I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

D I S C U S S I O N
P A P E R

0202

RATEMAKING BY GEOGRAPHICAL AREA:
A CASE STUDY USING THE BOSKOV AND VERRALL MODEL

N. BROUHNS, M. DENUIT, B. MASUY & R. VERRALL

http://www.stat.ucl.ac.be

RATEMAKING BY GEOGRAPHICAL AREA: A CASE STUDY USING THE BOSKOV AND VERRALL MODEL

Natacha Brouhns
Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve, Belgium
brouhns@stat.ucl.ac.be

Michel Denuit
Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve, Belgium
denuit@stat.ucl.ac.be

Bernard Masuy
Observatoire Social et Politique
Université Catholique de Louvain
Place Montesquieu, 1, B1
B-1348 Louvain-la-Neuve, Belgium
masuy@rspo.ucl.ac.be

Richard Verrall
Dpt. of Actuarial Science and Statistics
City University
Northampton Square
London EC1V 0HB, UK
r.j.verrall@city.ac.uk

January 24, 2002

Abstract

This paper uses a method proposed by Boskov & Verrall (1994) for premium rating by postcode area. The aim is to analyze geographical variation in claim frequencies in order to estimate the local risk in each geographical area and to build homogeneous rating regions. The method accounts for spatial correlation in an hierarchical Bayesian framework and uses computer-intensive MCMC methods for statistical inference.

Key words and phrases: insurance rating, spatial statistics, Poisson regression, postcodes, Gibbs sampler, hierarchical Bayes models.

1 Introduction and Motivation

1.1 The problem under study

This paper is devoted to spatial models for insurance rating. The aim of this article is to provide actuaries with a method for analyzing risk variation by geographic area. In particular, we purpose to predict the underlying risk in a geographical region, using claims data which are near or relatively near to a region of interest. The geographical location is contained within the postcode for each policy.

It is common in insurance of domestic property lines, such as householders' fire insurance, for instance, to let the risk premium per unit exposure vary with geographic area when all other risk factors are held constant. In automobile insurance, most companies have adopted a risk classification according to the geographical zone where the policyholder lives (urban / non urban for instance, or a more accurate splitting of the country according to Zip codes). The spatial variation may be related to geographic factors (e.g. traffic density or proximity to arterial roads in automobile insurance) or to socio-demographic factors (perhaps affecting theft rates in house insurance). In such cases it will be desirable to estimate the spatial variation in risk premium and to price accordingly.

The method described in this paper is based on statistical models for spatial data. Although these methods have been developed for other applications (image restoration using data from satellites), the techniques can also be used for risk assessment in an insurance context.

Spatial postcode methods for insurance rating attempt to improve estimates for the pure premiums of a potential policy given its geographical location (contained in the postcode). The aim is to extract information which is in addition to that contained in standard factors (like age or gender for instance). Often, claim characteristics tend to be similar in neighboring postcode areas (after other factors have been accounted for). The idea of postcode rating models is to exploit this spatial smoothness and allow information transfer to and from neighboring regions.

1.2 A simple example

We suppose that the region to be mapped is divided into n mutually exclusive districts and that the frequency risk is under investigation. Under the conventional approach to mapping risk, it is implicit that the risk in the ith district, θ_i, is an unknown parameter to be estimated. If Y_i and e_i, are, respectively, the observed and expected number of claims in the ith district, the standard maximum likelihood estimate of the θ_i under Poisson assumption for Y_i is $\hat{\theta}_i = \frac{Y_i}{e_i}$. However, taken together, the $\hat{\theta}_i$'s are not necessarily the best estimates of the θ_i's. These estimates suffer essentially from two main problems. First and foremost is the "small number problem", that is, the dependence of the statistical variation of the $\hat{\theta}_i$ on the risk exposure. The raw estimates $\widehat{\theta_i}$ are thus usually unreliable because some local areas contain only a few people at risk. This makes it a necessity to "borrow strength" or utilize information form the neighboring areas in order to produce smoothed estimates for the individual local areas. The second problem is that the $\hat{\theta}_i$'s totally ignore the spatial structure of the data. Both problems can be addressed through stochastic modelling of the

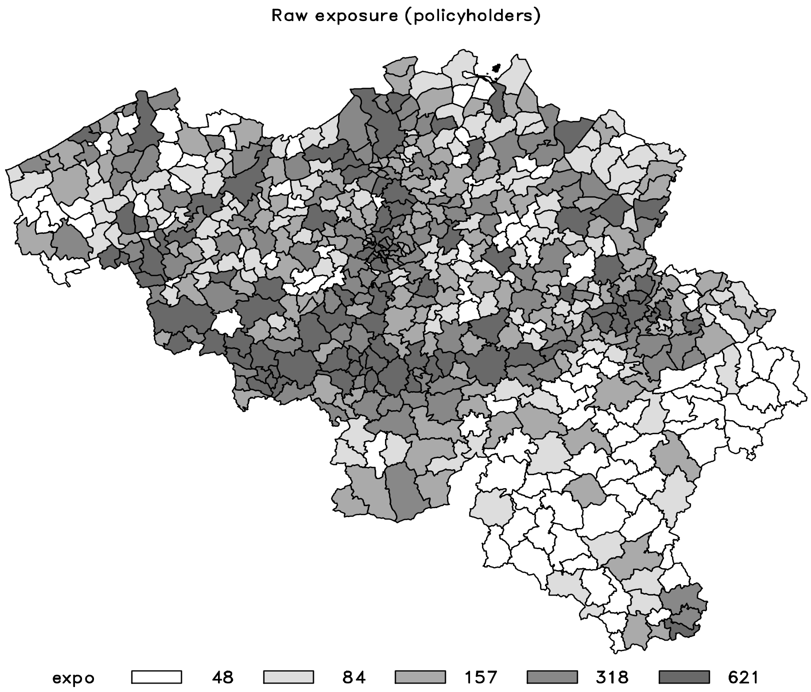

Figure 1.1.1: *Map of Belgium with exposures-to-risk (in persons).*

relative risk parameters.

To make these ideas more explicit, let us examine an example. The data used to illustrate the present article relate to the portfolio of a major Belgian insurance company. Specifically, we had at our disposal the number of claims Y_i reported by each of the 165,363 policyholders during the year 1997, together with explanatory variables (see Section 4 for a complete description). The data represented on the figures have been grouped into classes, which are divided by the 10, 25, 50, 75 and 90-th percentiles. Figure 1.1.1 displays the raw exposures e_i for each area (the number of policy-years, in our case) whilst Figure 1.1.2 shows crude rates $\widehat{\theta}_i$ (number of claims divided by exposures). However, the latter map is at best difficult to interpret and can even be seriously misleading because the $\widehat{\theta}_i$'s tend to be far more extreme in regions with smaller risk exposures (see Figures 1.1.1-1.1.2; the high rates in the North-West of Belgium (West Flanders) or in the South of the country correspond to districts for which we have less policyholders). Hence regions with the least reliable data will typically draw the main visual attention. This is one reason why it is difficult in practice to attempt any smoothing or risk assessment "by eye".

An unknown part of the variation of the $\widehat{\theta}_i$'s may be caused by geographically varying unobserved factors. The so-called disease mapping methods have been developed to give more reliable estimates of the geographical variation of claim rates. The general goal is to identify extra sample variation due to unobserved heterogeneity by filtering the Poisson sample variation.

