theta + X_{it}\beta + Z_{st}\gamma + b_{st} + c_i + \delta_t, \tag{3.1.5}$$

where

$$b_{st} = \omega_{st} + \varphi_{st}$$

and

$$\varphi_{st}|\varphi_{rt\neq st} \sim N(\bar{\varphi}_{st}, \sigma^2_{\varphi t}/m_s),$$

$$\omega_{st} \sim N(0, \sigma^2_{\omega t}).$$

Here, both the spatially dependent part and the random part of the regional effect depend on location *s* and time *t*, i.e., the model produces a separate effect for each county and each year (comparable to an interaction between dummy variables for county *s* and year *t*). This model has the advantage that we can estimate the spatial pattern for each year. However, it requires observations for each county in each year. Therefore, for **Model 4**, we restrict the dataset to counties with observations in each year. The adjusted dataset contains 54,723 out of 54,734 observations and 398 out of 401 counties.

3.2 Estimation

In this section we give a brief overview of the basics of Bayesian Methods and then detail the specifications under which the described models above are analyzed.

3.2.1 Bayesian Inference

In Bayesian inference, given the data, the main interest lies in learning about the distribution of the unknown parameters. In contrast to frequentist methods, such as Maximum Likelihood Estimation, the parameters are treated as stochastic quantities that follow a probability distribution.[11] The fundamental principle of Bayesian methods is summarized by Bayes formula:

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}, \tag{3.2.1}$$

where y is a random variable and θ is the unknown parameter. $p(y|\theta)$ is the likelihood function of the data. $p(\theta)$ is called a ***prior distribution***, which expresses prior belief or knowledge (e.g., from similar or older studies) about the distribution of the unknown parameter. $p(y)$ is the unconditional distribution of the data and serves as a normalizing constant to ensure that $p(\theta|y)$ integrates to one. $p(\theta|y)$ is called the ***posterior distribution*** and summarizes all knowledge of the unknown parameters given the data. Bayes formula may be regarded as a method to update prior knowledge about the parameters θ using the data y (Congdon, 2010).

---

[11] It is worth noting that the parameters are not regarded as random quantities - they are fixed but unknown. The incomplete knowledge about these parameters is assumed to be random (Gelman and Robert, 2012).

Bayesian inference is carried out by summarizing the posterior distribution. The point estimates depend on the appropriate “loss function”, e.g., if we want to minimize $E[(\theta-\hat{\theta})^2]$, the estimator is given by the posterior mean. For multimodal or non-normal distributions, the posterior median may be more appropriate (Banerjee et al., 2004).

Bayesian confidence intervals (usually called ***credible intervals***) are **Equal Tail (ET) Intervals** or **Highest Posterior Density (HPD)** intervals. For symmetric and unimodal posterior distributions, both intervals are identical. Otherwise the HPD interval will have shorter length, but it is more difficult to compute.

The ET interval is given as the interval $[q_L, q_U]$ where:

$$\int_{-\infty}^{q_L} p(\theta|y)d\theta = \frac{\alpha}{2}$$

and

$$\int_{q_U}^{\infty} p(\theta|y)d\theta = 1 - \frac{\alpha}{2}.$$

As usual, α denotes the confidence level. An attractive feature of the credible interval is that the often erroneous interpretation of confidence intervals in frequentist models does actually hold for them, namely the probability that θ lies in $[q_L, q_U]$ is $(1-\alpha)$.

A crucial point in Bayesian Analysis is the choice of the prior distribution. The posterior distribution depends on both the prior distribution and the likelihood. Likewise, the parameters of the posterior distribution (e.g., the mean) are a weighted average of the prior values and the likelihood-based values (for an example, see Banerjee et al., 2004).

This offers a way to formally include subjective beliefs or existing knowledge. However, often researchers prefer to not introduce any subjective elements, either since no additional information on the problem is available, or since the researcher prefers to remain as “objective” as possible. In this case, improper or uninformative priors may be chosen (e.g., a flat distribution over the entire real line). This leads to another problem: the determination of the posterior distribution.

3.2.2 The Gibbs sampler

An analytic derivation of the posterior through equation (3.2.1) is only possible for specific, “conjugate” prior distributions, i.e., when prior and posterior belong to the same distributional family. For non-conjugate, and especially for improper priors (i.e., distributions that do not integrate to anything finite), the required integrations—namely for $p(y)$—are intractable. In this case, **Markov**

Chain Monte Carlo (MCMC) methods may be used to obtain samples from the posterior distribution without knowing its exact analytical form.

The general idea of MCMC methods is to sample from a transition distribution

$$K(\theta^{(t)}|\theta^{(t-1)}, \dots, \theta^{(0)}) = K(\theta^{(t)}|\theta^{(t-1)}),$$

which has the posterior distribution $p(\theta|y)$ as its stationary distribution. $K(\cdot \mid \cdot)$ is a transition kernel, i.e. a probability distribution, which specifies the form of the transition from $\theta^{(t-1)}$ to $\theta^{(t)}$.[12] $\theta^{(0)}$ is a set of initial values for the parameters $\theta = (\theta_1, \dots, \theta_k)'$. After a "burn-in phase" of B samples, $\theta^{(t)}, t \geq B$ may be seen as a random sample from the posterior $p(\theta|y)$ (Congdon, 2010).

The Gibbs sampler described below generates samples from the joint posterior distribution by sampling from the full set of conditional distributions. Assume we are interested in k parameters $\theta = (\theta_1, \dots, \theta_k)'$. Then, for arbitrary initial values $\theta^{(0)}$, for *t=1,…T*, repeat:

Step 1: Draw ${\theta_1}^{(t)}$ from $p({\theta_1}^{(t)}|{\theta_2}^{(t-1)}, {\theta_3}^{(t-1)}, \dots, {\theta_k}^{(t-1)}, y)$.
Step 2: Draw ${\theta_2}^{(t)}$ from $p({\theta_2}^{(t)}|{\theta_1}^{(t)}, {\theta_3}^{(t-1)}, \dots, {\theta_k}^{(t-1)}, y)$.
…
Step k: Draw ${\theta_k}^{(t)}$ from $p({\theta_k}^{(t)}|{\theta_1}^{(t)}, {\theta_2}^{(t)}, \dots, {\theta_{k-1}}^{(t)}, y)$.

It can be shown that the k draws $\theta^{(t)}$ converge in distribution to a draw from the joint posterior $p(\theta|y)$ (Banerjee et al., 2004).

Samples from the conditional distributions may be obtained through other methods, such as the Metropolis-Hastings Algorithm. For an extensive treatment of MCMC methods, see Congdon (2010) or Robert and Casella (2005).

3.2.3 Specifications

Bayesian methods offer some important benefits. First, they offer a way to formally incorporate prior knowledge or beliefs into the analysis. Probably most important for our analysis, hierarchical and spatiotemporal models can be adapted quite naturally in a Bayesian framework (Congdon, 2010). Furthermore, Hierarchical Bayes Models impose a common prior distribution on group-level effects (i.e. individual and regional effects in our models). This allows for a more efficient estimation of the parameters by exploiting the similarity of the parameters, which is modeled through the common prior distribution. This effect is called "borrowing of strength" (Congdon, 2010). Bayesian Analysis itself is

---

[12] For example, if $K(\theta^{(t)}|\theta^{(t-1)})$ is specified as $N(\theta^{(t-1)}, \sigma^2)$, then the transition from $\theta^{(t-1)}$ to $\theta^{(t)}$ would be a draw from a normal distribution with mean $\theta^{(t-1)}$ and a certain variance σ^2.

especially suited for small samples - since all estimates are conditional upon the data given, we do not have to rely on asymptotic theory. Thus, we formulate our models as Hierarchical Bayes Models and estimate them using Gibbs sampling.

The prior distributions for our analysis are mostly uninformative. They allocate equal probability mass to all plausible values. In detail, we choose a flat prior $U(-\infty, \infty)$ for the intercept θ and a normal prior distribution with an inflated variance $N(0{,}10000)$ for the slopes $\beta_1, \dots, \beta_{15}$ and $\gamma_1, \dots, \gamma_{10}$ and the time fixed effects δ_t.

For our first model, the **Unstructured Regional Effects (URE) Model**, we use a normal prior distribution for the regional effects with a hyperparameter for the variance. For the second model, **the Intrinsic Conditional Autoregressive Regional Effects (ICAR) Model,** we use the ICAR prior as described in section 3.1.2. For the third and fourth model, the **Convolution Prior (CP)** and the **Spatiotemporal Convolution Prior (SCP)** models, we use an inflated normal distribution for ω_s and ω_{st} as well as an ICAR prior for φ_s and φ_{st}. All standard deviation parameters, including hyper-parameters, were assigned a $U(0{,}100)$ prior distribution. This distribution covers all plausible values - remember that the SF12 indicators lies within [0,100] - and assigns them equal probabilities.

In order to speed up convergence, we standardize all non-dummy variables, i.e., we subtract the overall mean and divide them by their standard deviation.

We run three parallel Markov chains with “dispersed”[13] initial values and monitored the values for the intercept, the slope parameters, the regional effects, and the variance parameters using **Trace Plots** and **Gelman-Rubin statistics** (Brooks and Gelman, 1998; Gelman and Rubin, 1992). The length of the burn-in period was determined for each model separately.

For models 1 and 2, we could not detect any departure from convergence after the 5,000th iteration. For models 3 and 4, the burn-in period took 15,000 iterations, after which the sampler seemed to have stabilized. After the burn-in period, we sampled 15,000 draws from the posterior distribution. In order to decrease autocorrelation and speed up “mixing”[14], we thinned the chain by storing only every 10th draw.

4 Results

---

[13] Every chain starts at different initial values. A simple way to monitor convergence of the sampler is to run parallel chains, which start at different initial values. If the chains “overlap” each other, i.e., the draws from the chains fluctuate around the same value, this could be interpreted as a sign of convergence. However, if the chains start at the same initial value, any overlap could be pure coincidence. We initialize the first chain at the mean values of the prior distributions and the second and third at the upper and lower end of the range of plausible values respectively.

