University of Innsbruck

Working Papers
in
Economics and Statistics

Applications of Multilevel Structured Additive Regression Models to Insurance Data

Stefan Lang and Nikolaus Umlauf

2010-01

Applications of Multilevel Structured Additive Regression Models to Insurance Data

Stefan Lang[1] and Nikolaus Umlauf[1]

University of Innsbruck, Department of Statistics
Universitätsstraße 15, A-6020 Innsbruck, Austria,
stefan.lang@uibk.ac.at and nikolaus.umlauf@uibk.ac.at

Abstract. Models with structured additive predictor provide a very broad and rich framework for complex regression modeling. They can deal simultaneously with nonlinear covariate effects and time trends, unit- or cluster specific heterogeneity, spatial heterogeneity and complex interactions between covariates of different type. In this paper, we discuss a hierarchical version of regression models with structured additive predictor and its applications to insurance data. That is, the regression coefficients of a particular nonlinear term may obey another regression model with structured additive predictor. The proposed model may be regarded as an extended version of a multilevel model with nonlinear covariate terms in every level of the hierarchy. We describe several highly efficient MCMC sampling schemes that allow to estimate complex models with several hierarchy levels and a large number of observations typically within a couple of minutes. We demonstrate the usefulness of the approach with applications to insurance data.

Keywords: Bayesian hierarchical models, multilevel models, P-splines, spatial heterogeneity

1 Introduction

The last 10 to 15 years have seen enormous progress in Bayesian semiparametric regression modeling based on MCMC simulation for inference. A particularly broad and rich framework is provided by generalized structured additive regression (STAR) models introduced in Fahrmeir et al. (2004) and Brezger and Lang (2006). STAR models assume that, given covariates, the distribution of response observations y_i, $i = 1, \dots, n$, belongs to an exponential family. The conditional mean μ_i is linked to a semiparametric additive predictor η_i by $\mu_i = h(\eta_i)$ where $h(\cdot)$ is a known response function. The predictor η_i is of the form

$$\eta_i = f_1(z_{i1}) + \ldots + f_q(z_{iq}) + \mathbf{x}_i'\boldsymbol{\gamma}, \qquad i = 1, \dots, n, \tag{1}$$

where $f_1, \dots, f_q$ are possibly nonlinear functions of the covariates $z_1, \dots, z_q$ and $\mathbf{x}_i'\boldsymbol{\gamma}$ is the usual linear part of the model. In contrast to pure additive models the nonlinear functions f_j are not necessarily smooth functions of some continuous (one-dimensional) covariates z_j. Instead, a particular covariate may for example indicate a time scale, a spatial index denoting the region or district a certain observation pertains to, or a unit- or cluster-index. Moreover, z_j may be two- or even three dimensional in order to model interactions between covariates. Summarizing, the functions f_j comprise usual nonlinear effects of continuous covariates, time trends and seasonal effects, two dimensional surfaces, varying coefficient terms, cluster- and spatial effects.

The nonlinear effects in (1) are modeled by a basis functions approach, i.e. a particular nonlinear function f of covariate z is approximated by a linear combination of basis or indicator functions

$$f(z) = \sum_{k=1}^{K} \beta_k B_k(z). \tag{2}$$

The B_k's are known basis functions and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)'$ is a vector of unknown regression coefficients to be estimated. Defining the $n \times K$ design matrix $\mathbf{Z}$ with elements $\mathbf{Z}[i,k] = B_k(z_i)$, the vector $\mathbf{f} = (f(z_1), \dots, f(z_n))'$ of function evaluations can be written in matrix notation as $\mathbf{f} = \mathbf{Z}\boldsymbol{\beta}$. Accordingly, for the predictor (1) we obtain

$$\boldsymbol{\eta} = \mathbf{Z}_1\boldsymbol{\beta}_1 + \ldots + \mathbf{Z}_q\boldsymbol{\beta}_q + \mathbf{X}\boldsymbol{\gamma}. \tag{3}$$

In this paper we discuss a hierarchical or multilevel version of STAR models. That is the regression coefficients $\boldsymbol{\beta}_j$ of a term f_j may themselves obey a regression model with structured additive predictor, i.e.

$$\boldsymbol{\beta}_j = \boldsymbol{\eta}_j + \boldsymbol{\varepsilon}_j = \mathbf{Z}_{j1}\boldsymbol{\beta}_{j1} + \ldots + \mathbf{Z}_{jq_j}\boldsymbol{\beta}_{jq_j} + \mathbf{X}_j\boldsymbol{\gamma}_j + \boldsymbol{\varepsilon}_j, \tag{4}$$

where the terms $\mathbf{Z}_{j1}\boldsymbol{\beta}_{j1}, \ldots, \mathbf{Z}_{jq_j}\boldsymbol{\beta}_{jq_j}$ correspond to additional nonlinear functions $f_{j1}, \ldots, f_{jq_j}$, $\mathbf{X}_j\boldsymbol{\gamma}_j$ comprises additional linear effects, and $\boldsymbol{\varepsilon}_j \sim N(\mathbf{0}, \tau_j^2\mathbf{I})$ is a vector of i.i.d. Gaussian errors. A third or even higher levels in the hierarchy are possible by assuming that the second level regression parameters $\boldsymbol{\beta}_{jl}$, $l = 1, \ldots, q_j$, obey again a STAR model. In that sense, the model is composed of a hierarchy of complex structured additive regression models.

The typical application for hierarchical STAR models are multilevel data where a hierarchy of units or clusters grouped at different levels is given. One of the main aspects of the paper are applications of multilevel STAR models to insurance data. In a first example, we apply our methods to analyze the amount of loss and claim frequency for car insurance data from a German insurance company. In our analysis in section 4.1 we will distinguish three levels: policyholders (level-1) are nested in districts (level-2) and districts are nested in counties (level-3). Our second example analyzes time-space trends for health insurance data.

2 Priors for the regression coefficient

We distinguish two types of priors: "direct" or "basic" priors for the regression coefficients $\boldsymbol{\beta}_j$ (or $\boldsymbol{\beta}_{jl}$ in a second level equation) and compound priors (4). We first briefly describe the general form of "basic" priors in the next subsection. Subsection 2.2 shows how the basic priors can be used as building blocks for the compound priors.