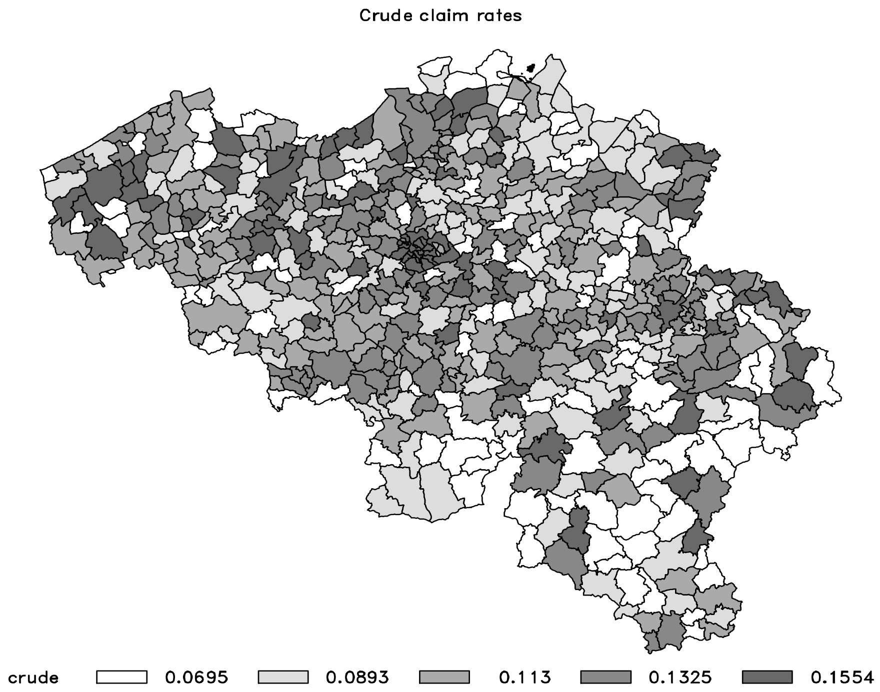

Figure 1.1.2: *Map of Belgium with crude rates.*

1.3 Hierarchical and empirical Bayes approaches

The simple example discussed above illustrates the unreliability of the raw estimates of the θ_i's. A convenient solution for estimating the θ_i's is to resort to Bayes methods, which connect the local areas, thereby enabling one to "borrow strength". The Bayesian statistical approach treats all unknown parameters appearing in a statistical model as random variables and derives their distribution conditional upon the known information. Hierarchical and empirical Bayes (HB and EB) procedures are particularly well suited to connect the local areas systematically through a model. The similarity between the two Bayesian procedures is that they both recognize the uncertainty due to not knowing the prior parameters (often called hyperparameters). But whereas the EB procedure estimates the hyperparameters from the marginal distributions of the observations, the HB method models the same by assigning second stage (and often diffuse) priors.

The EB approach of Clayton & Kaldor (1987) (see also Breslow & Clayton, 1993, and Leroux, 2000) shrinks the $\widehat{\theta}_i$'s toward a local or global mean, where the amount of shrinkage is determined by the reliability of the data of that particuliar region. The two smoothing options, local or global, seem to be appropriate if unobserved risk factors do or do not have a spatial structure, respectively. This approach was further generalized by Besag, York & Mollié (1991) who allowed for both spatially structured and unstructured heterogeneity in one model. In this paper, we resort to HB Poisson models for the analysis of spatial data, following Ghosh, Natarajan, Waller & Kim (1999).

1.4 A brief review of the literature

Whereas epidemiologists and environmetricians have been interested in spatial models for a long time, the actuarial literature is very poor in respect of ratemaking methods incorporating geographic components. TAYLOR (1989) used two-dimensional splines on a plane linked to the map of the region by a transformation chosen to match the features of the specific region. He applied this method to a data set from Sydney, Australia. BOSKOV & VERRALL (1994) highlighted some deficiencies in Taylor's model, and provided an alternative treatment which made use of the Gibbs sampler to implement a Bayesian revision of the observation in each area. The main advantage of the Bayesian framework is that it recognizes the magnitudes of sampling error and incorporates the concept of smoothing over neighboring areas. TAYLOR (1996) adopts a similar point of view and applied Whittaker graduation (a widely accepted actuarial technique which has also been shown to have a Bayesian interpretation; see TAYLOR (1989) and VERRALL (1993)). DIXON, KELSEY & VERRALL (2000) proposed an extension of BOSKOV & VERRALL (1994) model including weighting factors accounting for distances between regions.

1.5 Outline of the study

The paper is set out as follows. Section 2 contains the different elements belonging to the Boskov and Verrall model. In particular, it gives the definitions of prior and posterior distributions, a short introduction to Gibbs sampling and the application of this technique to geographical ratemaking. Section 3 aims to briefly present other approaches to the same problem, while section 4 describes a numerical illustration.

2 The Boskov and Verrall model

2.1 HB Poisson model for spatial data

We assume that we have a region divided into geographical areas numbered from 1 to n. Let x_i be the true risk in area i and $\boldsymbol{x}$ be the vector of risks over the whole region. The following HB spatial Poisson model is broad enough to cover most situations encountered by actuaries where a spatial structure needs to be incorporated. Let y_i denote the observed outcome relating to area i, and $\boldsymbol{y}$ be the corresponding vector. Conditional on $\boldsymbol{x}$, we assume that the y_i's are mutually independent with probability density function

$$f(y_i|x_i) = \exp(-x_i)\frac{x_i^{y_i}}{y_i!}.$$

Let us make explicit the components of x_i. The risk level for area i is assumed to be comprised of four terms

$$x_i = r_i \exp(\eta_i + u_i + v_i)$$

where

r_i is a known constant measuring the risk exposure in area i;

η_i is a linear predictor based on known factors whose coefficients are estimated using a standard Poisson regression model with r_i as offset (at this stage, the η_i's are treated as known constants)

u_i represents a component with significant spatial structure (it is the Markov random field term with spatially structured heterogeneity)

v_i represents unexplained variation which does not have a spatial structure.

Large values of v_i (relative to u_i) suggest exceptions to the smooth spatial pattern, or boundaries between two separate regions of stronger spatial correlation. It is the component u_i that is of interest in our analysis of the spatial structure of the data.

Regarding the incorporation of random effects, our modelling is similar to that of Besag et al. (1991), where each area effect is expressed as the sum of two random components. One set of error components are independent and identically distributed reflecting an unstructured pattern of heterogeneity, while the other set of error components exhibit spatial correlation among neighboring areas but uncorrelation among the other areas. This construction is usually referred to as "structured heterogeneity". The target is to find the posterior distribution of $\boldsymbol{u}$ given $\boldsymbol{y}$ and in particular the posterior means, medians, variances and covariances of this distribution.

Introducing $e_i = r_i \exp(\eta_i)$ the expected number of claims in region i ignoring the spatial effect and $\theta_i = \exp(u_i + v_i)$ the unknown relative risk in area i we have that $x_i = e_i\theta_i$.

2.2 Interpretation of the spatial correlation

One interpretation of spatial autocorrelation among claim rates is residual random variation due to unmeasured covariates. The partitioning of a model into deterministic (covariate effects) and stochastic (correlation) components is not unique, as shown in Cressie (1991, pp. 113-114). One goal of the statistical modeller is to accurately measure the effects of the covariates on the outcome while adequately accounting for residual correlation. If, as in our case, prediction is of primary importance, the goal is to portray the best estimates available from the data. For these applications, the correct partitioning into components is secondary to correct modelling of each component.