[14] I.e. the number of iterations needed for the chain to cover the whole support of the posterior distribution.

4.1 Regional Differences in Health

We estimate and plot the regional means of the SF12 variable to get a first impression of the magnitude and pattern of the regional differences. The results are displayed in the CHOROPLETH MAP in Figure 1 below. To produce this map, we order the values of the regional means and classify them into five categories, with the quintiles serving as cutoff-thresholds. Each category is assigned a shade of a color, with lighter shades corresponding to smaller values (lower quintiles) and darker shades corresponding to greater values (higher quintiles). Naturally, the intervals covered by each category differ; the highest and the lowest category cover a wider range of values than the medium category, since the interval covered by the medium category has higher probability mass.[15] However, the choice of categories is appropriate to give a visual impression of the spatial distribution of the regional means. Accordingly, regional deviations from the national mean range from -14 to 7 points in the SF12 health status measure.

[Insert Figure 1 about here]

However, these large differences are not very meaningful, since they could potentially be explained by differences in demographic factors, e.g. age or gender. We account for demographics by calculating age-sex-adjusted regional means by estimating Model 1 (see Section 3.1) using only age, age squared and gender as explanatory variables. The estimated regional effects are shown in the map in Figure 2 below. Note that the magnitude of the regional differences has decreased as compared to Figure 1. However, regional differences still amount to about 40 percent of the standard deviation of the SF12 measure. A comparison of Figure 1 and Figure 2 also shows that the estimated patterns are very similar, i.e., gender-age differences might partly explain the magnitude but not the distribution of the regional effects. Furthermore, the map gives first evidence that there are several clusters of positive and negative regional effects.

[Insert Figure 2 about here]

4.2 Model 1: Unstructured Regional Effects (URE)

Since age and gender cannot explain the regional differences observed in the mean values, we estimate the ***Unstructured Regional Effects (URE) Model*** (see Section 3.1.1), which includes the whole set of individual and regional predictors described in Section 2 (we will refer to this variant as Model 1a, see Section 4.4). Table 2 yields the parameter estimates for these variables. Please note that the coefficients are standardized for the non-dummy variables. Except for *number of children*, all variables are statistically significant at conventional levels.

---

[15] The distribution is approximately normal, i.e., unimodal and symmetric.

[Insert Table2 about here]

Looking at the sign of the coefficients, we note that, in general, males are healthier than females. This is in line with the stylized fact that women have greater health care expenditures, not only in childbearing years (Owens, 2008). Also, individuals who abstain from alcohol have a lower health status. This group is likely to contain respondents with current or past serious diseases. On the other hand, regular alcohol consumption is positively correlated with health; this might seem surprising, but is consistent with the literature (Ziebarth and Grabka, 2009). The number of hospital stays as well as the number of doctor visits in the last year—in other words: health care utilization—is negatively correlated with respondents' health. Concerning the size of the effects, gender, alcohol abstinence and smoking status show the largest associations with individuals' health, comparable to an age effect of 14 (male gender and alcohol abstinence) and 12 years (smoking).

Of the county-level variables, only the unemployment rate and the car density show a significant association with health. A one percentage point change in the unemployment rate can be compared to an age effect of half a life year.

However, the main focus of our analysis lies on the regional effects b_S and their spatial distribution that we discuss in detail below.

Spatial Pattern of the regional effects

The "posterior means"—the point estimates of the regional coefficients in the URE-Model—are plotted in form of a quantile map in Figure 3.

[Insert Figure 3 about here]

We observe the following: First, the county level health effects differ between -2.5 and 2.7.[16] This means that the overall SF12 mean (the intercept) varies between 46.3 and 51.5, depending on the county.[17] When compared to the age effect, interestingly, the regional effect has an influence on health that equals up to 31 years in both directions.

Second, it seems as if the counties in East Germany tend do have lower health values than the counties in West Germany. Overall, the map is dominated by clusters: We find clusters of high values in the Northwest (around Hanover and Hamburg), West (Cologne) and Southwest (Palatinate region). Clusters of smaller values can be found in the Southeast (Lower Bavaria), Center (Thuringia) and in the Northeast (Mecklenburg-Western Pomerania and Brandenburg).

---

[16] Of course not all of these effects are statistically significant. 95 of the 401 regional effects are significant at a 5 percent level and 155 are significant at a 10 percent level.

[17] Standard Errors and Monte Carlo Errors are excluded from the calculation.

Testing for Spatial Dependency using Moran's I

These clusters already suggest that the effects are not spatially randomly distributed and that there exists spatial dependency between neighboring counties. Therefore we calculate MORAN'S I to measure the strength of the spatial association and to formally test the hypothesis of no spatial autocorrelation.[18]

MORAN'S I can be interpreted as a spatial analogue for the lagged autocorrelation coefficient statistic in time series models and is defined by:

$$I = \frac{n \sum_s \sum_r w_{sr}(b_s - \bar{b})(b_r - \bar{b})}{(\sum_{s \neq r} w_{sr}) \sum_s (b_s - \bar{b})^2} \qquad (4.1.1)$$

b_s is the value of random variable $\mathbf{b}$ at location $s = 1, \ldots, n$, i.e., our regional effect. w_{sr} are the off-diagonal entries of the neighborhood matrix W. They represent the spatial connection between sites s and r.[19] W can be defined as a binary matrix; in this case, $w_{sr} = 1$ indicates that sites s and r are neighbors (Arbia, 2006). Another possibility would be to define w_{sr} as weights, e.g., proportional to the inverse distance.

Moran (1950) derives the asymptotic distribution of I; however, the variance expression is very complex. Banerjee et al. (2004) recommend carrying out a Monte Carlo permutation test by drawing a random sample of permutations[20] including the observed one. Then, the observed I can be positioned relative to the other permutations and a pseudo p-value may be derived.

We conduct the spatial analysis using a neighborhood matrix based on first-order adjacency. This means that sites are considered to be neighbors if they share a *common border*. This definition is frequently used in applications of spatial econometrics. It is advantageous since it results in a symmetric neighborhood matrix. The ICAR specifications described in section 3.1.2 and 3.1.3 require a symmetric neighborhood matrix in order to meet the conditions of the Hammersley-Clifford Theorem (Hammersley and Clifford, 1971). Therefore we choose the first-order adjacency definition for our spatial models.

Nevertheless, for illustrative purposes, we calculate MORAN'S *I* using eight additional neighborhood matrices based on ***three different definitions of neighborhood*** (Arbia, 2006) and test the significance of *I* using 10,000 permutations, including the observed values: The first definition is

---

[18] Arbia (2006) points out that MORAN'S I is not a proper statistical test since it does not consider an explicit alternative hypothesis. Since we use MORAN'S I as an exploratory tool before specifying our spatial models, this can actually be seen as an advantage, since we do not have to consider a single alternative hypothesis and instead explore several possible forms of spatial correlation. (Arbia, 2006).

[19] w_{ii} are usually set to 0, otherwise each site would be a neighbor of itself.

[20] The observed values of the variable are randomly assigned to the regions.

based on *adjacency*, i.e., we consider second- and third-order neighbors.[21] Our second definition is based on a *threshold-distance*, i.e., sites *i* and *j* are neighbors if the distance between their centroids is smaller than the threshold distance *d*. We choose *d* to equal 50 km, 68.2km, and 150 km. We include 68.2 km because it is the smallest distance—as such, each county has at least one neighbor. The last definition is based on a *Nearest-Neighbor algorithm* for 3, 5 and 7 nearest-neighbors. Note that all matrices are binary and that the adjacency- and threshold-based definitions lead to symmetric matrices.

[Insert Table 3 here]

Table 3 reports the number of non-zero links (i.e., connections between sites), the value for MORAN'S I and the empirical *p*-value. A comparison of the number of links shows that the Nearest-Neighbor matrices define very small neighborhoods, whereas distance-based matrices result in larger neighborhoods. In general, MORAN'S I is not very high but all values, with the exception of the third-order adjacency matrix, are significant. We also observe that the value of *I* decreases with the size of the neighborhood. The highest value is obtained for the 3-Nearest-Neighbors matrix, whereas the distance-based and higher-order adjacency matrices result in considerably smaller values. Our preferred first-order adjacency matrix shows a comparably high value of *I*, which is highly significant. This definition also results in small neighborhoods, although the number of evaluated neighborhood relationships is higher than for the 3-Nearest-Neighbors matrix.

Thus we conclude that the spatial dependence between regional health effects occurs on a small, local scale. All in all, we reject the hypothesis of independent regional effects and employ a first-order adjacency matrix for the spatial models as described in Section 3.

Local Indicators of Spatial Association (LISAs)

MORAN'S I is a measure of global spatial autocorrelation. In contrast, LOCAL INDICATORS OF SPATIAL ASSOCIATION (LISAS) are calculated for each region individually and can be used to test the significance of the spatial association for a specific county. We calculate local *I*'s as proposed by Anselin (1995). Again, we use a Monte Carlo permutation test to obtain pseudo *p*-values. The *p*-values for each county are depicted in Figure 4. The map corresponds to an adjacency-based neighborhood matrix. The classes correspond to different types of clusters. They are formed according to the significance of the spatial association, the sign of the regional effect as well as the sign of the spatial correlation. "High-high" stands for counties with positive regional effects and positive spatial correlations, i.e., the regional effects of its neighbors are also positive (Anselin, 2005). We see that for most clusters, the spatial association is significant on a local level.

---

[21] Second-order neighbors are sites that share a common border with first-order neighbors of site *i*, but not with site *i* itself ("neighbors of neighbors").

[Insert Figure 4 about here]

4.2 Model 2: Intrinsic Conditional Autoregressive Regional Effects (ICAR)

In the next step, we estimate the ICAR-Model to obtain smoothed estimates of the regional effects (see Section 3.1.2). Model 2 produces a more distinct spatial pattern and is a model with smoother estimates.