2.1 General form of basic priors

In a frequentist setting, overfitting of a particular function $\mathbf{f} = \mathbf{Z}\boldsymbol{\beta}$ is avoided by defining a roughness penalty on the regression coefficients, see for instance Belitz and Lang (2008) in the context of structured additive regression. The standard are quadratic penalties of the form $\lambda\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta}$ where $\mathbf{K}$ is a penalty matrix. The penalty depends on the smoothing parameter λ that governs the amount of smoothness imposed on the function f.

In a Bayesian framework a standard smoothness prior is a (possibly improper) Gaussian prior of the form

$$p(\boldsymbol{\beta}|\tau^2) \propto \left(\frac{1}{\tau^2}\right)^{rk(\mathbf{K})/2} \exp\left(-\frac{1}{2\tau^2}\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta}\right) \cdot I(\mathbf{A}\boldsymbol{\beta} = \mathbf{0}), \qquad (5)$$

where $I(\cdot)$ is the indicator function. The key components of the prior are the penalty matrix $\mathbf{K}$, the variance parameter τ^2 and the constraint $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$.

The structure of the penalty or prior precision matrix $\mathbf{K}$ depends on the covariate type and on our prior assumptions about smoothness of f. Typically the penalty matrix in our examples is rank deficient, i.e. $rk(\mathbf{K}) < K$, resulting in a partially improper prior.

The amount of smoothness is governed by the variance parameter τ^2. A conjugate inverse Gamma prior is employed for τ^2 (as well as for the overall variance parameter σ^2 in models with Gaussian responses), i.e. $\tau^2 \sim IG(a, b)$ with small values such as $a = b = 0.001$ for the hyperparameters a and b resulting in an uninformative prior on the log scale. The smoothing parameter λ of the frequentist approach and the variance parameter τ^2 are connected by $\lambda = \sigma^2/\tau^2$.

The term $I(\mathbf{A}\boldsymbol{\beta} = \mathbf{0})$ imposes required identifiability constraints on the parameter vector. A straightforward choice is $\mathbf{A} = (1, \ldots, 1)$, i.e. the regression coefficients are centered around zero. A better choice in terms of interpretability and mixing of the resulting Markov chains is to use a weighted average of regression coefficients, i.e. $\mathbf{A} = (c_1, \ldots, c_K)$. As a standard we use $c_k = \sum_{i=1}^n B_k(z_i)$ resulting in the more natural constraint $\sum_{i=1}^n f(z_i) = 0$.

Specific examples for modeling nonlinear terms are one or two dimensional P-splines for nonlinear effects of continuous covariates, or Gaussian Markov random fields and Gaussian fields (kriging) for modeling spatial heterogeneity, see Brezger and Lang (2006) for details.

2.2 Compound priors

In the vast majority of cases a compound prior is used if a covariate $z_j \in \{1,\dots,K\}$ is a unit- or cluster index and z_{ij} indicates the cluster observation i pertains to. Then the design matrix $\mathbf{Z}_j$ is a $n \times K$ incidence matrix with $\mathbf{Z}_j[i,k] = 1$ if the i-th observation belongs to cluster k and zero else. The $K \times 1$ parameter vector $\boldsymbol{\beta}_j$ is the vector of regression parameters, i.e. the k-th element in $\boldsymbol{\beta}$ corresponds to the regression coefficient of the k-th cluster. Using the compound prior (4) we obtain an additive decomposition of the cluster specific effect. The covariates z_{jl}, $l = 1,\dots,q_j$, in (4) are cluster specific covariates with possible nonlinear cluster effect. By allowing a full STAR predictor (as in the level-1 equation) a rather complex decomposition of the cluster effect $\boldsymbol{\beta}_j$ including interactions is possible. A special case arises if cluster specific covariates are not available. Then the prior for $\boldsymbol{\beta}_j$ collapses to $\boldsymbol{\beta}_j = \boldsymbol{\varepsilon}_j \sim N(\tau_j^2 \mathbf{I})$ and we obtain a simple i.i.d. Gaussian cluster specific random effect with variance parameter τ_j^2.

Another special situation arises if the data are grouped according to some discrete geographical grid and the cluster index z_{ij} denotes the geographical region observation i pertains to. For instance, in our applications on insurance data in section 4 for every observation the district of the policyholders residence is given. Then the compound prior (4) models a complex spatial heterogeneity effect with possibly nonlinear effects of region specific covariates z_{jl}.

In a number of applications geographical information and spatial covariates are given at different resolutions. For instance, in our case studies on insurance problems, the districts (level-2) are nested within counties (level-3). This allows to model a spatial effect over two levels of the form

$$\begin{aligned}\boldsymbol{\beta}_j &= \mathbf{Z}_{j1}\boldsymbol{\beta}_{j1} + \mathbf{Z}_{j2}\boldsymbol{\beta}_{j2} + \dots + \boldsymbol{\varepsilon}_j,\\ \boldsymbol{\beta}_{j1} &= \mathbf{Z}_{j11}\boldsymbol{\beta}_{j11} + \mathbf{Z}_{j12}\boldsymbol{\beta}_{j12} + \dots + \boldsymbol{\varepsilon}_{j1}.\end{aligned}$$

Here, the first covariate z_{j1} in the district specific effect is another cluster indicator that indicates the county in which the districts are nested. Hence

$\mathbf{Z}_{j1}$ is another incidence matrix and $\boldsymbol{\beta}_{j1}$ is the vector of county specific effects modeled through the level-3 equation.

Other possibilities for compound priors can be found in Lang et al. (2010).

3 Sketch of MCMC Inference

In the following, we will describe a Gibbs sampler for models with Gaussian errors. The non-Gaussian case can be either traced back to the Gaussian case via data augmentation, see e.g Frühwirth-Schnatter et al. (2008), or is technically similar (Brezger and Lang, 2006).

For the sake of simplicity we restrict the presentation to a two level hierarchical model with one level-2 equation for the regression coefficients of the first term $\mathbf{Z}_1\boldsymbol{\beta}_1$. That is, the level-1 equation is $\mathbf{y} = \boldsymbol{\eta}+\boldsymbol{\varepsilon}$ with predictor (3) and errors $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{W}^{-1})$ with diagonal weight matrix $\mathbf{W} = diag(w_1, \dots, w_n)$. The level-2 equation is of the form (4) with $j = 1$.