The components u_i and v_i of the x_i's can be interpreted as surrogates for unknown or unobserved covariates. The u_i's represent variables that, if observed, would display substantial spatial structure in that values for a pair of contiguous zones would be generally much more alike than for two arbitrary zones. The v_i's represent unstructured variables.

In practice, it will often be the case that either $\boldsymbol{u}$ or $\boldsymbol{v}$ dominates the other but which one will not usually be known in advance. If $\boldsymbol{u}$, then the estimated risks $\boldsymbol{x}$ will display spatial structure. If $\boldsymbol{v}$, then the effect will be to shrink the estimated risks towards the overall mean.

2.3 Definition of neighborhoods

A key step in model specification is the definition of neighbors, i.e. those areas whose claim rates are correlated with that of a given area. A traditional definition of neighbors includes all areas contiguous to a given area. Specifically, we define for each i the set δ_i of areas in the neighborhood of the ith area; in our setting, δ_i can be interpreted as postcode areas which are adjacent, or close, to the ith area. Typically, δ_i is the set of postcodes which border on region i in the spirit of Markov random fields (Besag, 1974).

2.4 Prior distribution for the random effects

As usual, the first stage of a Bayesian analysis is to specify a prior probability density for $\boldsymbol{x}$. The latter should support the local regularities that are believed to exist. In particular, we usually anticipate that nearby regions are likely to be more similar than those further apart. In assessing the local behavior of any prior distribution, the most useful characteristic is the conditional density of x_i given all other region values x_j, $j \neq i$. Usually, we want this to depend only on the values at a few areas in the immediate vicinity of i, that is in δ_i.

Let us now formulate our prior beliefs concerning $\boldsymbol{x}$ as a joint distribution for $\boldsymbol{u}$ and $\boldsymbol{v}$. We assume that the risk at the ith region depends only on regions which are in the neighborhood δ_i of the ith region. Following Besag et al. (1991, 1995), we assume that the prior conditional density of the spatial component u_i can be factorized into components representing the dependencies on each of the neighboring regions. Hence, we opt for a pairwise difference prior of the form

$$p(u_i|u_j, j \neq i, \tau) \propto \frac{1}{\tau^{1/2}} \exp\left(-\frac{1}{2\tau}\sum_{j\in\delta_i}(u_i - u_j)^2\right).$$

It reflects the fact that the spatial dependence will reduce as the distance between the regions increase. This choice naturally imposes smoothing on the u_k's (since large differences between neighboring regions are penalized).

The joint distribution for $\boldsymbol{u}$ is

$$p(\boldsymbol{u}|\tau) \propto \frac{1}{\tau^{n/2}} \exp\left(-\frac{1}{2\tau}\sum_{i\sim j}(u_i - u_j)^2\right) \tag{2.1}$$

where $i \sim j$ denotes i and j are contiguous. This is a Gaussian intrinsic autoregression and has conditional moments

$$\mathbb{E}[u_i|u_j, j \neq i] = \frac{1}{\#\delta_i}\sum_{j\in\delta_i} u_j = \overline{u}_i \text{ and } \mathbb{V}\text{ar}[u_i|u_j, j \neq i] = \frac{\tau}{\#\delta_i}$$

where $\#\delta_i$ is the cardinality of δ_i. Note that $\tau \to 0$ implies constant u_i's whereas τ large implies correspondingly large but spatially structured variations.

Compared to the simplest models (building exchangeability among all the local areas, and shrinking the individual local area effects towards a global value), the present model incorporates the geographical structure of the map. More specifically, estimates of the u_i's are strongly influenced by their neighbors, and only indirectly influenced by estimates of all other areas of the map. As a result, the individual estimates shrink more towards a local than towards a global mean value. Inclusion of the covariates also eliminates exchangeability among the neighboring areas as such.

We also assume that u_i and v_i are independent, and that the v_i's are mutually independent and conform to a normal prior distribution, i.e.

$$p(v_i|\lambda) = \lambda^{-1/2} \exp\left(-\frac{1}{2\lambda} v_i^2\right).$$

Thus, $\boldsymbol{v}$ is a realization of a Gaussian white noise with unknown variance λ. Note that $\lambda \to 0$ implies $\boldsymbol{v} = \mathbf{0}$ whereas λ large implies substantial but unstructured extra Poisson variability.

The two scale parameters τ and λ which determine the variances of the u_i's and of the v_i's must also be given a priori distribution. A suitable choice for this prior, which is close to the uninformative distribution but which avoids technical difficulties is

$$\mathrm{prior}(\lambda, \tau) \propto \exp\left(-\frac{\epsilon}{2\tau} - \frac{\epsilon}{2\lambda}\right)$$

for some small positive constant ϵ (0.01, say); see Besag et al. (1991) for more details. It is worth mentioning that Gamma priors for λ and τ were proposed by Ghosh et al. (1999).

Note that in the prior distribution of $\boldsymbol{x}$, induced by those of $\boldsymbol{u}$ and $\boldsymbol{v}$, the conditional density of x_i depends on all other x_j's, and not merely on those in contiguous areas.

2.5 Posterior distribution

If the x_i's were the only unknowns then inferences about $\boldsymbol{x}$ should be based on the posterior density of $\boldsymbol{x}$ given $\boldsymbol{y}$, that is

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{x}) p(\boldsymbol{x}).$$

The most obvious Bayesian point estimate of $\boldsymbol{x}$ is that which maximizes $p(\boldsymbol{x}|\boldsymbol{y})$; this the maximum a posteriori estimate (MAPE) of $\boldsymbol{x}$. This is particularly attractive when $p(\boldsymbol{x}|\boldsymbol{y})$ has a unique maximum but loses its appeal and is extremely difficult to locate if there are many local maxima (as is often the case in real situations). Furthermore the determination of the MAPE of $\boldsymbol{x}$ provides no assessment of precision.

For this reason, we prefer to make inference empirically by collecting many realizations from the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$. Next, we discuss implementation of the Bayes procedures via Markov Chain Monte Carlo (MCMC). In particular, we use the Gibbs sampler which requires generating samples from the full conditionals, in order to generate samples from the marginal posterior $p(\boldsymbol{x}|\boldsymbol{y})$.

Now, having split each x_i into several parts, we rewrite $p(\boldsymbol{x}|\boldsymbol{y})$ taking the structure of the x_i's into account. This yields

$$\begin{aligned} p(\boldsymbol{u}, \boldsymbol{v}, \tau, \lambda|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{v}, \tau, \lambda) p(\boldsymbol{u}, \boldsymbol{v}, \tau, \lambda) \\ &= \left\{\prod_{i=1}^{n} f(y_i|x_i)\right\} p(\boldsymbol{u}|\tau) p(\boldsymbol{v}|\lambda) \mathrm{prior}(\tau, \lambda) \\ &\propto \left\{\prod_{i=1}^{n} \exp(-x_i) \frac{x_i^{y_i}}{y_i!}\right\} \tau^{-n/2} \exp\left(-\frac{1}{2\tau} \sum_{i \sim j} (u_i - u_j)^2\right) \\ &\quad \lambda^{-n/2} \exp\left(-\frac{1}{2\lambda} \sum_{i=1}^{n} v_i^2\right) \mathrm{prior}(\tau, \lambda). \end{aligned}$$

2.6 The Gibbs sampler

Fully Bayesian analyses are computationally infeasible, and various approximations have been utilized instead. This changed a decade ago with computer-intensive MCMC simulation methods like Metropolis-Hastings algorithms and the Gibbs sampler. Soon after, a number of researchers began to utilize MCMC methods in actuarial contexts. For details and numerous references, see e.g. Scollnick (1996).