The individual-level coefficient estimates of Model 2 are very robust as compared to Model 1a.[22] Not surprisingly, the county-level variables show some variation across the two models, which is due to the different assumptions upon the spatial structure. Therefore, in models 3 and 4, we improve upon the modeling of the spatial representation before interpreting the parameters of the regional covariates.[23]

The map in Figure 5 shows the regional effects for ***Model 2*** ("ICAR model based on adjacency").

[Insert Figure 5 about here]

First, it should be noted that the strength of the regional effects is even larger than in the URE-Model (Model 1a). Here, the regional effects vary between -3 and 3.3 as compared to -2.5 and 2.7 in Model 1a.

Second, the spatial pattern displayed in Figure 3 is very similar to the spatial pattern produced by Model 1a in Figure 3, but considerably smoother. Both maps show several large clusters and suggest that there exists a pattern in the spatial distribution.

Third, in West Germany, unobserved but systematic county-level effects show a strong positive association with residents' health. Contrarily, in East Germany, county-level effects show a strong negative association with residents' health. Remember that we control for a rich set of socio-economic individual background characteristics such as age, gender, marital, and employment status as well as health-behavior and the degree of health care utilization. In addition, ten county-level covariates net out persistent differences across counties due to unemployment, urbanization and population density. Given this modeling approach, it is surprising and staggering that we still find a clear East-West health pattern, 20 years after the German reunification.

4.3. Model 3: Convolution Prior (CP)

---

[22] These estimates are available upon request.

[23] A possible source of this sensitivity could be omitted county-level variables, which introduce spatial correlation. Cressie (1993) points out that modeling spatial autocorrelation can mitigate this problem.

Now, we run Model 3 using a convolution prior as described in Section 3.1.3. This model has the advantage that it can distinguish between unstructured (i.e., random) regional effects and spatially dependent regional effects. If we plot the overall regional effect, the resulting maps look very similar to the ICAR-Model (Model 2). However, the values of the spatially dependent random effects appear to be significantly smoother. Figure 6 illustrates the results of Model 3.

[Insert Figure 6 about here]

Figure 6 presents a map of the spatially dependent part of the regional effect φ_s. We observe a prominent cluster of high values in the Northwest (Lower Saxony, Schleswig-Holstein) and clusters of lower values in the East (Bavaria, Thuringia and Brandenburg). This confirms our findings above and reinforces a clear East-West German spatial health pattern that is likely to be a legacy of the 40 year long division of Germany.[24]

[Insert Figure 7 about here]

Figure 7 shows a LISA cluster map for Model 3. The counties with significant local *I* are color-coded with respect to the type of spatial correlation. We identify three clusters of high values, mainly in the Northwestern part of Germany and three clusters of low values in the Eastern part. This again supports the notion of an East-West trend in the spatially dependent component. Deviations from this trend in Eastern Bavaria and Western Mecklenburg-West Pomerania are likely due to spillover effects.

The finding of low regional health patterns in regions with former communist governments during the Cold War is consistent with the existing literature. For example Baltagi et al. (2012) and Bonneux et al. (2010) report significantly lower life expectancy in Eastern Europe, specifically Hungary, Czech Republic, Slovak Republic and Poland, when compared to Western Europe. In addition, Baltagi et al. (2012) find significant spillover effects in the healthcare production process across neighboring countries. Treurniet et al. (2004) report higher avoidable mortality rates for Hungary and the Czech Republic as compared to Western Europe between 1980 and 1997. However, we are not aware of any

---

[24] Note that, theoretically, there could be a plenitude of reasons for this distinct pattern. However, all of them could at least be interpreted as an indirect consequence of the 40 year long division of Germany. For example, it has been shown that unemployment has a causal impact on individuals' health and well-being (Sullivan and van Wachter, 2009; Kassenboehmer and Haisken-DeNew, 2009). The structurally higher unemployment rate in East Germany could be cited as one potential explanation for the East-West health differential. However, note that we control for the individual employment status as well as the county-level unemployment rate. Even in East Germany, county-level unemployment rates vary tremendously from 6.9 to 26.2 percent. Another potential explanation might refer to migration. Since the German reunification, about 3 million primarily young and healthy East Germans migrated to West Germany. However, many of those have returned to their home states since the economic conditions have improved considerably in last years
Furthermore, according to the SOEP data in our sample, only 6.6 percent or 962 individuals living in West Germany in 2006 were born in East Germany. If we subtract these individuals from the West German sample, the average SF12 decreases only slightly from 50.08 to 50.05. Similarly, if we add individuals who migrated from East to West Germany to the East German sample, the average SF12 increases only slightly from 49.01 to 49.2. All these differences are not statistically significant. Thus it is very unlikely that this marked difference has been caused by migration.

paper that spatially models and compares the health effects within Germany—a country with a common history, language, and culture that had been divided for 40 years.

4.4 Model 4: Spatiotemporal Convolution Prior (SCP)

Finally we estimate the Spatiotemporal Convolution Prior (SCP) Model in order to investigate changes in the spatial pattern over time. The results are in Table 4. First, the parameter estimates of the individual covariates are robust when we compare them to the findings from our first model, the URE-Model in Table 2.

Second, most of the county-level covariates are not significantly different from zero and are similar in size to the findings from our first URE-Model in Table 2. Third, among county-level covariates, the county-level unemployment rate shows by far the strongest association with individual-level health. In the SCP-Model, a 1 percentage point higher unemployment rate is associated with a reduction in health by 0.06 points. This equals an age effect of 1 year. This point estimate is highly significant at the 1 percent level. The second most important county-level correlate of individual health is the average household income in the county, which is significant at the 10 percent level. Inhabitants of counties with a €1,000 higher household income have on average 0.7 points more on their SF12 scale. This equals an age effect of 10 years.

[Insert Table 4 about here]

Figure 8 illustrates the spatial patterns for each year. Note that only the spatially dependent part of the regional effect is displayed. First of all, the size of the effects varies across years, but they largely overlap: in 2006, the values fall between -1.7 and 2, in 2008, they fall between -1.3 and 0.9 and in 2010, they fall between -1.6 and 1.

[Insert Figure 8 about here]

Second, the spatial pattern itself varies slightly across time. However, since the majority of the regional effects are not significant, this is probably just random variation. In those counties with an especially small sample size, the results are also highly sensitive to panel attrition. Even more importantly, the clusters with significant spatial dependence, e.g., the regions in the Northwest (Hanover), Southwest (Rhineland-Palatinate), Center (Thuringia) and Southeast (Bavaria), are stable across the years. Maps of the pseudo p-values of the regional effects[25] (φ_{st}) show that the regional effect in these clusters is significant in all years, whereas most of the effects that change across the years are insignificant.

---

[25] The figures are available upon request.

Finally, to explicitly test for the East-West-German health pattern, we include an East-West dummy variable in Model 4. The dummy is “1” for counties in pre-1989 communist East Germany and “0” for counties in capitalist West Germany. The highly significant coefficient estimate equals -0.79 and represents one of the strongest health predictors. Its impact is comparable to an age impact of more than half a standard deviation or up to 10 life years.

4.4. Model Fit and Robustness

We use the Deviance Information Criterion (DIC) introduced by Spiegelhalter et al. (2002) to compare our candidate models. The DIC is a criterion for model selection similar to the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), i.e., it trades off model fit[26] and complexity[27]. As with AIC and BIC, the model with the lowest DIC should be preferred. In contrast to AIC and BIC, the DIC is valid for hierarchical models and can be easily computed as a by-product of MCMC methods. However, the scale of the DIC has no meaning by itself since it includes a term that depends solely on the data. Therefore, only differences between DIC can be interpreted. However, there is no distinct rule when differences in DIC are considered significant. We use the rule of thumb suggested by Spiegelhalter et al. (2002) who suggest that differences larger or equal to 10 should be considered significant.

The results in Table 5 indicate that the ICAR-Model 2 and the Convolution Prior-Model 3 perform less well, both in terms of model fit and complexity. This is no surprise: we impose additional structural assumptions; hence we obtain more complex models. A comparison with Model 4 shows that this Spatiotemporal Model offers a much better model fit than the purely spatial models (see Table 5). This suggests the existence of space-time interactions. The gain in model fit more than compensates for the higher complexity of Model 4. Surprisingly, the model fit (Dbar) is even better than for the Unstructured Regional Effects Model (Model 1a). This implies that our Model 4 should be preferred among all estimated models. Although the DIC for the SCP-Model 4 is larger than the DIC for the URE-Model 1a, this effect is solely driven by the complexity of the model (pD).

[Insert Table 5 about here]

Table 5 also gives the estimated deviance, the BIC and the R² for all models. The BIC shows a different pattern than the DIC; however it should be noted that the BIC measures model complexity solely through a function of the nominal number of parameters. This means it does not account for the more efficient estimation of the individual and regional effects, i.e. the “borrowing of strength”.

---

[26] Model fit is measured through the posterior expected deviance.

[27] Complexity refers to the effective number of parameters calculated as the difference between posterior expected deviance and the deviance at the posterior means.

The R² shows that our preferred SCP-Model 4 explains 46 percent of the observed overall variation in health status.

To assess the robustness of our results, we estimate two additional variants of the URE-Model: the “reduced covariates” variant does not include information on health care utilization, health-related behavior and employment status. The “full covariates” variant includes all individual covariates, but not the regional predictors. As can be inferred from Table 5, the model fit for Model 1b (reduced set of only individual-level covariates) is much worse than the URE-Model 1a, i.e., the excluded individual covariates explain a lot of the observed variation in health.

Model 1c (all individual-level covariates but no regional predictors) results in a very similar model fit as the URE-Model 1a including regional covariates. In other words, the regional predictors have little to no explanatory power. However, it should be noted that the “full covariates” variant includes regional effects, which might have picked up some of the variation explained by the regional covariates. A comparison of the pattern of the regional effects shows that our conclusions about the spatial structure are not affected by the exclusion of the regional covariates.