The parameters are updated in blocks where each vector of regression coefficients $\boldsymbol{\beta}_j$ ($\boldsymbol{\beta}_{1l}$ in a second level of the hierarchy) of a particular term is updated in one (possibly large) block followed by updating the regression coefficients $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}_1$ of linear effects and the variance components τ_j^2, τ_{1l}^2, σ^2. The next subsection 3.1 sketches updates of regression coefficients $\boldsymbol{\beta}_j$, $\boldsymbol{\beta}_{1l}$ of nonlinear terms. Updates of the remaining parameters are straightforward.

3.1 Full conditionals for regression coefficients of nonlinear terms

The full conditionals for the regression coefficients $\boldsymbol{\beta}_1$ with the compound prior (4) and the coefficients $\boldsymbol{\beta}_j$, $j = 2, \dots, q$, $\boldsymbol{\beta}_{1l}$, $l = 1, \dots, q_1$ with the basic prior (5) are all multivariate Gaussian. The respective posterior precision $\boldsymbol{\Sigma}^{-1}$ and mean $\boldsymbol{\mu}$ is given by

$$
\begin{aligned}
\boldsymbol{\Sigma}^{-1} &= \tfrac{1}{\sigma^2}\left(\mathbf{Z}_1'\mathbf{W}\mathbf{Z}_1 + \tfrac{\sigma^2}{\tau_1^2}\mathbf{I}\right), \quad \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \tfrac{1}{\sigma^2}\mathbf{Z}_1'\mathbf{W}\mathbf{r} + \tfrac{1}{\tau_1^2}\boldsymbol{\eta}_1, \; (\boldsymbol{\beta}_1),\\
\boldsymbol{\Sigma}^{-1} &= \tfrac{1}{\sigma^2}\left(\mathbf{Z}_j'\mathbf{W}\mathbf{Z}_j + \tfrac{\sigma^2}{\tau_j^2}\mathbf{K}_j\right), \; \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \tfrac{1}{\sigma^2}\mathbf{Z}_j'\mathbf{W}\,\mathbf{r}, \qquad (\boldsymbol{\beta}_j), \qquad (6)\\
\boldsymbol{\Sigma}^{-1} &= \tfrac{1}{\tau_1^2}\left(\mathbf{Z}_{1l}'\mathbf{Z}_{1l} + \tfrac{\tau_1^2}{\tau_{1l}^2}\mathbf{K}_{1l}\right), \; \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} = \tfrac{1}{\tau_1^2}\mathbf{Z}_{1l}'\,\mathbf{r}_1, \qquad (\boldsymbol{\beta}_{1l}),
\end{aligned}
$$

where $\mathbf{r}$ is the current partial residual and $\mathbf{r}_1$ is the "partial residual" of the level-2 equation. More precisely, $\mathbf{r}_1 = \boldsymbol{\beta}_1 - \tilde{\boldsymbol{\eta}}_1$ and $\tilde{\boldsymbol{\eta}}_1$ is the predictor of the level-2 equation excluding the current effect of z_{1l}.

MCMC updates of the regression coefficients takes advantage of the following key features:

Sparsity: Design matrices $\mathbf{Z}_j, \mathbf{Z}_{1l}$ and penalty matrices $\mathbf{K}_j, \mathbf{K}_{1l}$ and with it cross products $\mathbf{Z}_j'\mathbf{W}\mathbf{Z}_j, \mathbf{Z}_{1l}'\mathbf{Z}_{1l}$ and posterior precision matrices in (6) are often sparse. The sparsity can be exploited for highly efficient computation of cross products, Cholesky decompositions of posterior precision matrices and for fast solving of relevant linear equation systems.

Reduced complexity in the second or third stage of the hierarchy: Updating the regression coefficients $\boldsymbol{\beta}_{1l}$, $l = 1, \ldots, q_1$, in the second (or third level) is done conditionally on the parameter vector $\boldsymbol{\beta}_1$. This facilitates updating the parameters for two reasons. First the number of "observations" in the level-2 equation is equal to the length of the vector $\boldsymbol{\beta}_1$ and therefore much less than the actual number of observations n. Second the full conditionals for $\boldsymbol{\beta}_{1l}$ are Gaussian regardless of the response distribution in the first level of the hierarchy.

Number of different observations smaller than sample size: In most cases the number m_j of different observations $z_{(1)}, \ldots, z_{(m_j)}$ in $\mathbf{Z}_j$ (or m_{1l} in $\mathbf{Z}_{1l}$ in the level-2 equation) is much smaller than the total number n of observations. The fact that $m_j \ll n$ may be utilized to considerably speed up computations of the cross products $\mathbf{Z}_j'\mathbf{W}\mathbf{Z}_j$, $\mathbf{Z}_{1l}'\mathbf{Z}_{1l}$, the vectors $\mathbf{Z}_j'\mathbf{W}\,\mathbf{r}$, $\mathbf{Z}_{1l}'\,\mathbf{r}_1$ and finally the updated vectors of function evaluations $\mathbf{f}_j = \mathbf{Z}_j\boldsymbol{\beta}_j$, $\mathbf{f}_{1l} = \mathbf{Z}_{1l}\boldsymbol{\beta}_{1l}$.

Full details of the MCMC techniques can be found in Lang et al. (2010).

3.2 Alternative sampling scheme based on a transformed parametrization

An alternative sampling scheme works with a transformed parametrization such that the cross product of the design matrix and the penalty matrix of a nonlinear term are diagonal resulting in a diagonal posterior precision matrix.

We describe the alternative parametrization for a particular nonlinear function f with design matrix $\mathbf{Z}$ and parameter vector $\boldsymbol{\beta}$ with general prior (5).

Let $\mathbf{Z}'\mathbf{W}\mathbf{Z} = \mathbf{R}\mathbf{R}'$ be the Cholesky decomposition of the cross product of the design matrix and let $\mathbf{Q}\mathbf{S}\mathbf{Q}'$ be the singular value decomposition of $\mathbf{R}^{-1}\mathbf{K}\mathbf{R}^{-T}$. The diagonal matrix $\mathbf{S} = diag(s_1, \ldots, s_K)$ contains the eigenvalues of $\mathbf{R}^{-1}\mathbf{K}\mathbf{R}^{-T}$ in ascending order. The columns of the orthogonal matrix $\mathbf{Q}$ contain the corresponding eigenvectors. Columns 1 through $rk(\mathbf{K})$ form a basis for the vector space spanned by the columns of $\mathbf{R}^{-1}\mathbf{K}\mathbf{R}^{-T}$. The remaining columns are a basis of the nullspace.