The model complexity, high dimensionality and multimodality of the problem rule out any classical optimization routines, but it is possible to set up a Markov chain whose stationary distribution is consistent with the posterior distribution. One standard approach which produces such a Markov chain is based on a variant of the Metropolis alogorithm called the Gibbs sampler. It will enable us to exploit conditional densities to obtain realizations from the posterior density. After obtaining a sufficient number of realizations, we may use the empirical density generated to find MAPE's.

Moreover, the MCMC approach provides the actuary with much more information than just MAPE. Indeed, for each area a sample from the posterior distribution of x_i is obtained, which allows for interval estimates and precision evaluation. Note that we could also estimate $\boldsymbol{u}$, $\boldsymbol{v}$, τ and λ by approximation to their posterior means

$$\widehat{\boldsymbol{u}} = \mathbb{E}[\boldsymbol{u}|\boldsymbol{y}],\ \ \widehat{\boldsymbol{v}} = \mathbb{E}[\boldsymbol{v}|\boldsymbol{y}],\ \ \widehat{\tau} = \mathbb{E}[\tau|\boldsymbol{y}] \text{ and } \widehat{\lambda} = \mathbb{E}[\lambda|\boldsymbol{y}]$$

obtained from the Gibbs sampler (instead of MAPE).

Let us now describe the Gibbs procedure more formally. Consider a random vector $\boldsymbol{Z}$ with joint probability density function h. In a Bayesian context, some of the components of $\boldsymbol{Z}$ are model parameters, while others may represent unobserved past or future data. Suppose h is so complicated and analytically intractable that it does not permit independent random draws. In this case a MCMC simulation method may be used.

The main idea behind a MCMC method is to simulate realizations from a Markov chain which has h as its stationary distribution. The resulting random draws $\boldsymbol{Z}^{(1)}, \boldsymbol{Z}^{(2)}, \ldots$ are no longer independent, but under mild regularity conditions (as described in the Appendix of Smith & Roberts (1993), for example), the value of $\boldsymbol{Z}^{(t)}$ tends in distribution to that of a random draw from h as t becomes moderately large.

The question is now how to simulate realizations from a Markov chain with h as its stationary distribution. Among the existing competing methods, the simple Gibbs sampler appears to be powerful enough for our purposes. Let us denote as $\boldsymbol{Z}_i$ either the ith component of $\boldsymbol{Z}$ alone, or more generally a block of several components of $\boldsymbol{Z}$ grouped together. Let k be the number of blocks partitioning $\boldsymbol{Z}$, i.e. $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_k)$. The full conditional distribution of the jth block given the remaining variables is $h(\boldsymbol{z}_j|\boldsymbol{z}_i, i \neq j)$. Given an arbitrary vector of starting values $\boldsymbol{Z}^{(0)}$, the first iteration of the Gibbs sampler proceeds by making random draws from the full conditional distribution as follows:

$$\begin{array}{rcl}
\boldsymbol{Z}_1^{(1)} & \sim & h(\cdot|\boldsymbol{Z}_2^{(0)},\ldots,\boldsymbol{Z}_k^{(0)}) \\
\boldsymbol{Z}_2^{(1)} & \sim & h(\cdot|\boldsymbol{Z}_1^{(1)},\boldsymbol{Z}_3^{(0)},\ldots,\boldsymbol{Z}_k^{(0)}) \\
\vdots & \vdots & \vdots \\
\boldsymbol{Z}_j^{(1)} & \sim & h(\cdot|\boldsymbol{Z}_1^{(1)},\ldots,\boldsymbol{Z}_{j-1}^{(1)},\boldsymbol{Z}_{j+1}^{(0)},\ldots,\boldsymbol{Z}_k^{(0)}) \\
\vdots & \vdots & \vdots \\
\boldsymbol{Z}_k^{(1)} & \sim & h(\cdot|\boldsymbol{Z}_1^{(1)},\ldots,\boldsymbol{Z}_{k-1}^{(1)}).
\end{array}$$

This completes a single iteration of the algorithm and defines a transition from $\boldsymbol{Z}^{(0)}$ to $\boldsymbol{Z}^{(1)}$. After t such iterations we have $\boldsymbol{Z}^{(t)}$.

Determining how long a MCMC simulation should be run is a function of the particular application. Usually, several tens of thousands of iterations are enough. In any case, the first portion of the simulated Markov chain is discarded in order to reduce the effect of the starting values. An ad hoc but useful test of convergence is obtained by running several simulations in parallel, with different starting values, and then comparing the results: the number of iterations must be increased if the results look rather different.

2.7 Applications of Gibbs sampler to geographic ratemaking

Let us now apply this method to our problem. At each step, a value for x_i is sampled from the density $p(x_i|\delta_i,\boldsymbol{y})$. The values of the risk parameters in all regions other than i included in δ_i are assumed fixed at their current values in this step. This step involves sampling from each of the distributions subsumed into x_i, i.e. for u_i, v_i, τ and λ. When each x_i has been updated a single cycle of the algorithm is complete, as is one step of a Markov chain. Initial values of the parameters must be supplied.

Let $\boldsymbol{u}_{-i}$ denote all the values in $\boldsymbol{u}$ except u_i. Then, the marginal posterior of u_i is given by

$$\begin{array}{rcl}
p(u_i|\boldsymbol{u}_{-i},\boldsymbol{v},\tau,\lambda,\boldsymbol{y}) & \propto & f(y_i|x_i)p(u_i|\boldsymbol{u}_{-i},\tau) \\
& \propto & f(y_i|x_i)\exp\left(-\dfrac{1}{2\tau}\displaystyle\sum_{j\in\delta_i}(u_i-u_j)^2\right) \\
& \propto & f(y_i|x_i)\exp\left(-\dfrac{\#\delta_i}{2\tau}(u_i-\overline{u}_i)^2\right) \\
& \propto & \exp\left(-e_i\exp(u_i+v_i)+u_iy_i-\dfrac{\#\delta_i}{2\tau}(u_i-\overline{u}_i)^2\right).
\end{array}$$

Similarly, the marginal posterior of v_i is of the form

$$\begin{array}{rcl}
p(v_i|\boldsymbol{v}_{-i},\boldsymbol{u},\tau,\lambda,\boldsymbol{y}) & \propto & f(y_i|x_i)p(v_i|\lambda) \\
& \propto & f(y_i|x_i)\exp\left(-\dfrac{1}{2\lambda}v_i^2\right) \\
& \propto & \exp\left(-e_i\exp(u_i+v_i)+v_iy_i-\dfrac{1}{2\lambda}v_i^2\right).
\end{array} \tag{2.2}$$

To get a sample from distribution whose density function in given in (2.2), we need a method such as adaptive rejection sampling (the form of (2.2) makes direct methods useless) proposed by Gilks & Wild (1992).

Now, let us examine a posteriori distributions for the hyperparameters: for τ,

$$\begin{aligned} p(\tau|\boldsymbol{u},\boldsymbol{v},\lambda,\boldsymbol{y}) &\propto p(\boldsymbol{u}|\boldsymbol{v},\lambda,\tau,\boldsymbol{y})p(\tau|\boldsymbol{v},\lambda,\boldsymbol{y}) \\ &\propto \tau^{-n/2}\exp\left\{-\frac{1}{2\tau}\left(\epsilon+\sum_{i\sim j}(u_i-u_j)^2\right)\right\} \end{aligned}$$

and for λ

$$\begin{aligned} p(\lambda|\boldsymbol{u},\boldsymbol{v},\tau,\boldsymbol{y}) &\propto p(\boldsymbol{v}|\boldsymbol{u},\tau,\boldsymbol{y})p(\lambda|\boldsymbol{u},\tau,\boldsymbol{y}) \\ &\propto \lambda^{-n/2}\exp\left\{-\frac{1}{2\lambda}\left(\epsilon+\sum_{i=1}^{n}v_i^2\right)\right\} \end{aligned}$$

which can be sampled using standard techniques designed for chi-squared distributions.