We also check for multicollinearity. The entries of the correlation matrix for the individual variables are not larger than 0.3. Among the regional covariates, the population density, the degree of urbanization and the area used for recreation are highly correlated. However, since the exclusion of the regional covariates does not affect our conclusions, multicollinearity does not pose a problem.

5 Conclusion

This paper combines representative individual-level household panel data and register county-level data to model and estimate spatial health effects on individual health. Methodologically, it makes use of hierarchical models and Bayesian methods.

In a first step, we examine regional effects for spatial associations using adjacency-based, distance-based and Nearest-Neighbor-based definitions of neighborhoods. In all cases, we find a highly significant spatial dependency of individual health. In the next step, we impose structural ICAR assumptions onto our models and re-estimate them, using an ICAR prior based on adjacency and a convolution prior based on adjacency. For both models we find a trade-off between model fit and smoother estimates. We find the model fit for our Spatiotemporal Model with separate regional effects for each year to be much better than the model fit for our basic Regional Effects Model without structural assumptions.

In terms of content, this paper shows that the regional association with residents’ health is systematic and strong. The general county-level predictor has the strongest impact among all individual- and county-level predictors considered. The regional impact is comparable to an age effect on health of up to 30 years: this means that—after considering a rich set of socio-demographic background information for an average respondent at the age of 50—the regional impact alone may shift a respondent's health either down to the health of a 20 year old or up to the health of a 80 year old.

Interestingly and surprisingly, even 20 years after the German reunification, we still find a clear East-West spatial health pattern. This finding could be interpreted as the long-lasting health effect of the 40 year long division of Germany into a Communist and a Capitalist Part. The long-lasting legacy of Communism on health equals an age effect of up to 10 life years.

Although most outcomes of interest reveal a clear spatially-dependent structure, applied researchers are reluctant to model these spatial dependencies explicitly. Using Germany and a quasi-objective and continuous measure of population health as an example, this paper provides a framework for applied researchers on how to use Hierarchical Bayes Models and estimation techniques. A next step would be to establish an identification strategy in order to test whether these regional effects have a causal interpretation or arise due to selection processes or omitted variable bias. Clearly, more research on and applications of modeling techniques for spatial patterns would be fruitful.

6 References

Andersen, H. H., Mühlbacher, A., Nübling, M., Schupp, J., and Wagner, G. G., 2007. Computation of standard values for physical and mental health scale scores using the SOEP version of sf12v2. *Schmollers Jahrb*, 127(1), 171-182.

Anselin, L., 1995. Local indicators of spatial association – LISA. *Geogr Anal*, 27(2), 93-115.

Anselin, L., 2005. *Exploring Spatial Data with GeoDa™: A Workbook*. Available at http://www2.geog.ucl.ac.uk/~jessletzbichler/GEOG2003/geodaworkbook.pdf, last accessed on May 1, 2012.

Arbia, G., 2006. *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Springer. Berlin Heidelberg.

Baltagi, B. H., and Moscone, F., 2010. Health care expenditure and income in the OECD reconsidered: Evidence from panel data, *Econ Model*, 27(4), 804-811.

Baltagi, B.H., Moscone, F. and Tosetti, E., 2012. Medical technology and the production of health care. *Empir Econ*, 42(2), 395-411.

Banerjee, S., Carlin, B. P. and Gelfand, A. E., 2004. *Hierarchical Modeling and Analysis for Spatial Data.* Chapman & Hall/CRC. Boca Raton.

Besag, J. , 1974. Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J R Stat Soc*, B 36, 192-236.

Besag, J. and Kooperberg, C., 1995. On conditional and intrinsic autoregressions. *Biometrika*, 82(4),733-746.

Besag, J., York, J.C. and Mollié, A., 1991. Bayesian image restoration with two applications in spatial statistics (with discussion). *Ann I Stat Math*, 43(1), 1-59.

Bonneux, L.G., Huisman, C. and de Beer, J.A., 2010. Mortality in 272 European regions, 2002–2004. An update. *Eur J Epidemiol*, 25(2), 77-85.

Brooks, S.P. and Gelman, A., 1998. General Methods for Monitoring Convergence of Iterative Simulations. *J Comput Graph Stat*, 7(4), 434-455.

Browning, C.R., Cagney, K.A. and Wen, M., 2003. Explaining variation in health status across space and time: implications for racial and ethnic disparities in self-rated health. *Soc Sci Med*, 57(7), 1221-1235.

Bryk, A.S. and Raudenbush, S.W., 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Sage Publications. Newbury Park. London. New Delhi.

Burchardi, K.B. and Hassan, T.A., 2013. The economic impact of social ties: evidence from German reunification. *Quarterly Journal of Economics*, forthcoming.

Burkhauser, R.V. and Cawley, J., 2008. Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *J Health Econ*, 27(2), 519-529.

Coneus, K. and Spieß, K.C., 2012. Pollution exposure and child health: Evidence for infants and toddlers in Germany. *J Health Econ*, 31(1), 180-196.

Congdon, P.D., 2010. *Applied Bayesian Hierarchical Methods*. Chapman & Hall/CRC. Boca Raton. London. New York.

Costa-Font, J. and Pons-Novell, J., 2007. Public health expenditures and spatial interactions in a decentralized national health system. *Health Econ*, 16(3), 291–306.

Cressie, N. A.C., 1993. *Statistics for Spatial Data.* Wiley. New York.

Cressie, N. A.C. and Wikle, C. K., 2011. *Statistics for Spatio-Temporal Data*. Wiley, Hoboken.

Dartmouth Atlas of Health Care, 2012. Available at www.dartmouthatlas.org, last accessed May 2, 2012.

Durlauf, S. N., 2003. Neighbourhood effects, in: J. V. Henderson / J.-F. Thisse(eds.), *Handbook of Regional and Urban Economics*, edition 1, volume 4, chapter 50, 2173-2242.

Federal Agency for Cartography and Geodesy, 2012. Maps of the Administrative Units of Germany, available at http://www.geodatenzentrum.de/geodaten/, last accessed February 6, 2013.

Federal Institute for Research on Building, Urban Affairs and Spatial Development, 2012. *INKAR 2011. Indicators and Maps on Spatial Development*, CD-ROM.

Felder, S. and Tauchmann, H., 2013. Federal state differentials in the efficiency of health production in Germany: an artifact of spatial dependence? *Eur J Health Econ*, 14(1), 21-39.

Frakes, M., 2013. The Impact of Medical Liability Standards on Regional Variations in Physician Behavior: Evidence from the Adoption of National-Standard Rules. *Am Econ Rev*, 103(1), 257-76.

Frick, J. R. and Ziebarth, N.R., 2013. Welfare-Related Health Inequality: Does the Choice of Measure Matter?, *Eur J Health Econ*, forthcoming.

Fuchs, V.R., McClellen, M.B. and Skinner, J.S., 2004. Area differences in utilization of medical care and mortality among U.S. elderly. In: Wise, D.A. (ed.): *Perspectives on the economics of aging*, University of Chicago Press, Chicago, 367-413.

Gelman, A. and Robert, C.P., 2013. “Not only defended but also applied”: The perceived absurdity of Bayesian inference. *Am Stat*, forthcoming.

Gelman, A. and Rubin, D.B., 1992. Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci*, 7(4), 457-511.

Hammersley, J. M. and Clifford, P., 1971. Markov field on finite graphs and lattices. Unpublished, available at http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf, last accessed on August 7, 2012.

Kassenboehmer, S. C. and Haisken-DeNew, J.P., 2009.You're Fired! The Causal Negative Effect of Entry Unemployment on Life Satisfaction, *Econ J*, 119(536), 448-462.

Knies, G. and & Spieß, C.K., 2007. Regional data in the German Socio-Economic Panel Study (SOEP), Data Documentation 17, DIW Berlin, German Institute for Economic Research, available at http://ideas.repec.org/e/pkn24.html, last accessed May 2, 2012.

Knies, G., Burgess, S. and Propper, C., 2008. Keeping up with the Schmidt's – an empirical test of relative deprivation theory in the neighbourhood context. *Schmollers Jahrb*, 128(1), 75-108.

Hox, J., 2002. Multilevel Analysis. Techniques and Applications. Lawrence Erlbaum Associates, Publishers. Mahwah (New Jersey), London.

Mollié, A., 1996. Bayesian mapping of disease. In A. Lawson, A. Biggeri, D. Bohning, E.Lesaffre, J.F. Viel, and R. Bertollini, (editors): *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*. Chapman and Hall. London. 359-379.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. *Biometrika*, 37(1), 17-23.

Moscone, F. ,Knapp, M. and Tosetti, E., 2007. Mental health expenditure in England: A spatial panel approach. *J Health Econ*, 26(4), 842-864.

Moscone, F. and Tosetti, E., 2010. Health expenditure and income in the United States. *Health Econ*, 19(12), 385-1403.

Mueller, I., Vounatsou, P., Allen, B.J. and Smith, T., 2001. Spatial patterns of child growth in Papua New Guinea and their relation to environment, diet, socio-economic status and subsistence activities. *Ann Hum Biol*, 28(3), 263-280.

Ocaña-Riola, R. and Mayoral-Cortés, J.M., 2010. Spatio-temporal trends of mortality in small areas of Southern Spain. *BMC Public Health*, 10:26.

OECD Project on Income Distribution and Poverty, 2009. What are equivalence scales?. Availabe online at http://www.oecd.org/dataoecd/61/52/35411111.pdf, last accessed on April 30, 2012.

Owens, G. M., 2008. Gender differences in health care expenditures, resource utilization, and quality of Care. *J Manag Care Pharm*, 14(3), 2-6.

Pesaran, M.H., 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4), 967-1012.