Then the decomposition $\boldsymbol{\beta} = \mathbf{R}^{-T}\mathbf{Q}\tilde{\boldsymbol{\beta}}$ yields

$$\mathbf{Z}\boldsymbol{\beta} = \mathbf{Z}\mathbf{R}^{-T}\mathbf{Q}\tilde{\boldsymbol{\beta}} = \tilde{\mathbf{Z}}\tilde{\boldsymbol{\beta}},$$

where the transformed design matrix $\tilde{\mathbf{Z}}$ is defined by $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{R}^{-T}\mathbf{Q}$.

We now obtain for the cross product

$$\mathbf{Z}'\mathbf{W}\mathbf{Z} = \mathbf{Q}'\mathbf{R}^{-1}\mathbf{Z}'\mathbf{W}\mathbf{Z}\mathbf{R}^{-T}\mathbf{Q} = \mathbf{Q}'\mathbf{Q} = \mathbf{I}$$

and for the penalty

$$\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}'\mathbf{Q}'\mathbf{R}^{-1}\mathbf{K}\mathbf{R}^{-T}\mathbf{Q}\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}'\mathbf{S}\tilde{\boldsymbol{\beta}}$$

with the new diagonal penalty matrix $\mathbf{S}$ given by the singular value decomposition of $\mathbf{R}^{-1}\mathbf{K}\mathbf{R}^{-T}$, see above.

Summarizing, we obtain the equivalent formulation $\mathbf{f} = \tilde{\mathbf{Z}}\tilde{\boldsymbol{\beta}}$ for the vector of function evaluations based on the transformed design matrix $\tilde{\mathbf{Z}}$ and the transformed parameter vector $\tilde{\boldsymbol{\beta}}$ with (possibly improper) Gaussian prior

$$\tilde{\boldsymbol{\beta}} \,|\, \tau^2 \sim N(\mathbf{0}, \tau^2\mathbf{S}^-).$$

The result of the transformation is that the prior precision or penalty matrix $\mathbf{S}$ is diagonal resulting in a diagonal posterior precision matrix. More specifically, the full conditional for $\tilde{\boldsymbol{\beta}}$ is Gaussian with k-th element μ_k, $k = 1, \ldots, K$, of the mean vector $\boldsymbol{\mu}$ given by

$$\mu_k = \frac{1}{1 + \lambda\, s_k} \cdot u_k,$$

where $\lambda = \sigma^2/\tau^2$ and u_k is the k-th element of the vector $\mathbf{u} = \tilde{\mathbf{Z}}'\mathbf{W}\mathbf{r}$ with $\mathbf{r}$ the partial residual. The covariance matrix $\boldsymbol{\Sigma}$ is diagonal with diagonal elements

$$\boldsymbol{\Sigma}[k,k] = \frac{\sigma^2}{1 + \lambda\, s_k}.$$

More details on this alternative sampling scheme can be found in Lang et al. (2010).

The main advantage of the transformation is that it provides fast MCMC inference even in situations where the posterior precision is relatively dense as is the case for many surface estimators. The prime example is a Gaussian random field (kriging) which is almost intractable in the standard parametrization.

4 Applications to insurance data

4.1 Car insurance data

The analyzed data set contains individual observations for a sample of policyholders with full comprehensive car insurance for one year. Regression analyzes for claim probabilities and amount of loss were carried out separately for different types of damage: traffic accidents, breakage of glass and theft. Here we report only results for claim probabilities of one type (specific type not mentioned to guarantee anonymity of the data source).

Claim probabilities were analyzed with a multilevel structured additive probit model $y_i \sim B(\pi_i)$ with three hierarchy levels for the probability $\pi_i = \Phi(\eta_i)$ that a damage occurred:

$$\begin{aligned}
\textbf{level-1} \quad \boldsymbol{\eta} &= \cdots + \mathbf{f}_1(nclaim) + \mathbf{f_2}(g) + \mathbf{f}_3(dist) \\
&= \cdots + \mathbf{Z}_1\boldsymbol{\beta}_1 + \mathbf{Z}_2\boldsymbol{\beta}_2 + \mathbf{Z}_3\boldsymbol{\beta}_3 \\
\textbf{level-2} \quad \boldsymbol{\beta}_2 &= \boldsymbol{\varepsilon}_2 \\
\textbf{level-2} \quad \boldsymbol{\beta}_3 &= \mathbf{f}_{31}(dist) + \mathbf{f}_{32}(county) + \mathbf{f}_{33}(dens) + \boldsymbol{\varepsilon}_3 \\
&= \mathbf{Z}_{31}\boldsymbol{\beta}_{31} + \mathbf{Z}_{32}\boldsymbol{\beta}_{32} + \mathbf{Z}_{33}\boldsymbol{\beta}_{33} + \boldsymbol{\varepsilon}_3 \\
\textbf{level-3} \quad \boldsymbol{\beta}_{32} &= \boldsymbol{\varepsilon}_{32}
\end{aligned}$$

The level-1 equation consists of a nonlinear function f_1 of the covariate "no-claims bonus" (*nclaim*) and of nonlinear effects of three other continuous covariates (indicated through the dots, results not shown to guarantee anonymity of the data provider). All nonlinear effects are modeled using P-splines. Additionally a random effect of the "car classification" (g) measured by scores from 10-40 and a spatial random effect of the districts (*dist*) in Germany is included. For "car classification" a simple i.i.d random effect with $\boldsymbol{\varepsilon}_2 \sim N(0, \sigma_2^2)$ is assumed, see the first level-2 equation. The spatial random effect is modeled through the other level-2 equation and is composed of a spatially correlated effect $\mathbf{f}_{31}$ using a Markov random fields prior, another spatial random effect of the counties, and a smooth nonlinear effect of the population density (*dens*). The "error term" in the district effect is a i.i.d random effect, i.e. $\boldsymbol{\varepsilon}_3 \sim N(0, \sigma_3^2)$. For the county specific effect in the third level equation a simple i.i.d random effect without further covariates is assumed, i.e. $\boldsymbol{\varepsilon}_{32} \sim N(0, \sigma_{32}^2)$. One of the advantages of our approach is that we are able to model spatial heterogeneity at different resolutions (here district and county level). This allows a very detailed modeling of spatial heterogeneity and provides further insight into the problem.