In practice, we typically run the Gibbs sampler for an initial period of 1 000 cycles and then collect information from a further 10 000 cycles of which we store every 10th or 20th for the subsequent construction of approximate interval estimates. The posterior means are estimated by the corresponding sample means. Since the joint posterior density of $\boldsymbol{u}$ and $\boldsymbol{v}$ given τ, λ and $\boldsymbol{y}$, $p(\boldsymbol{u},\boldsymbol{v}|\tau,\lambda,\boldsymbol{y})$, is log-concave and differentiable, it possesses a single maximum. We are thus in a position to determine the MAPE $\boldsymbol{u}^*$ and $\boldsymbol{v}^*$ of $\boldsymbol{u}$ and $\boldsymbol{v}$, given $\widehat{\tau}$, $\widehat{\lambda}$ and $\boldsymbol{y}$. We have that

$$\sum_{i=1}^{n}v_i^*=0 \text{ and } \sum_{i=1}^{n}e_i\exp(u_i^*+v_i^*)=\sum_{i=1}^{n}y_i.$$

The latter equality is very important for actuarial practice, since it ensures that the resulting tarriff is fair.

All the process can be performed using the software developed in the Department of Actuarial Science and Statistics of the City University, which produces a user friendly implementation of the Boskov and Verrall approach to spatial modelling. The outputs are the MAPE's of u_i and v_i for each postcode region and the expected claim frequency for each region. This software can freely be downloaded from:

http://www.staff.city.ac.uk/r.j.verrall/spatial

3 Other approaches

3.1 Mixture models

Let us mention that simpler techniques based on multi-component mixture models can be used to account for the small number problem alone. Such methods do not incorporate the spatial structure in the data; see e.g. Schlattmann and Böhning (1993).

3.2 Kriging method

The use of (2.1) to model spatial dependence has been criticized by RAFTERY & BANFIELD (1991). Indeed, specifying (2.1) is sensible for spatial arrays not too dissimilar to rectangular arrays. An alternative specification has been developed in geostatistics as the basis of the kriging method. This implements the idea that dependence decreases with distance. Let us describe the kriging approach developed by DIGGLE ET AL. (1998).

Let $\mathcal{R} = \{R(\boldsymbol{s}),\ \ \boldsymbol{s} \in \mathbb{R}^2\}$ be a stationary Gaussian process with

$$\mathbb{E}[R(\boldsymbol{s})] = 0 \text{ and } \mathbb{C}\text{ov}[R(\boldsymbol{s}), R(\boldsymbol{s}')] = \sigma^2 \rho(\boldsymbol{s} - \boldsymbol{s}').$$

Conditionally on the unobserved process $\mathcal{R}$, observations $Y_1, Y_2, \ldots, Y_n$ at sample location $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n$ are Poisson distributed with the corresponding values $R(\boldsymbol{s}_1), R(\boldsymbol{s}_2), \ldots, R(\boldsymbol{s}_n)$ appearing as offset terms in the linear predictor. Conditionally on $\mathcal{R}$, the sample variables $Y_1, Y_2, \ldots, Y_n$ are mutually independent with densities

$$f(\cdot|R(\boldsymbol{s}_i)) = f(\cdot|m_i)$$

specified by the values of the conditional expectations $m_i = \mathbb{E}[Y_i|R(\boldsymbol{s}_i)]$. Now,

$$h(m_i) = R(\boldsymbol{s}_i) + \eta_i$$

for some link function h and linear predictor η_i.

The kriging predictor of $R(\boldsymbol{s})$ is

$$\widehat{R}(\boldsymbol{s}) = \mathbb{E}[R(\boldsymbol{s})|\boldsymbol{Y}].$$

The above assumptions specify explicit forms for the unconditional distribution of $\mathcal{R}$ and from the conditional distribution of $\boldsymbol{Y}$ given $\mathcal{R}$. DIGGLE ET AL. (1998) used MCMC methods to simulate from the conditional distribution of $\mathcal{R}$ given $\boldsymbol{Y}$, and hence to estimate any functional associated with the conditional distribution.

3.3 Smoothing geographical data

Most smoothers apply some sort of average to the y_i and its neighbors, for example, a simple disk average, a disk-weighted average, a linear combination interpretable as the prediction from a regression equation, a median, or a combination of medians. KAFADAR (1994) offers a description and quantitative comparison of commonly applied smoothers. KAFADAR (1996) examined smoothing for geographical data in the form of ratio; see also TALBOT, KULLDORFF, FORAND & HALEY (2000).

MÜLLER, STADTMÜLLER & TABNAK (1997) considered counting data available only in geographically aggregated form. Instead of ascribing the aggregated measurement to the centroid of each area (and then applying standard nonparametric smoothing procedure or kriging methods), these authors proposed smoothing methods taking into account the aggregated nature of the data.

4 Numerical illustration: spatial analysis of a Belgian automobile portfolio

4.1 Presentation of the problem

We have at our disposal observations relating to the 165,363 policies comprised in the portfolio of a major company operating in Belgium (for the year 1997).

Specifically, we have the number Y_i of claims reported by each of the policyholders, together with the following explanatory variables:

Agec : Age of the car (in years).

- $1 = \leq 1$
- $2 = > 1$ and ≤ 3
- $1 = > 3$ and ≤ 6
- $2 = > 6$ and ≤ 9
- $1 = > 9$

Ageph : Age of the policyholder (at the last policy renewal, in years).

- $1 = < 25$
- $2 = \geq 25$ and < 35
- $1 = \geq 35$ and < 60
- $2 = \geq 60$

BM : Bonus-malus in the beginning of the period (in the official Belgian 23-level scale).

- $1 = < 3$
- $2 = \geq 3$ and < 11
- $2 = \geq 11$

Coverage : Type of coverage.

- 1 = TPL only
- 2 = TPL + fire and theft
- 3 = TPL + comprehensive coverage

Duration : Duration of the policy (in days).

Fleet : Fleet code.

- 1 = car belonging to a fleet
- 2 = car belonging not to a fleet

Four : Four-wheel drive code.

- 0 = normal car
- 1 = four-wheel drive

Fuel : Type of fuel.

- 1 = gasoline
- 2 = diesel
- 3 = LPG
- 3 = Other

Monovol : Monovolume type.

- 0 = car of normal type
- 1 = car of monovolume type

Nclaims : Number of claims for the policy in the year 1997.

Power : Power of the engine.

- 1 = < 50
- 2 = ≥ 50 and < 90
- 2 = ≥ 90

Sex : Gender of the policyholder.

- 1 = woman
- 2 = man
- 3 = company
- 4 = society

Sport : Sport code.

- 1 = sport model car
- 1 = 2 = normal model car

Use : Use of the car.

- 1 = private use (leisure and commuting)
- 2 = professional use

Zip : Population code.