RAND, 1995. *RAND. User's Manual for the Medical Outcomes Study (MOS) Core Measures of Health-Related Quality of Life*. (by Hays, R. D., Sherbourne, C. D, Mazel, R.). RAND Monograph Report, Santa Monica, Available at: http://www.rand.org/pubs/monograph_reports/MR162.html, last accessed February 2, 2012

Riva, M.; Gauvin, L. and Barnett, T.A., 2007. Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. *J Epidemiol Commun H*, 61(10), 853-861.

Robert, C.P. and Casella, G., 2005. *Monte Carlo Statistical Methods*. 2nd Edition. Springer. New York.

Spiegelhalter, D.J., Best, N., Carlin, B.P. and van der Linde, A., 2002. Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc,* Series B, 64(4), 583-639.

Sullivan, D. and Wachter, T.v., 2009. Job Displacement and Mortality: An Analysis Using Administrative Data, *Q J Econ*, 124(3): 1265-1306.

Sundmacher, L., Gaskins, M., Hoffman, K, and Busse R., 2012. Spatial distribution of avoidable cancer mortality. *J Public Health*, 20(3), 279-288.

Treurniet, H.F., Boshuizen, H.C. and Harteloh, P.P.M., 2004. Avoidable mortality in Europe (1980-1997): a comparison of trends. *J Epidemiol Comm H,* 58(4), 290-295.

Voigtländer, S., Berger, U. and Razum, O., 2010. The impact of regional and neighbourhood deprivation on physical health in Germany: a multilevel study. *BMC Public Health*, 10:403.

Wennberg, J. and Gittelsohn, A., 1973. Small area variations in health care delivery, *Science,* 182 (4117), 1102-1108.

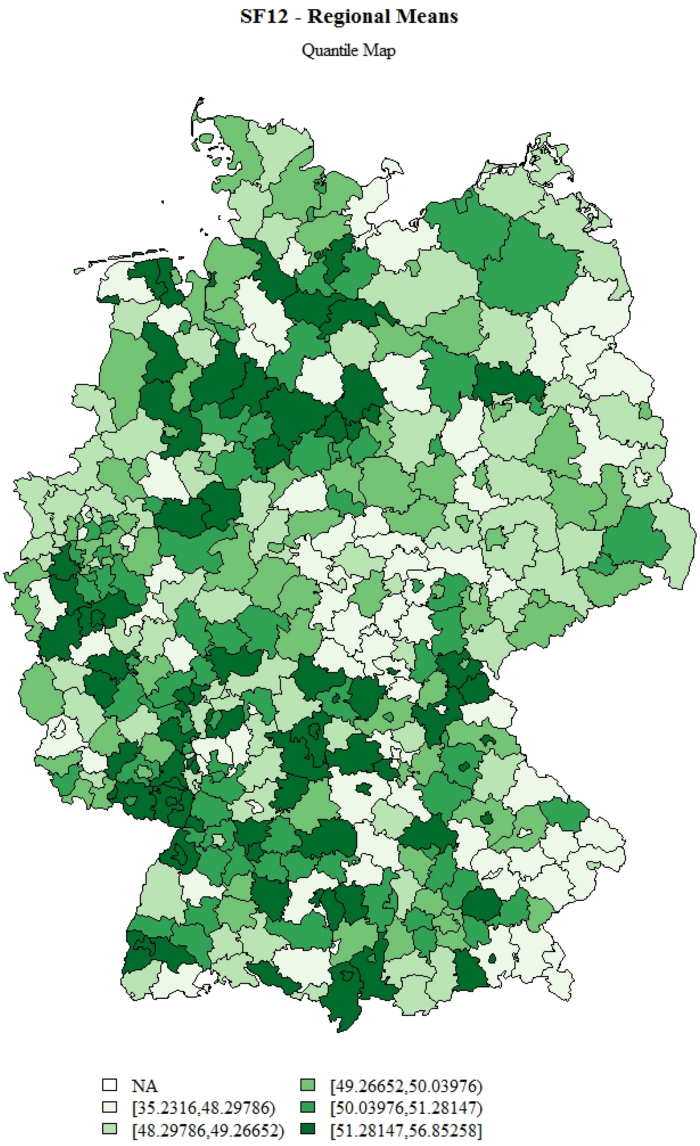
Wennberg, J. E.,Fisher, E. and Skinner, J.S., 2002. Geography and the debate over Medicare reform. *Health Affair* 21 (2, Suppl. Web Exclusive): W96–114.

Wagner, G. G., Frick, J. R., and Schupp, J., 2007. The German Socio-Economic Panel study (SOEP) - evolution, scope and enhancements. *Schmollers Jahrb*, 127(1), 139-169.

Ziebarth, N.R.and M.M Grabka, 2009. In Vino Pecunia? The Association Between Beverage-Specific Drinking Behavior and Wages. *J Labor Res*, 30(3), 219-244.

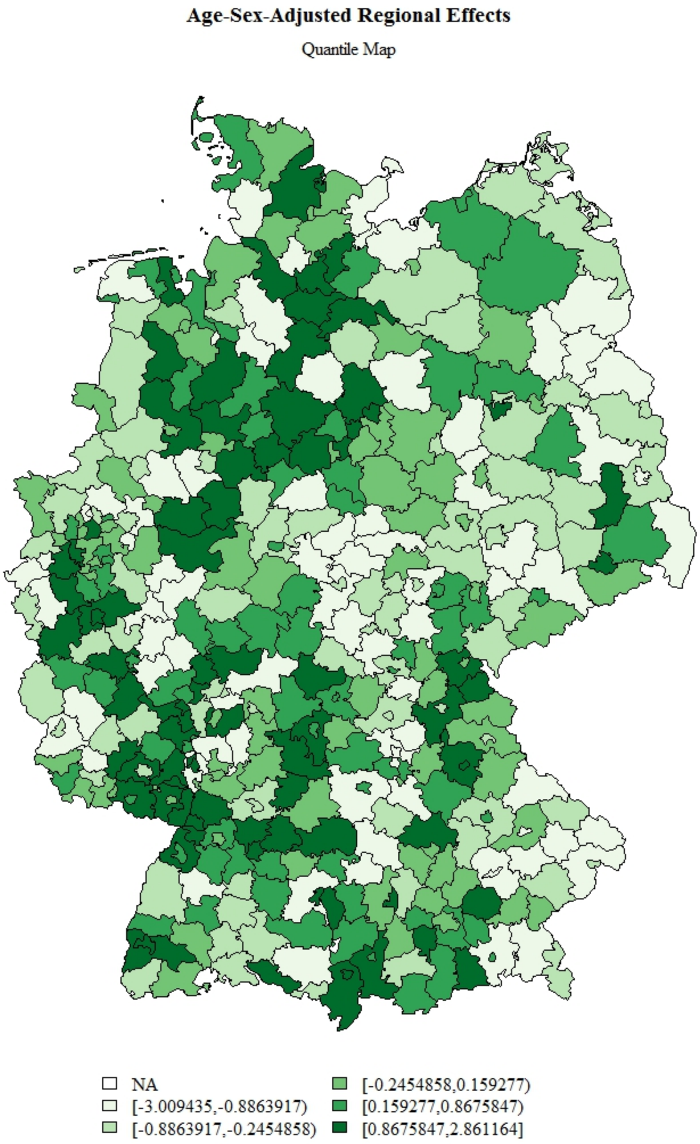
Tables and Figures

Figure 1



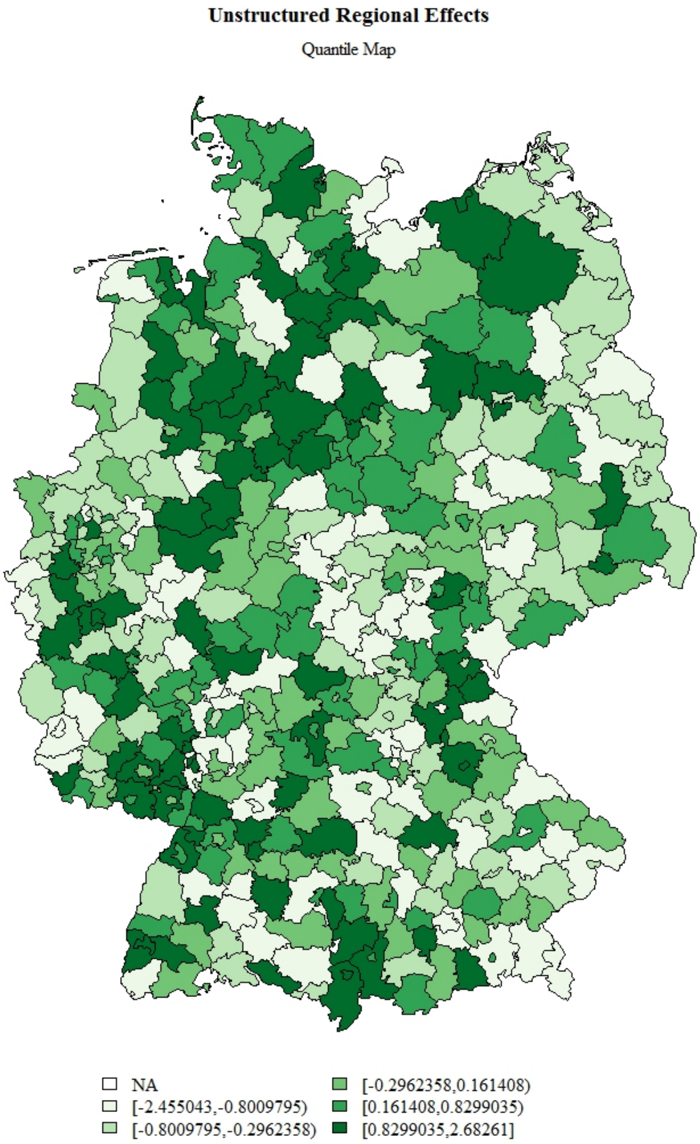
Source: SOEP v28, own calculations.Means of SF12 by county are displayed. The county borders reflect the territorial statuses as of January 1, 2012. The values of the variables are divided into five classes; the quintiles of the SF12 distribution serve as cutoff points. Each county is colored in a shade according to the class of the respective value of the variable. Lighter shades stand for lower values and darker shades for higher values. Areas without observations are depicted in white.