Results for the effects of "no-claims bonus" and "car classification" are given in figure 1 showing a monotonically decreasing effect for *nclaim*. Since higher scores for car classification roughly correspond to "bigger cars" the random effect for g is more or less increasing with scores (with notable exceptions for car groups 32 and 34).

A visualization of the spatial effect $\boldsymbol{\beta}_3$ can be found in figure 2. It is composed of the spatially smooth district effect $\mathbf{f}_{31}(dist)$, the district i.i.d. random effect $\boldsymbol{\varepsilon}_3$, the county random effect $\mathbf{f}_{32}(county)$ and the nonlinear effect $\mathbf{f}_{33}(dens)$ of population density. The i.i.d district random effect is very small while the other effects are considerably stronger and roughly of equal size (all effects not shown to save space). Inspecting the total spatial effect in figure 2 reveals a clear north south pattern with lower damage probabilities in the north, in particular the less densely populated north eastern part of Germany, and higher probabilities in the south. We nicely see the effect of the "population density" *dens* as the most densely populated urban areas of Germany are mostly colored in dark grey or black indication higher damage probabilities as in the rural areas. The effect $\mathbf{f}_{33}(dens)$ itself is almost linearly increasing (not shown).

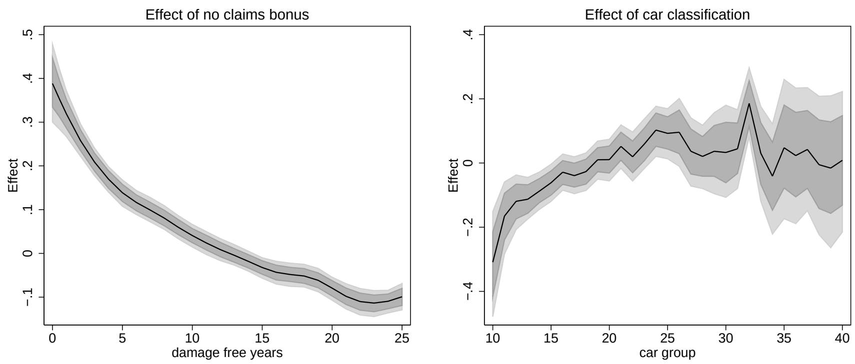

Fig. 1. Car insurance: Effect of "no claims bonus" and "car classification" random effect including 80% and 95% pointwise credible intervals.

4.2 Health insurance data

In our second example we exemplify modeling of space-time interactions using data from a German private health insurance company. In a consulting case the main interest was on analyzing the dependence of treatment costs on covariates with a special emphasis on modeling the spatio-temporal development. We distinguish several types of treatment costs. In this demonstrating example, we present results for "treatment with operation" in hospital. We assumed a two level Gaussian model for the log treatment costs $C_{it} \sim N(\eta_{it}, \sigma^2)$ for policyholder i at time t and with predictor

$$\eta_{it} = \cdots + f_1(A_{it}) + f_2(t, county_{it}) + f_3(D_{it}),$$

where f_1 is a nonlinear effect of the policyholders age modeled via P-splines, f_2 represents county specific nonlinear time trends modeled again using P-splines, and f_3 is a district specific spatial random effect modeled in a second

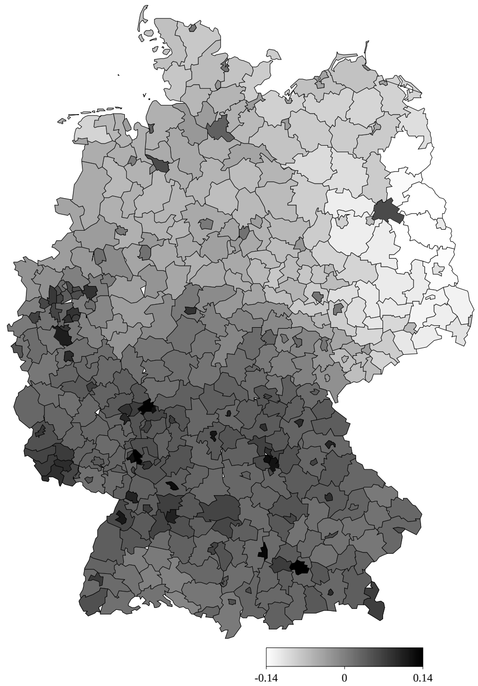

Fig. 2. Car insurance: Visualization of the spatial random effect.

level equation (not shown here). The time-space interaction f_2 is regularized by assuming a common variance parameter for the otherwise unrestricted curves.

Figure 3 displays the county specific time trends showing considerable variation from county to county. For comparison figure 4 shows for the counties in the last row of figure 3 the time trend if the different curves are *not* regularized through common variance parameters, i.e. *different* variance parameters are assumed for each curve. Obviously the curves are much more wiggled and the credible intervals show some instability.

References

BELITZ, C. and LANG, S. (2008): Simultaneous selection of variables and smoothing parameters in structured additive regression models. *Computational Statistics and Data Analysis 53, 61-81.*

BREZGER, A. and LANG, S. (2006): Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis 50, 967-991.*

FAHRMEIR, L., KNEIB, T. and LANG, S. (2004): Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica 14, 731-761.*

FRÜHWIRTH-SCHNATTER, S., FRÜHWIRTH, R., HELD, L. and RUE, H. (2008): Improved Auxiliary Mixture Sampling for Hierarchical Models of Non-Gaussian Data. *IFAS, University of Linz.*

LANG, S., UMLAUF, N., KNEIB, T. and WECHSELBERGER, P. (2010): Multilevel Generalized Structured Additive Regression.

