

Examining the Structure of Spatial Health Effects using Hierarchical Bayes Models

Peter Eibich¹, Nicolas R. Ziebarth²

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 15

¹ DIW Berlin, Socio-Economic Panel Study (SOEP), Berlin, Germany

peibich@diw.de

² Cornell University, Policy Analysis and Management (PAM), Ithaca (NY), USA

nrz2@cornell.edu

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

1 Motivation

The strong and systematic impact of regional and neighborhood effects on individual health is long recognized by applied researchers [1]. A large set of studies reports large regional differences in health and health care consumption [2]. Understanding these regional differences may help to identify their driving forces. It may also help to implement policy measures that eliminate health inequalities that can be attributed to neighborhood influences. However, the vast majority of studies neglect spatial dependencies between areas [3]. Not taking spatial

patterns in the empirical models into account means spatial independence is implicitly assumed. 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.

2 Empirical Specification

2.1 Data

In this paper we model and estimate the spatial health pattern in Germany. The dataset combines individual panel data from the German Socio-Economic Panel Study (SOEP) with administrative information at the county-level. As dependent variable, we use the generic, quasi-objective and continuous SF12 health measure. This measure is generated by a specific algorithm on the basis of 12 different health-related question. Although it is a self-reported measure of health, it is less subjective than the standard 5-categorical Self-Assessed Health (SAH) measure, since several different dimensions of health are taken into account. The SF12 can be interpreted as a single quasi-objective measure of an individual's health status.[4] We control for a rich set of individual- and county-level variables, which can potentially explain differences between individuals and regions. On the individual-level we include demographic factors, education and labor market participation, health behavior and health care utilization into our model. Additionally, we control for ten county-level predictors, e.g. the degree of urbanization, the unemployment rate or the average per capita income, which exhibit alot of variation.

2.2 Methods

Our econometric models combine three different methodological approaches - hierarchical models, spatial econometrics and Bayesian inference. A three-stage hierarchical model is used to estimate the effect of predictors from different hierarchical levels on individual endpoints and account for correlation within counties. Moran's I is used as a measure of spatial dependency between counties.¹ We employ three different definitions of neighborhood in order to explore the structure of the spatial dependence [5]. Furthermore Intrinsic Conditional Autoregressive (ICAR) models are specified to incorporate the spatial structure into our econometric models [6]. These models assume that the spatial dependence can be represented through a Markov Random Field, i.e. the value of an area depends only on the values of its neighbors. Bayesian methods are used to partly compensate for small sample sizes and the models are estimated by Markov Chain Monte Carlo (MCMC) methods.

¹It can be regarded as an analog to the lagged autocorrelation coefficient in Time Series Analysis.

We estimate four candidate models and compare them using the Deviance Information Criterion (DIC) [7].² The first model assumes spatial independence between counties and imposes a normal prior distributions with inflated variances on the county-level effects. The second and third model allow for spatial dependence through an ICAR and a convolution prior respectively. The fourth model allows for space-time interactions.

3 Results

Our findings reveal highly significant spatial dependencies. The strong and systematic county-level impact is comparable to an age effect on health of up to 31 years. Furthermore, the results show several significant clusters of positive and negative health effects. These clusters are stable across the 2006 to 2010 time period (see Figure 1 below). Even 20 years after the German reunification, we detect a clear spatial East-West health pattern that equals an age impact on health of up to 9 life years.

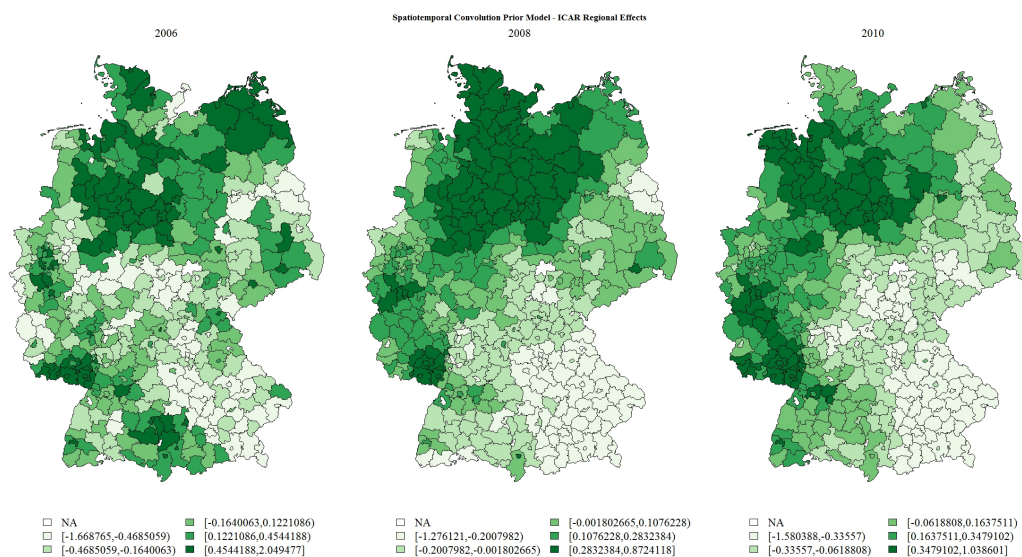


Figure 1: Source: SOEP v28, own calculations. This graph represents the spatial correlation according to Model 4 for each year separately. Displayed is only the spatially dependent part of the regional effect. The county borders reflect the territorial statuses as of January 1, 2012. The values of the 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.

²The DIC is a criterion for model selection similar to the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), i.e., it trades of model fit and model complexity. 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.

References

- [1] J. Wennberg, A. Gittelsohn. **Small area variations in health care delivery.** *Science*; 1973; 182(4117); pp. 1102-1108.
- [2] L. Sundmacher, M. Gaskins, K. Hoffman, R. Busse. **Spatial distribution of avoidable cancer mortality.** *Journal of Public Health*; 2012; 20(3); pp. 279-288.
- [3] M. Riva, L. Gauvin, T.A. Barnett. **Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998.** *Journal of Epidemiology and Community Health*; 2007; 61(10); pp. 853-861.
- [4] 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*; 1995; Available at: http://www.rand.org/pubs/monograph_reports/MR162.html, last accessed April 5, 2013.
- [5] P.A.P. Moran. **Notes on continuous stochastic phenomena.** *Biometrika*; 1950; 37(1); pp. 17-23.
- [6] J. Besag, J.C. York, A. Molli. **Bayesian image restoration with two applications in spatial statistics (with discussion).** *Annals of the Institute of Statistical Mathematics*; 1991; 43(1); pp. 1-59.
- [7] D.J. Spiegelhalter, N. Best, B.P. Carlin, A. van der Linde. **Bayesian measures of model complexity and fit (with discussion).** *Journal of the Royal Statistical Society*; 2002; 64(4); pp. 583-639.