Figure 2



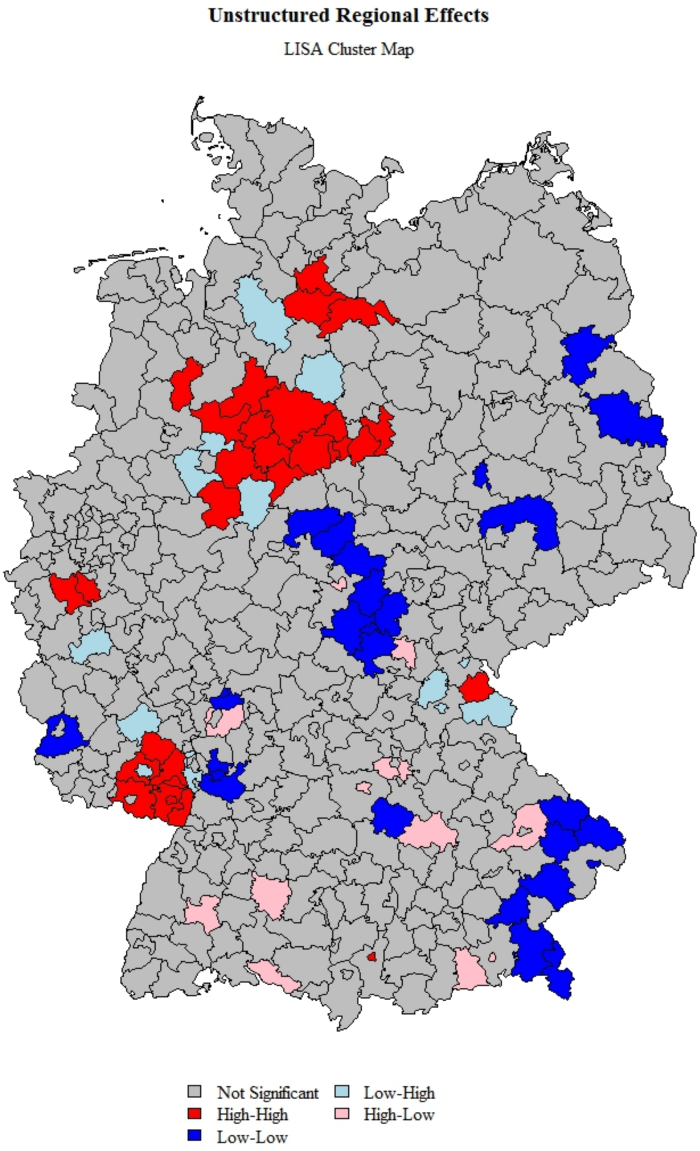
Source: SOEP v28, own calculations. County-level effects are obtained by regressing SF12 on age, gender, individual and a set of county dummies . The county borders reflect the territorial statuses as of January 1, 2012. The values of the SF12 variable are divided into five classes; the quintiles of the distribution serve as cutoff points. Each county is colored in a shade according to the class of the respective value of the variable. Lighter shades stand for lower values and darker shades for higher values. Areas without observations are depicted in white.

Figure 3



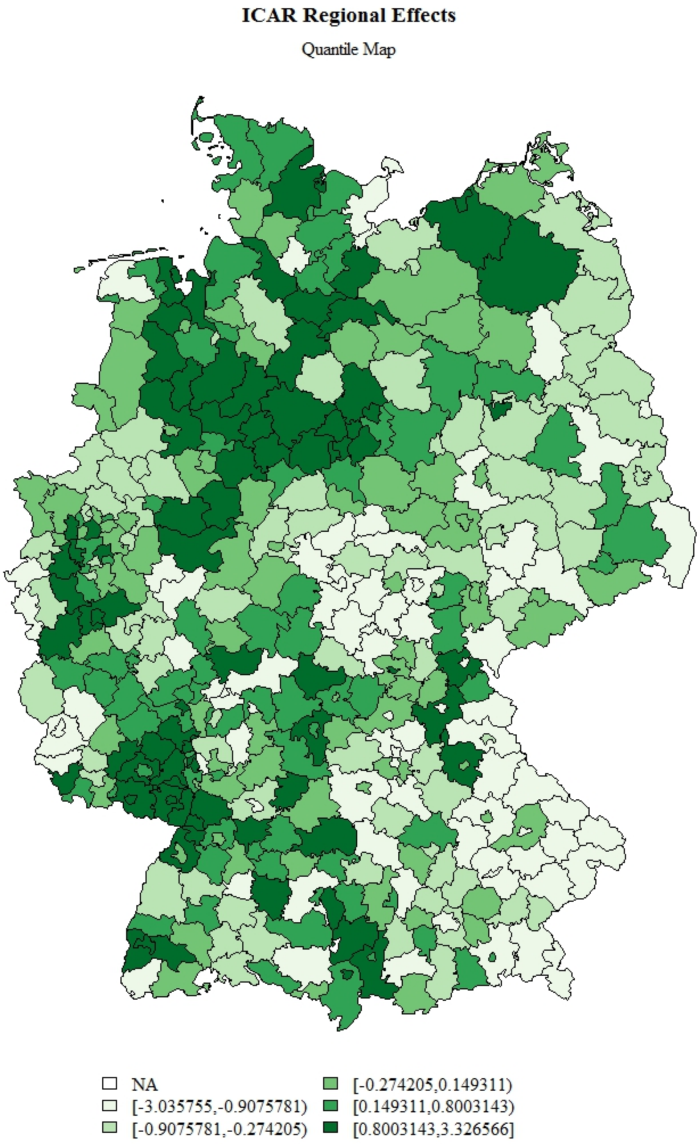
Source: SOEP v28, own calculations. Estimated county-level effects according to Model 1a (URE-Model) are displayed. The county borders reflect the territorial statuses as of January 1, 2012. The values of the SF12 variable are divided into five classes; the quintiles of the distribution serve as cutoff points. Each county is colored in a shade according to the class of the respective value of the variable. Lighter shades stand for lower values and darker shades for higher values. Areas without observations are depicted in white.

Figure 4



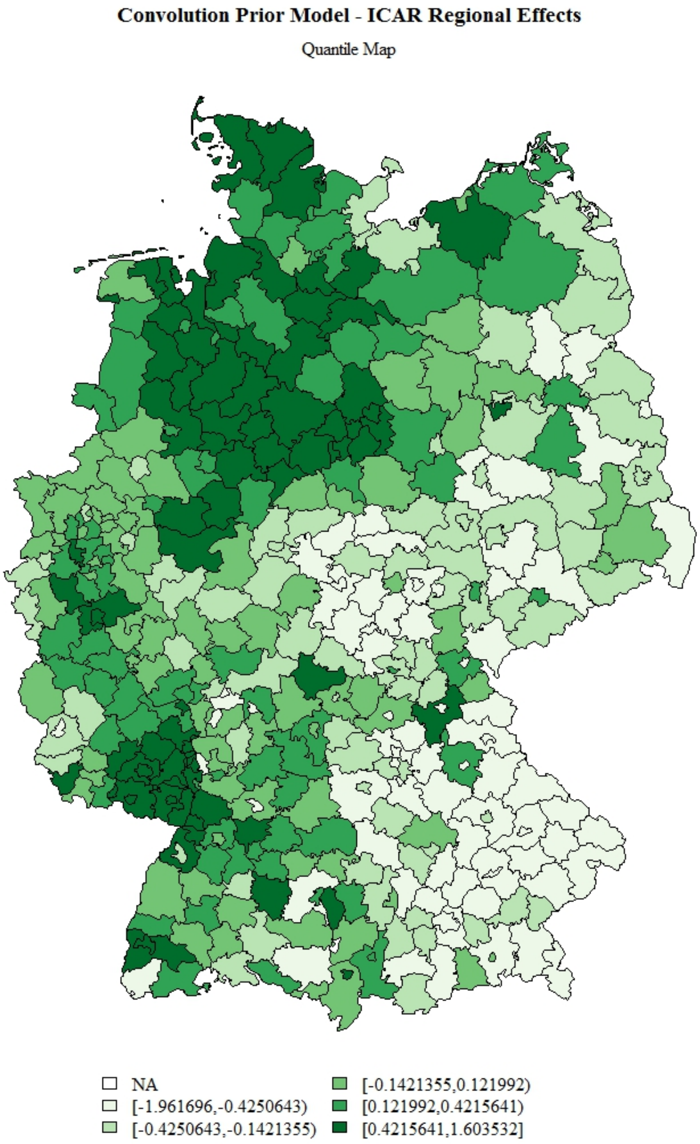
Source: SOEP v28, own calculations. This graph represents the spatial correlation according toModel 1a (URE-Model), estimated by local Moran's I with a first-order adjacency matrix. The areas are color-coded with respect to the sign of their l regional effect and the sign and significance of the spatial autocorrelation. “High-High” means that both the regional effect and the spatial autocorrelationare positive, i.e., the neighboring counties have also positive regional effects. In contrast, “High-Low” implies that the regional effect is positive, but the neighboring counties have negative regional effects, i.e. the spatial correlation is negative.