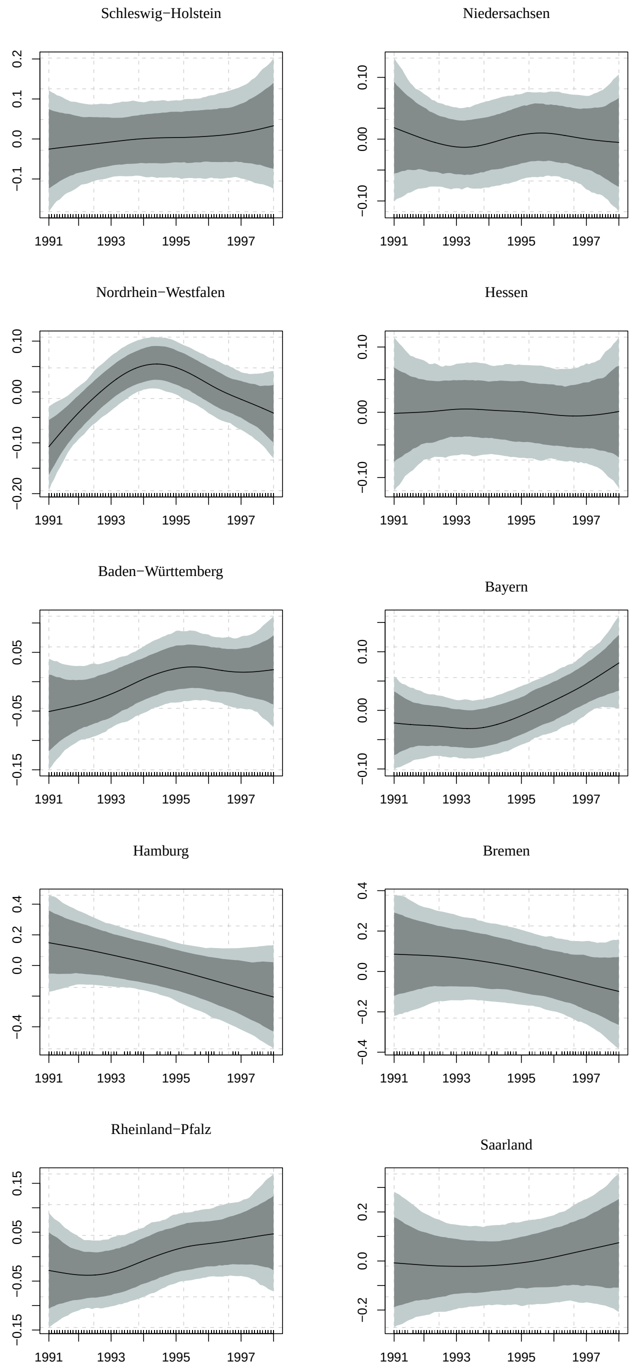

Fig. 3. Health insurance: Visualization of the space-time interaction $f_2(t, county_{it})$.

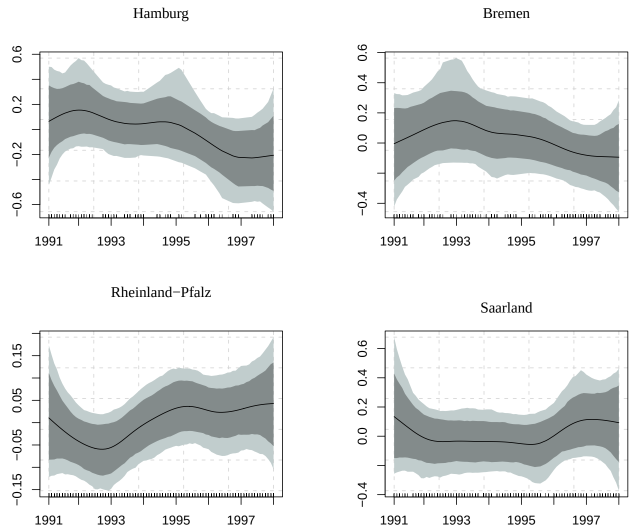

Fig. 4. Health insurance: Visualization of the space-time interaction for Hamburg, Bremen, Rheinland-Pfalz and Saarland if different variance parameters are used for the county specify time trends.

University of Innsbruck – Working Papers in Economics and Statistics
Recent papers

2010-01 **Stefan Lang and Nikolaus Umlauf:** Applications of Multilevel Structured Additive Regression Models to Insurance Data

---

2009-29 **Loukas Balafoutas:** How much income redistribution? An explanation based on vote-buying and corruption

2009-28 **Rudolf Kerschbamer, Matthias Sutter and Uwe Dulleck:** The Impact of Distributional Preferences on (Experimental) Markets for Expert Services

2009-27 **Adrian Beck, Rudolf Kerschbamer, Jianying Qiu and Matthias Sutter:** Car Mechanics in the Lab - Investigating the Behavior of Real Experts on Experimental Markets for Credence Goods

2009-26 **Michael Kirchler, Jürgen Huber and Thomas Stöckl:** Bubble or no Bubble - The Impact of Market Model on the Formation of Price Bubbles in Experimental Asset Markets

2009-25 **Rupert Sausgruber and Jean-Robert Tyran:** Tax Salience, Voting, and Deliberation

2009-24 **Gerald J. Pruckner and Rupert Sausgruber:** Honesty on the Streets - A Natural Field Experiment on Newspaper Purchasing

2009-23 **Gerlinde Fellner, Rupert Sausgruber and Christian Traxler:** Testing Enforcement Strategies in the Field: Legal Threat, Moral Appeal and Social Information

2009-22 **Ralph-C. Bayer, Elke Renner and Rupert Sausgruber:** Confusion and Reinforcement Learning in Experimental Public Goods Games

2009-21 **Sven P. Jost:** Transfer Pricing Risk Awareness of Multinational Corporations - Evidence from a Global Survey

2009-20 **Andrea M. Leiter and Engelbert Theurl:** The Convergence of Health Care Financing Structures: Empirical Evidence from OECD-Countries

2009-19 **Francesco Feri and Miguel A. Meléndez-Jiménez:** Coordination in Evolving Networks with Endogenous Decay

2009-18 **Harald Oberhofer:** Firm growth, European industry dynamics and domestic business cycles

2009-17 **Jesus Crespo Cuaresma and Martin Feldkircher:** Spatial Filtering, Model Uncertainty and the Speed of Income Convergence in Europe

2009-16 **Paul A. Raschky and Manijeh Schwindt:** On the Channel and Type of International Disaster Aid

2009-15 **Jianying Qiu:** Loss aversion and mental accounting: The favorite-longshot bias in parimutuel betting

2009-14 **Siegfried Berninghaus, Werner Güth, M. Vittoria Levati and Jianying Qiu:** Satisficing in sales competition: experimental evidence

2009-13 **Tobias Bruenner, Rene Levinský and Jianying Qiu:** Skewness preferences and asset selection: An experimental study

2009-12 **Jianying Qiu and Prashanth Mahagaonkar:** Testing the Modigliani-Miller theorem directly in the lab: a general equilibrium approach

2009-11 **Jianying Qiu and Eva-Maria Steiger:** Understanding Risk Attitudes in two Dimensions: An Experimental Analysis