- 0 = <2,739 inhabitants
- 1 = number of inhabitants between 2,739 and 19,300

- 2 = number of inhabitants between 19,300 and 38,315
- 3 = >38,315 inhabitants

The methodology is as follows. In a first stage, available explanatory variables are incorporated to the policyholders' claim frequencies with the help of a Poisson regression model. In a second stage, the data are aggregated by districts and overdispersion is accounted for by introduction of a random effect (splitted into a spatially structured part u_i and a spatially unstructured one v_i). For each district, the raw exposure r_i is the number of policy-years from individual data, and the expected frequency e_i is the sum of the annual claim frequencies of policyholders living in that area. The Boskov and Verrall model is then used to recover the spatial structure of the claims pattern. The $\hat{u}_i$'s provided by the model quantify the geographic risk of each district and can be used to design the geographical ratemaking strategy of the company.

4.2 First stage: Poisson regression

The characteristics of each policy are used to predict individual claim frequencies by means of a Poisson regression model. The offset used is the natural logarithm of the risk exposure (in policy-years). Relevant exogeneous variables are selected with the aid of a Type 3 Analysis (a backward-type selection technique implemented in the SAS procedure GENMOD). At each step, the least significant variable is removed from the model, provided this deletion does not significantly deteriorate the quality of the fit. Table 4.1 gives a summary of the removed variables while Table 4.2 contains the variables finally retained in the model with their significance level. We end up with a linear predictor based on nine variables, all highly significant.

Step	Variable removed	DF	Chi-Square	Pr > ChiSq
1	SPORT	1	0.74	0.3912
2	SEX	2	3.65	0.1616
3	USE	1	3.44	0.0638

Table 4.1: Summary of Type 3 Analysis.

Variable retained	DF	Chi-Square	Pr > ChiSq
AGEC	4	73.78	<.0001
AGEPH	3	305.8	<.0001
BM	2	1058.6	<.0001
POWER	2	55.64	<.0001
FUEL	3	140.91	<.0001
COVERAGE	2	14.53	0.0007
FLEET	1	10.27	0.0014
ZIP	3	266.05	<.0001
FOUR	1	5.38	0.0204

Table 4.2: Significance levels of the nine variables retained in the final model.

4.3 Second stage: aggregation of the data by district and spatial analysis

For each district of Belgium, we define Y_i as the sum of the claim numbers reported by all the policyholders living there. Similarly, we build r_i and e_i. The purpose of Figure 4.4.1 is to represent the geographical variation of the $\hat{\theta}_i$'s ($= \frac{y_i}{e_i}$) when exogeneous variables have been taken into account (i.e. the e_i's have been obtained by a Poisson regression incorporating exogeneous information). Specifically, the ratios of actual to expected claim numbers are now mapped (these are a sort of residuals) so that the observed variations should be attributed to spatial effects.

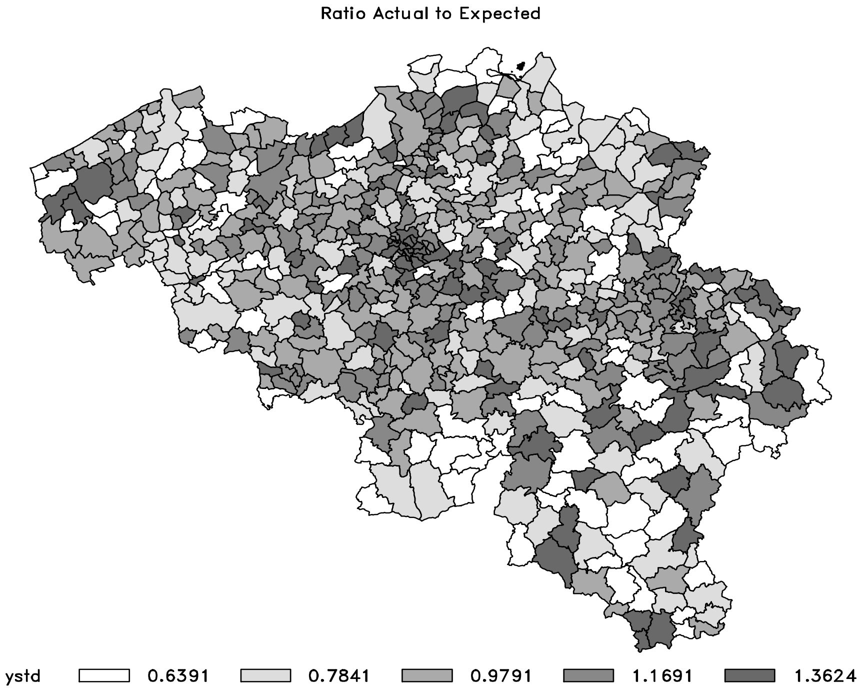

Figure 4.4.1: *Ratios actual to expected.*

The parameters of the procedure are the number of burn-ins, the number of samples and the frequency for sampling. Partial results (run with parameters 10,000-1,000-10) are given in tabular form in Table 4.3; they relate to the province of Antwerp.

In this case, we see clearly that the second estimation is much closer to the observed values than the first one. Moreover, the v_i's are small compared to the u_i's, so that the spatially structured heterogenity dominates the unstructured one. In order to figure out the global claim pattern, we have prepared Figures 4.4.3 and 4.4.4: the first one displays expected claim frequencies $r_i \exp(\eta_i)$ and the second one exhibits the $r_i \exp(\eta_i + u_i + v_i)$'s. The MAPE's of the u_i's are given in Figure 4.4.2; they serve as basis for the determination of the geographical variations in premium amounts. Clearly, several regions with high, medium and low values of the u_i's emerge. Finally, the graph of Figure 4.4.5 illustrates the improvement in fitting produced by spatial smoothing.

District code	District Name	Nbr. of Policies	Actual Nbr of Claims	Nbr of Claims by lin. pred.	Nbr of Claims with geo. rating	U	V
11002	ANTWERPEN 1	4677	638	647.42	650.19	0.025	-0.020
11055	HALLE	633	77	77.30	77.69	0.006	-0.001
11055	ZOERSEL	633	77	77.30	77.69	0.006	-0.001
11023	KAPELLEN	525	74	62.60	67.47	0.064	0.011
11008	BRASSCHAAT	425	62	50.43	55.73	0.089	0.010
11040	SCHOTEN	389	50	46.45	50.06	0.075	0.000
11037	RUMST	376	44	40.83	40.77	-0.007	0.005
11013	EDEGEM	302	35	36.11	37.51	0.042	-0.004
11039	SCHILDE	273	46	32.87	37.33	0.113	0.014
11044	STABROEK	289	39	31.67	35.68	0.114	0.006
11005	BOOM	297	31	31.14	30.91	-0.008	0.000
11009	BRECHT	228	40	28.33	30.46	0.057	0.016
11029	MORTSEL	240	34	28.32	31.31	0.096	0.004
11018	HEMIKSEM	245	28	26.23	26.99	0.027	0.002
11056	ZWIJNDRECHT	241	32	26.04	27.72	0.055	0.007
11024	KONTICH	207	26	25.02	26.03	0.039	0.000
11030	NIEL	237	33	24.79	26.25	0.046	0.011
11001	AARTSELAAR	224	29	24.21	25.38	0.041	0.006
11022	KALMTHOUT	222	18	24.18	20.80	-0.146	-0.005
11035	RANST	190	19	20.27	20.39	0.008	-0.002
11016	ESSEN	186	9	19.81	13.85	-0.350	-0.008
11052	WOMMELGEM	166	20	18.17	19.58	0.074	0.001
11038	SCHELLE	152	8	16.19	15.41	-0.037	-0.012
11007	BORSBEEK	137	21	13.63	15.54	0.122	0.009
11054	ZANDHOVEN	112	11	12.76	12.25	-0.038	-0.002
11004	BOECHOUT	112	14	11.68	12.49	0.065	0.003
11053	WUUSTWEZEL	104	7	10.95	10.51	-0.036	-0.006
11021	HOVE	95	19	10.34	11.73	0.114	0.012
11057	MALLE	91	13	10.17	9.93	-0.029	0.005
11050	WIJNEGEM	80	11	8.08	8.91	0.094	0.003
11025	LINT	72	6	7.73	7.89	0.023	-0.003