Figure 5



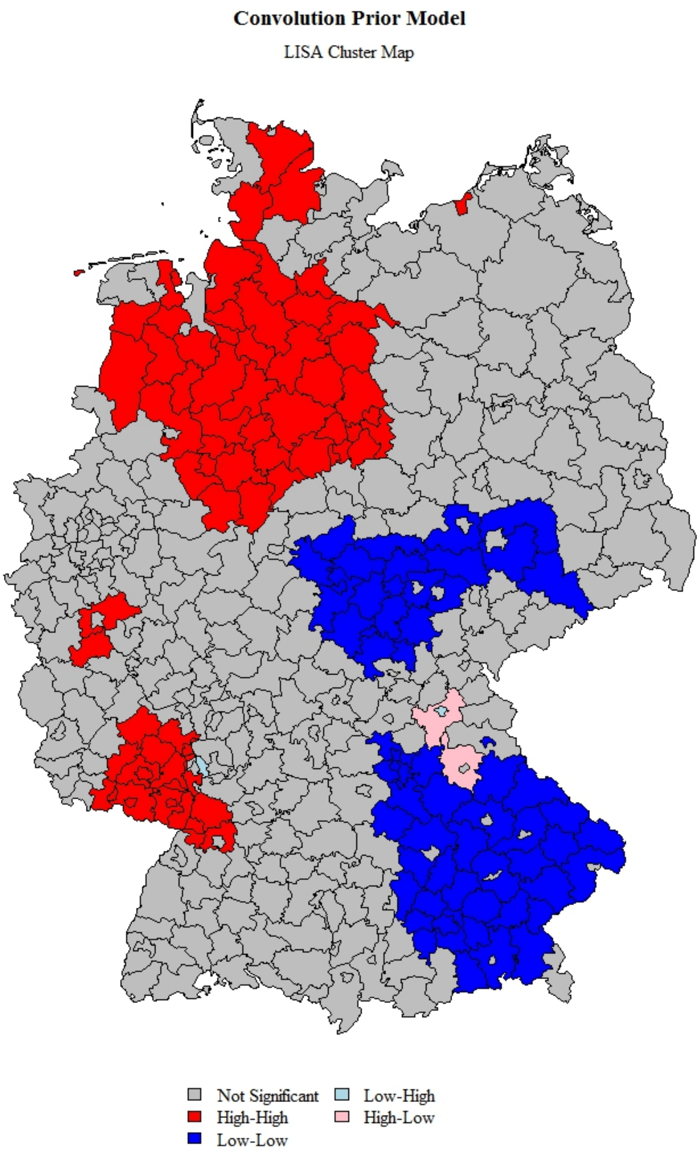
Source: SOEP v28, own calculations. This graph represents the spatial correlation according to Model 2 (ICAR-Model). The county borders reflect the territorial statuses as of January 1, 2012. The values of the SF12 variable are divided into five classes; the quintiles of the distribution serve as cutoff points. Each county is colored in a shade according to the class of the respective value of the variable. Lighter shades stand for lower values and darker shades for higher values. Areas without observations are depicted in white.

Figure 6



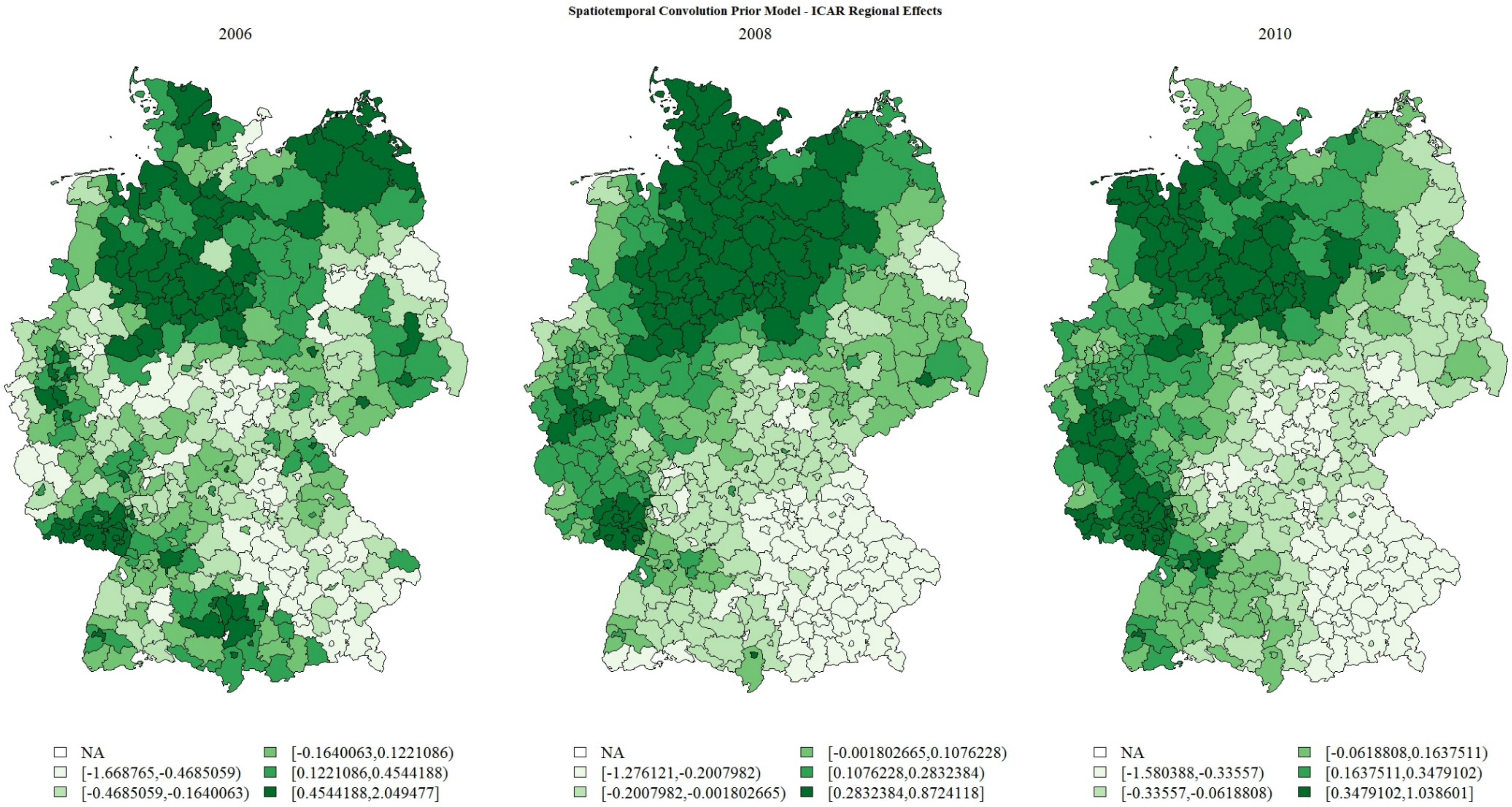
Source: SOEP v28, own calculations. This graph represents the spatial correlation according to Model 3 (CP-Model). Shown is only the spatially dependent part φ_s. The county borders reflect the territorial statuses as of January 1, 2012. The values of the SF12 variable are divided into five classes; the quintiles of the distribution serve as cutoff points. Each county is colored in a shade according to the class of the respective value of the variable. Lighter shades stand for lower values and darker shades for higher values. Areas without observations are depicted in white.

Figure 7



Source: SOEP v28, own calculations. This graph represents the spatial correlation according to Model 3 (CP-Model), estimated by local Moran's I with a first-order adjacency matrix. The spatially dependent regional effect φ_s is shown. is The areas are color-coded with respect to the sign of their regional effect and the sign and significance of the spatial autocorrelation. "High-High" means that both the regional effect and the spatial autocorrelation are positive, i.e., the neighboring counties have also positive regional effects. In contrast, "High-Low" implies that the regional effect is positive, but the neighboring counties have negative regional effects, i.e., the spatial correlation is negative.

Figure 8



Source: SOEP v28, own calculations.This graph represents the spatial correlation according to Model 4 (SCP-Model) for each year separately. Displayed is only the spatially dependent part φ_{st}. . The county borders reflect the territorial statuses as of January 1, 2012. The values of the SF12 variable are divided into five classes; the quintiles of the distribution serve as cutoff points. Each county is colored in a shade according to the class of the respective value of the variable. Lighter shades stand for lower values and darker shades for higher values. Areas without observations are depicted in white.

Table 1: Summary Statistics

Variable	Mean	Standard Deviation	Min	Max	N
Dependent Variable					
SF12 - generic health measure	49.73	7.12	22.33	61.62	54,734
A: Individual Characteristics					
Demographics					
Age	49.31	17.58	17	100	54,734
Number of Children under 14	0.36	0.75	0	8	54,734
Sex	0.48	0.50	0	1	54,734
Marital Status	0.60	0.49	0	1	54,734
Education & Labor Market Participation					
Monthly Household Income - Equivalence scale	1,776.74	1,263.56	0.00	66,666.00	54,734
Not working	0.43	0.50	0	1	54,734
University degree	0.21	0.41	0	1	54,734
No vocational training	0.23	0.42	0	1	54,734
Health-related Behavior					
No alcohol consumption	0.13	0.34	0	1	54,734
Regular alcohol consumption	0.17	0.38	0	1	54,734
Smoking	0.28	0.45	0	1	54,734
Body Mass Index (BMI)	25.96	4.60	12.59	73.05	54,734
Health Care Utilization					
Number of stays in hospital during the past year	0.16	0.62	0	42	54,734
Number of doctor visits during the past year	9.80	15.18	0	396	54,734

…continued on next page

…Table 1 continued

B: Regional Characteristics

Area size	891.73	721.58		35.70	5,811.47	1,200
Population density	521.92	676.67		38.62	4,355.28	1,200
Proportion of area used for settlement and transport	20.78	15.48		4.60	76.10	1,200
Proportion of area suitable for recreation	1.98	2.39		0.20	15.00	1,200
Cars per 1,000 inhabitants	524.81	57.10		316.10	880.60	1,200
Inhabitants per physician	684.82	183.44		255.20	1,224	1,200
Unemployment rate	9.02	4.51		1.90	26.20	1,200
Percentage of immigrants	7.25	4.50		0.70	25.80	1,200
Average available income per month	1,534	206		1,108	2,702	1,200
Number of overnight stays per inhabitant	4.75	5.68		0.00	43.30	1,200

Sources: SOEP v28, INKAR 2011.