2009-10 **Erwann Michel-Kerjan, Paul A. Raschky and Howard C. Kunreuther:** Corporate Demand for Insurance: An Empirical Analysis of the U.S. Market for Catastrophe and Non-Catastrophe Risks

2009-09 **Fredrik Carlsson, Peter Martinsson, Ping Qin and Matthias Sutter:** Household decision making and the influence of spouses' income, education, and communist party membership: A field experiment in rural China

2009-08 **Matthias Sutter, Peter Lindner and Daniela Platsch:** Social norms, third-party observation and third-party reward

2009-07 **Michael Pfaffermayr:** Spatial Convergence of Regions Revisited: A Spatial Maximum Likelihood Systems Approach

2009-06 **Reimund Schwarze and Gert G. Wagner:** Natural Hazards Insurance in Europe – Tailored Responses to Climate Change Needed

2009-05 **Robert Jiro Netzer and Matthias Sutter:** Intercultural trust. An experiment in Austria and Japan

2009-04 **Andrea M. Leiter, Arno Parolini and Hannes Winner:** Environmental Regulation and Investment: Evidence from European Industries

2009-03 **Uwe Dulleck, Rudolf Kerschbamer and Matthias Sutter:** The Economics of Credence Goods: On the Role of Liability, Verifiability, Reputation and Competition. *Revised version forthcoming in American Economic Review.*

2009-02 **Harald Oberhofer and Michael Pfaffermayr:** Fractional Response Models - A Replication Exercise of Papke and Wooldridge (1996)

2009-01 **Loukas Balafoutas:** How do third parties matter? Theory and evidence in a dynamic psychological game.

2008-27 **Matthias Sutter, Ronald Bosman, Martin Kocher and Frans van Winden:** Gender pairing and bargaining – Beware the same sex! *Revised version published in Experimental Economics, Vol. 12 (2009): 318-331.*

2008-26 **Jesus Crespo Cuaresma, Gernot Doppelhofer and Martin Feldkircher:** The Determinants of Economic Growth in European Regions.

2008-25 **Maria Fernanda Rivas and Matthias Sutter:** The dos and don'ts of leadership in sequential public goods experiments.

2008-24 **Jesus Crespo Cuaresma, Harald Oberhofer and Paul Raschky:** Oil and the duration of dictatorships.

2008-23 **Matthias Sutter:** Individual behavior and group membership: Comment. *Revised Version published in American Economic Review, Vol.99 (2009): 2247-2257.*

2008-22 **Francesco Feri, Bernd Irlenbusch and Matthias Sutter:** Efficiency Gains from Team-Based Coordination – Large-Scale Experimental Evidence. *Revised and extended version forthcoming in American Economic Review.*

2008-21 **Francesco Feri, Miguel A. Meléndez-Jiménez, Giovanni Ponti and Fernando Vega Redondo:** Error Cascades in Observational Learning: An Experiment on the Chinos Game.

2008-20 **Matthias Sutter, Jürgen Huber and Michael Kirchler:** Bubbles and information: An experiment.

2008-19 **Michael Kirchler:** Curse of Mediocrity - On the Value of Asymmetric Fundamental Information in Asset Markets.

2008-18 **Jürgen Huber and Michael Kirchler:** Corporate Campaign Contributions as a Predictor for Abnormal Stock Returns after Presidential Elections.

2008-17 **Wolfgang Brunauer, Stefan Lang, Peter Wechselberger and Sven Bienert:** Additive Hedonic Regression Models with Spatial Scaling Factors: An Application for Rents in Vienna.

2008-16 **Harald Oberhofer, Tassilo Philippovich:** Distance Matters! Evidence from Professional Team Sports. *Extended and revised version forthcoming in Journal of Economic Psychology.*

2008-15 **Maria Fernanda Rivas and Matthias Sutter:** Wage dispersion and workers' effort.

2008-14 **Stefan Borsky and Paul A. Raschky:** Estimating the Option Value of Exercising Risk-taking Behavior with the Hedonic Market Approach. *Revised version forthcoming in Kyklos.*

2008-13 **Sergio Currarini and Francesco Feri:** Information Sharing Networks in Oligopoly.

2008-12 **Andrea M. Leiter:** Age effects in monetary valuation of mortality risks - The relevance of individual risk exposure.

2008-11 **Andrea M. Leiter and Gerald J. Pruckner:** Dying in an Avalanche: Current Risks and their Valuation.

2008-10 **Harald Oberhofer and Michael Pfaffermayr:** Firm Growth in Multinational Corporate Groups.

2008-09 **Michael Pfaffermayr, Matthias Stöckl and Hannes Winner:** Capital Structure, Corporate Taxation and Firm Age.

2008-08 **Jesus Crespo Cuaresma and Andreas Breitenfellner:** Crude Oil Prices and the Euro-Dollar Exchange Rate: A Forecasting Exercise.

2008-07 **Matthias Sutter, Stefan Haigner and Martin Kocher:** Choosing the carrot or the stick? – Endogenous institutional choice in social dilemma situations. Revised version forthcoming in *Review of Economic Studies*.

2008-06 **Paul A. Raschky and Manijeh Schwindt:** Aid, Catastrophes and the Samaritan's Dilemma.

2008-05 **Marcela Ibanez, Simon Czermak and Matthias Sutter:** Searching for a better deal – On the influence of group decision making, time pressure and gender in a search experiment. *Revised version published in Journal of Economic Psychology, Vol. 30 (2009): 1-10.*

2008-04 **Martin G. Kocher, Ganna Pogrebna and Matthias Sutter:** The Determinants of Managerial Decisions Under Risk.

2008-03 **Jesus Crespo Cuaresma and Tomas Slacik:** On the determinants of currency crises: The role of model uncertainty. *Revised version accepted for publication in Journal of Macroeconomics.*

2008-02 **Francesco Feri:** Information, Social Mobility and the Demand for Redistribution.

2008-01 **Gerlinde Fellner and Matthias Sutter:** Causes, consequences, and cures of myopic loss aversion - An experimental investigation. *Revised version published in The Economic Journal, Vol. 119 (2009), 900-916.*

2007-31 **Andreas Exenberger and Simon Hartmann:** The Dark Side of Globalization. The Vicious Cycle of Exploitation from World Market Integration: Lesson from the Congo.