Table 4.3: Results Antwerp

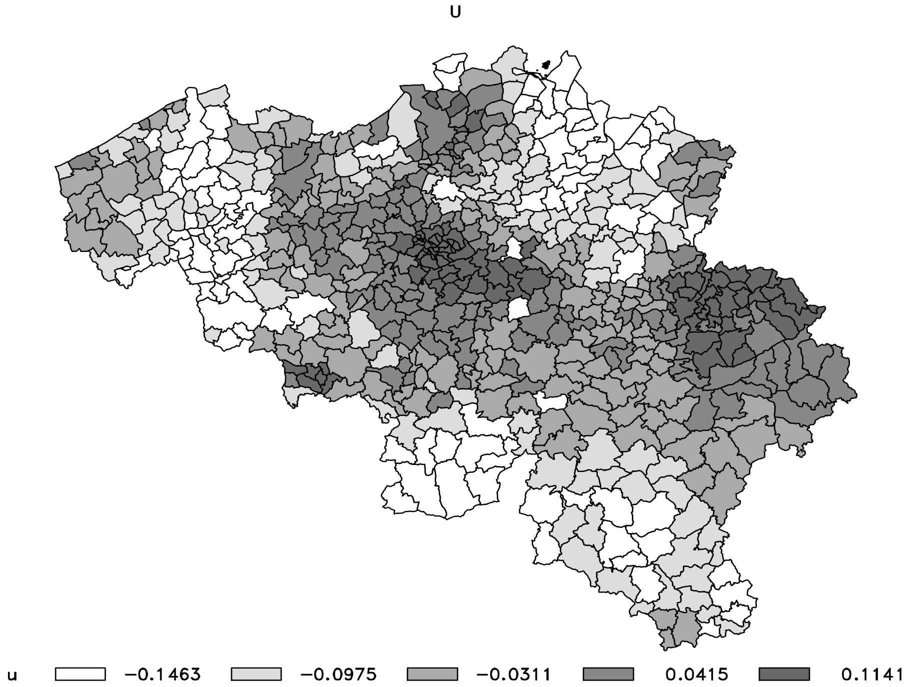

Figure 4.4.2: *MAPE of the u_i's.*

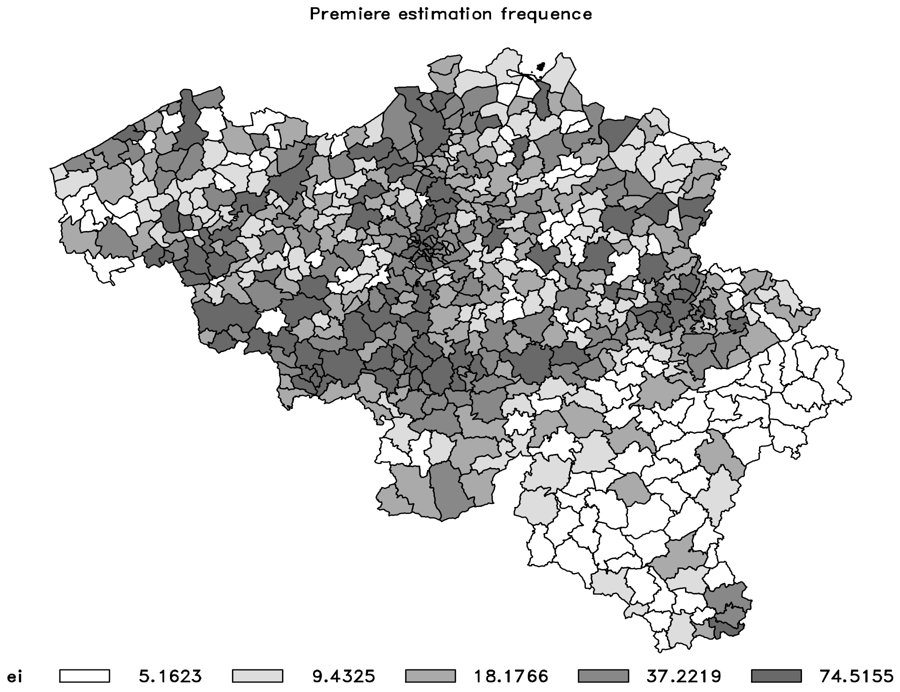

Figure 4.4.3: *Expected number of claims $r_i \exp(\eta_i)$ by district.*

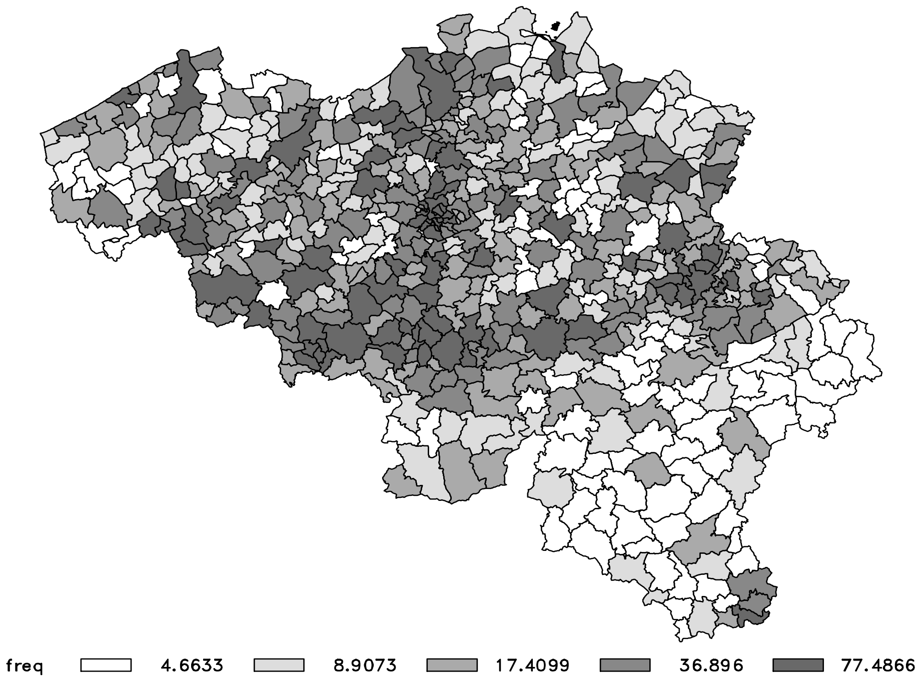

Figure 4.4.4: *Expected number of claims* $r_i \exp(\eta_i + u_i + v_i)$ *by district.*