Table 2: Coefficient Estimates for the Covariates in the Unstructured Regional Effects model 1a

Individual Variables	**Mean**	**Rescaled**	**Significance**	**S.E**	**2.5% Percentile**	**97.5% Percentile**
Age	-1.28614	-0.07318	***	0.21973	-1.71253	-0.85139
Squared age	-0.76608	-0.00043	***	0.21782	-1.19953	-0.32943
Male gender	1.15463	-	***	0.07329	1.01200	1.29505
Marital status	0.45292	-	***	0.07475	0.30395	0.59975
Number of children	0.03134	0.04179		0.03295	-0.03253	0.09429
Household income	0.40063	0.00032	***	0.02942	0.34480	0.46010
Not working	-0.22421	-	***	0.06488	-0.34771	-0.09879
University degree	0.92998	-	***	0.09170	0.75272	1.10652
No vocational training	-0.49042	-	***	0.08668	-0.66135	-0.32025
No alcohol consumption	-1.15209	-	***	0.07846	-1.30900	-1.00200
Regular alcohol consumption	0.27163	-	***	0.06932	0.13580	0.40715
Smoking	-0.93692	-	***	0.06831	-1.07252	-0.80115
Number of hospital stays	-0.39360	-0.63484	***	0.02162	-0.43680	-0.34995
Number of doctor’s visits	-1.54285	-0.10162	***	0.02404	-1.58953	-1.49500
BMI	-0.58620	-0.12744	***	0.03465	-0.65426	-0.51850
Regional Variables	**Mean**	**Rescaled**	**Significance**	**S.E**	**2.5% Percentile**	**97.5% Percentile**
Population density	-0.34820	-0.00051		0.38277	-1.08500	0.40919
Area size	0.03382	0.00005		0.10444	-0.16450	0.24360
Unemployment rate	-0.15172	-0.03363	*	0.07959	-0.30705	0.00582
Overnight stays per inhabitant	0.01491	0.00262		0.06630	-0.11497	0.14507
Area used for settlement and transport	0.43681	0.02822		0.37855	-0.31600	1.16110
Average household income	-0.03855	-0.00019		0.10137	-0.23785	0.15916
Percentage of migrants	0.17928	0.03981		0.14995	-0.11715	0.47351
Area used for recreation	0.16984	0.07106		0.22085	-0.26198	0.58815
Cars per 1,000 Inhabitants	0.26274	0.00460	*	0.13626	0.00316	0.52831
Inhabitants per physician	-0.05119	-0.00028		0.10126	-0.24611	0.15271

…continued on next page

…Table 2 continued

Time effects						
Dummy 2006	0.50183	-	***	0.08470	0.33695	0.66626
Dummy 2008	0.43759	-	***	0.04889	0.34304	0.53536
Constant	48.80483	-		0.11614	48.58000	49.04000
σ_C	4.64475	-		0.03011	4.58600	4.70400
σ_B	1.15744	-		0.06381	1.03500	1.28805
σ	3.87646	-		0.01589	3.84500	3.90800

Source: SOEP v28, INKAR 2011, own calculations. Number of observations N=54,734. Number of draws from the Posterior Distribution D=4,500. Column 1 gives the posterior mean (i.e., the point estimate) for the 15 individual predictors, the 10 county-level covariates, the aggregate time effects, the constant and the standard deviations of the dependent variable, (σ), the individual effects, (σ_C) , and the regional effects, (σ_B). Note that the continuous variables are standardized and therefore the measurement units are standard deviations. Column 2 gives a rescaled coefficient estimate, i.e., the effect size corresponding to the original measurement unit (left blank for dummy variables). Column 3 gives the posterior probability that the parameter has a different sign than the point estimate: ***=<0.1%, **=<1%, *<5% and .=<10%. This should not be confused with p-values and Null Hypothesis Significance Tests (see Section 3.2).. Column 4 gives the standard errors. Columns 5-6 give the 95%-Equal Tail Credibility Interval around the median.

Table 3: Moran's I for Different Definitions of Neighborhood

Weights	**Links**	**I**	**p-value**
first-order adjacency	2084	0.18	0.00
second-order adjacency	4880	0.04	0.02
third-order adjacency	7580	0.01	0.20
threshold distance - 50 km	3694	0.14	0.00
threshold distance - 68.2 km	27092	0.10	0.00
threshold distance - 150 km	8410	0.04	0.00
3 nearest neighbors	1201	0.19	0.00
5 nearest neighbors	2000	0.16	0.00
7 nearest neighbors	2798	0.15	0.00

Source: SOEP v28, INKAR 2011, own calculations; Moran's I for nine different weight matrices are displayed. Column 1 gives the number of non-zero links, i.e., spatial connections between sites. Column 2 gives the estimate of Moran's I and Column 3 the corresponding empirical p-value, which is derived by a Monte-Carlo permutation test with 10,000 permutations.

Table 4: Coefficient Estimates for the Spatiotemporal Convolution Prior Model (Model 4)

Individual Variables	**Mean**	**Rescaled**	**Significance**	**S.E.**	**2.5% Percentile**	**97.5% Percentile**
Age	-1.31845	-0.07502	***	0.22216	-1.74505	-0.87964
Squared age	-0.72661	-0.00041	***	0.22116	-1.16100	-0.30579
Male gender	1.15315	-	***	0.07353	1.01100	1.29600
Marital status	0.45354	-	***	0.07562	0.30253	0.60046
Number of children	0.04076	0.05434		0.03399	-0.02592	0.10550
Household income	0.40632	0.00032	***	0.03050	0.34630	0.46680
Not working	-0.22098	-	***	0.06611	-0.34926	-0.08959
University degree	0.94829	-	***	0.09365	0.76884	1.13400
No vocational training	-0.48708	-	***	0.08601	-0.65340	-0.31753
No alcohol consumption	-1.13863	-	***	0.08060	-1.29500	-0.98110
Regular alcohol consumption	0.29659	-	***	0.07199	0.15875	0.43975
Smoking	-0.93549	-	***	0.06751	-1.06700	-0.80338
Number of hospital stays	-0.39267	-0.63333	***	0.02206	-0.43500	-0.34935
Number of doctor visits	-1.53342	-0.10100	***	0.02458	-1.58053	-1.48400
BMI	-0.59296	-0.12890	***	0.03478	-0.66155	-0.52605
Regional Variables						
Population density	-0.12437	-0.00018		0.23794	-0.59186	0.34502
Area size	0.02590	0.00004		0.07055	-0.11696	0.16090
Unemployment rate	-0.26264	-0.05822	**	0.09601	-0.45265	-0.08067
Overnight stays per inhabitant	0.04596	0.00809		0.04654	-0.04375	0.13840
Area used for settlement and transport	0.40089	0.02590	.	0.25923	-0.11016	0.91161
Average household income	0.14007	0.00068	*	0.07974	-0.01849	0.29135
Percentage of migrants	-0.00837	-0.00186		0.11039	-0.22330	0.21206
Area used for recreation	0.01464	0.00612		0.14160	-0.26128	0.28741
Cars per 1,000 inhabitants	0.12540	0.00220	.	0.09220	-0.05458	0.30590
Inhabitants per physician	-0.03260	-0.00018		0.07264	-0.17587	0.10881

…continued on next page

…Table 4 continued

Time effects						
Dummy 2006	0.66759	-	***	0.10636	0.46214	0.88351
Dummy 2008	0.44027	-	***	0.07733	0.29075	0.59506
Constant	48.72238	-	***	0.10723	48.51000	48.94000
σ_C	4.69884	-	***	0.03098	4.63700	4.76100
σ_B / σ_ω	*Results for each year available upon request*					
σ_ϕ	*Results for each year available upon request*					
σ	3.85114	-	***	0.01556	3.82000	3.88100

Source: SOEP v28, INKAR 2011, own calculations; number of observations N=54,723. Number of draws from the posterior distribution D=4,500. Column 1 gives the posterior mean (i.e., the point estimate) for the 15 individual predictors, the 10 county-level covariates, the aggregate time effects, the constant and the standard deviations of the dependent variable (σ) and the individual effects (σ_B). Keep in mind that the continuous variables are standardized and therefore the measurement units are standard deviations. Column 2 gives a rescaled coefficient estimate, i.e., the effect size corresponding to the original measurement unit (left blank for dummy variables). Column 3 gives the posterior probability that the parameter has a different sign than the point estimate: ***=<0.1%, **=<1%, *<5% and .=<10%. This should not be confused with p-values and null hypothesis significance tests (see Section 3.2). Column 4 gives the standard errors. Columns 5 and 6 give the 95%-Equal Tail Credibility Interval around the median.

Table 5: Model Selection Criteria for all Candidate Models

Model	Dbar	pD	DIC	Deviance	BIC	R²
Age-Sex-Adjusted Regional Effects	305,969	18,462.6	324,432	305,974.8	418,835.4	0.44388
URE-Model 1a	303,642	17,593.8	321,236	303,641.4	416,606.2	0.45555
URE-Model 1b (reduced individual-level covariate set)	306,220	18,280.2	324,500	306,219.2	419,103.4	0.44258
URE-Model 1c (no regional-level covariates)	303,644	17,592.5	321,237	303,646.3	416,563.7	0.45556
ICAR-Model 2	304,630	18,554.2	323,184	304,629.8	417,594.6	0.45544
CP-Model 3	304,617	18,538.5	323,156	304,619.4	419,484.2	0.45537
SCP-Model 4	303,216	18,451.6	321,667	303,221.8	425,686.8	0.45911

Source: SOEP v28, INKAR 2011, own calculations. Column 1 gives the measure of model fit used for the calculation of the DIC, i.e., a function of the deviance and the data. Column 2 gives the measure of model complexity used for the DIC, i.e., the number of effective parameters. Column 3 states the Deviance Information Criterion (DIC). Column 4 gives the estimated deviance for each model and Column 5 states the Bayesian Information Criterion (BIC). The last column gives the unadjusted R².