2007-30 **Andrea M. Leiter and Gerald J. Pruckner:** Proportionality of willingness to pay to small changes in risk - The impact of attitudinal factors in scope tests. *Revised version forthcoming in Environmental and Resource Economics.*

2007-29 **Paul Raschky and Hannelore Weck-Hannemann:** Who is going to save us now? Bureaucrats, Politicians and Risky Tasks.

2007-28 **Harald Oberhofer and Michael Pfaffermayr:** FDI versus Exports. Substitutes or Complements? A Three Nation Model and Empirical Evidence.

2007-27 **Peter Wechselberger, Stefan Lang and Winfried J. Steiner:** Additive models with random scaling factors: applications to modeling price response functions.

2007-26 **Matthias Sutter:** Deception through telling the truth?! Experimental evidence from individuals and teams*. Revised version published in The Economic Journal, Vol. 119 (2009), 47-60.*

2007-25 **Andrea M. Leiter, Harald Oberhofer and Paul A. Raschky:** Productive disasters? Evidence from European firm level data. *Revised version forthcoming in Environmental and Resource Economics.*

2007-24 **Jesus Crespo Cuaresma:** Forecasting euro exchange rates: How much does model averaging help?

2007-23 **Matthias Sutter, Martin Kocher and Sabine Strauß:** Individuals and teams in UMTS-license auctions. *Revised version with new title "Individuals and teams in auctions" published in Oxford Economic Papers, Vol. 61 (2009): 380-394***).**

2007-22 **Jesus Crespo Cuaresma, Adusei Jumah and Sohbet Karbuz:** Modelling and Forecasting Oil Prices: The Role of Asymmetric Cycles. *Revised version accepted for publication in The Energy Journal.*

2007-21 **Uwe Dulleck and Rudolf Kerschbamer:** Experts vs. discounters: Consumer free riding and experts withholding advice in markets for credence goods. *Revised version published in* *<u>International Journal of Industrial Organization</u>*, *Vol. 27, Issue 1 (2009): 15-23.*

2007-20 **Christiane Schwieren and Matthias Sutter:** Trust in cooperation or ability? An experimental study on gender differences*. Revised version published in* *<u>Economics Letters</u>, Vol. 99 (2008): 494-497.*

2007-19 **Matthias Sutter and Christina Strassmair:** Communication, cooperation and collusion in team tournaments – An experimental study. *Revised version published in: <u>Games and Economic Behavior</u>, Vol.66 (2009), 506-525.*

2007-18 **Michael Hanke, Jürgen Huber, Michael Kirchler and Matthias Sutter:** The economic consequences of a Tobin-tax – An experimental analysis. *Revised version forthcoming in <u>Journal of Economic Behavior and Organization</u>.*

2007-17 **Michael Pfaffermayr:** Conditional beta- and sigma-convergence in space: A maximum likelihood approach. *Revised version forthcoming in <u>Regional Science and Urban Economics</u>.*

2007-16 **Anita Gantner:** Bargaining, search, and outside options. *Published in: <u>Games and Economic Behavior</u>, Vol. 62 (2008), pp. 417-435.*

2007-15 **Sergio Currarini and Francesco Feri:** Bilateral information sharing in oligopoly.

2007-14 **Francesco Feri:** Network formation with endogenous decay.

2007-13 **James B. Davies, Martin Kocher and Matthias Sutter:** Economics research in Canada: A long-run assessment of journal publications. *Revised version published in: <u>Canadian Journal of Economics</u>, Vol. 41 (2008), 22-45.*

2007-12 **Wolfgang Luhan, Martin Kocher and Matthias Sutter:** Group polarization in the team dictator game reconsidered. *Revised version published in: <u>Experimental Economics</u>, Vol. 12 (2009), 26-41.*

2007-11 **Onno Hoffmeister and Reimund Schwarze:** The winding road to industrial safety. Evidence on the effects of environmental liability on accident prevention in Germany.

2007-10 **Jesus Crespo Cuaresma and Tomas Slacik:** An "almost-too-late" warning mechanism for currency crises. *(Revised version accepted for publication in <u>Economics of Transition</u>*)

2007-09 **Jesus Crespo Cuaresma, Neil Foster and Johann Scharler:** Barriers to technology adoption, international R&D spillovers and growth.

2007-08 **Andreas Brezger and Stefan Lang:** Simultaneous probability statements for Bayesian P-splines.

2007-07 **Georg Meran and Reimund Schwarze:** Can minimum prices assure the quality of professional services? (*Accepted for publication in <u>European Journal of Law and Economics</u>*)

2007-06 **Michal Brzoza-Brzezina and Jesus Crespo Cuaresma:** Mr. Wicksell and the global economy: What drives real interest rates?.

2007-05 **Paul Raschky:** Estimating the effects of risk transfer mechanisms against floods in Europe and U.S.A.: A dynamic panel approach.

2007-04 **Paul Raschky and Hannelore Weck-Hannemann:** Charity hazard - A real hazard to natural disaster insurance. *Revised version forthcoming in: <u>Environmental Hazards</u>.*

2007-03 **Paul Raschky:** The overprotective parent - Bureaucratic agencies and natural hazard management.

2007-02 **Martin Kocher, Todd Cherry, Stephan Kroll, Robert J. Netzer and Matthias Sutter:** Conditional cooperation on three continents. *Revised version published in: <u>Economics Letters</u>, Vol. 101 (2008): 175-178.*

2007-01 **Martin Kocher, Matthias Sutter and Florian Wakolbinger:** The impact of naïve advice and observational learning in beauty-contest games.

University of Innsbruck

Working Papers in Economics and Statistics

2010-01

Stefan Lang and Nikolaus Umlauf

Applications of Multilevel Structured Additive Regression Models to Insurance Data

Abstract

Models with structured additive predictor provide a very broad and rich framework for complex regression modeling. They can deal simultaneously with nonlinear covariate effects and time trends, unit- or cluster specific heterogeneity, spatial heterogeneity and complex interactions between covariates of different type. In this paper, we discuss a hierarchical version of regression models with structured additive predictor and its applications to insurance data. That is, the regression coefficients of a particular nonlinear term may obey another regression model with structured additive predictor. The proposed model may be regarded as an extended version of a multilevel model with nonlinear covariate terms in every level of the hierarchy. We describe several highly efficient MCMC sampling schemes that allow to estimate complex models with several hierarchy levels and a large number of observations typically within a couple of minutes. We demonstrate the usefulness of the approach with applications to insurance data.

ISSN 1993-4378 (Print)
ISSN 1993-6885 (Online)