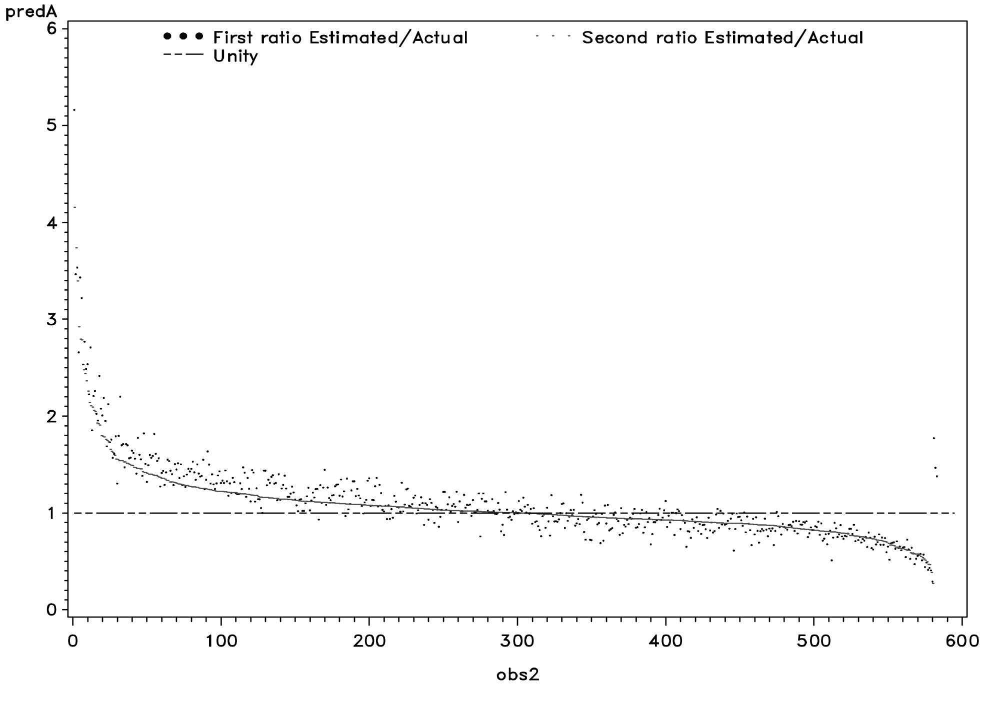

Figure 4.4.5: *Comparison of fitted claim frequencies with and without spatial effect.*

To end with, we have run the program with with different parameter values, to check for the stability of the results. As an illustration, we present the outputs produced with high values of the parameters (50,000-10,000-20). A summary of the results is given in Table 4.4 for the districts in the province of Antwerp. We see here also that the results show a satisfying stability.

Commune NIS	Commune Name	Actual Nbr of Claims	1st Run (short)	2d Run (long)
11001	AARTSELAAR	29	25.38	25.37
11002	ANTWERPEN 1	638	650.19	650.21
11004	BOECHOUT	14	12.49	12.49
11005	BOOM	31	30.91	30.90
11007	BORSBEEK	21	15.54	15.53
11008	BRASSCHAAT	62	55.73	55.71
11009	BRECHT	40	30.46	30.45
11013	EDEGEM	35	37.51	37.50
11016	ESSEN	9	13.85	13.87
11018	HEMIKSEM	28	26.99	26.99
11021	HOVE	19	11.73	11.72
11022	KALMTHOUT	18	20.80	20.80
11023	KAPELLEN	74	67.47	67.45
11024	KONTICH	26	26.03	26.03
11025	LINT	6	7.89	7.88
11029	MORTSEL	34	31.31	31.30
11030	NIEL	33	26.25	26.24
11035	RANST	19	20.39	20.39
11037	RUMST	44	40.77	40.77
11038	SCHELLE	8	15.41	15.42
11039	SCHILDE	46	37.33	37.31
11040	SCHOTEN	50	50.06	50.05
11044	STABROEK	39	35.68	35.67
11050	WIJNEGEM	11	8.91	8.90
11052	WOMMELGEM	20	19.58	19.57
11053	WUUSTWEZEL	7	10.51	10.51
11054	ZANDHOVEN	11	12.25	12.25
11055	HALLE	77	77.69	77.68
11055	ZOERSEL	77	77.69	77.68
11056	ZWIJNDRECHT	32	27.72	27.72
11057	MALLE	13	9.93	9.93

Table 4.4: Comparison of expected numbers of claims generated by two different runs.

5 Conclusion

The present work aims to provide actuaries with a practical method for incorporating spatial effects in a price list. Most companies operating in Belgium have now included some sort of geographical effect in their tariff, ranging from the simple urban-rural dichotomy to elaborated discount systems according to provinces or districts. HB models offer the appropriate framework to quantify the geographically structured heterogeneity. Outputs of the model are easily integrated in a ratemaking. A crucial advantage of the approach developped here is that it is in line with a priori risk classification: both steps are embedded in a Poisson model recognizing the integer nature of claim frequencies and the results are thus coherent. Of course, the Boskov and Verrall method has competitors; we purpose to compare different approaches in a forthcoming work.

Acknowledgements

Natacha Brouhns and Michel Denuit thank the Université Catholique de Louvain for financial support through a “Fonds Spéciaux de Recherche” research grant.

The authors also acknowledge the Observatoire Social et Politique of the Université Catholique de Louvain for the technical support with the realization of the maps. They also benefited from stimulating discussions with Professor Richard Kelsey.

References

[1] Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society - Series B* **36**, 192-236.

[2] Besag, J.E., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 1-20.

[3] Besag, J.E., Green, P., Higdon, D., & Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science* **10**, 3-66.

[4] Boskov, M. & Verrall, R.J. (1994). Premium rating by geographical area using spatial models. *ASTIN Bulletin* **24**, 131-143.

[5] Breslow, N.E., & Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9-25.

[6] Clayton, D., & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics **43**, 671-681.

[7] Cressie, N. (1991). *Statistics for Spatial Data*. Wiley, New York.

[8] Diggle, P.J., Tawn, J.A., & Moyeed, R.A. (1998). Model based geostatistics. *Applied Statistics* **47**, 299-350.

[9] Dixon, M., Kelsey, R. & Verrall, R. (2000). Postcode insurance rating: spatial modelling and performance evaluation. Paper presented at the 4th IME Congress, Barclona.

[10] Gilks, W.R. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Appl. Statist.* **41**, 337-348.

[11] Ghosh, M., Natarajan, K., Waller, L.A., & Kim, D. (1999). Hierarchical Bayes GLM's for the analysis of spatial data: An application to disease mapping. *Journal of Statistical Planning and Inference* **75**, 305-318.

[12] Kafadar, K. (1996). Smoothing geographical data, particularly rates of disease. *Statistics in Medicine* **15**, 2539-2560.

[13] Kafadar, K. (1994). Choosing among two-dimensional smoothers in practice. *Computational Statistics and Data Analysis* **18**, 419-439.

[14] Leroux, B.G. (2000). Modelling spatial disease rates using maximum likelihood. *Statistics in Medicine* **19**, 2321-2332.

[15] Müller, H., Stadtmüller, U., & Tabnak, F. (1997). Spatial smoothing of geographically aggregated data, with application to the construction of incidence maps. *Journal of the American Statistical Association* **92**, 61-71.

[16] Raftery, A.E., & Banfield, J.D. (1991). Stopping the Gibbs sampler, the use of morphology, and other issues in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43**, 32-43.

[17] Schlattmann, P., & Böhning, D. (1993). Mixture models and disease mapping. *Statistics in Medicine* **12**, 1943-1950.

[18] Scollnik, D.P.M. (1996). An introduction to Markov Chain Monte Carlo methods and their actuarial applicaitons. *Proceedings of the Casualty Actuarial Society* **83**, 114-165.

[19] Smith, A.F.M. & Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *J. Royal Statist. Soc., Series B* **55**, No 1.

[20] Talbot, T.O., Kulldorff, M., Forand, S.P., & Haley, V.B. (2000). Evaluation of spatial filters to create smooth maps of health data. *Statistics in Medicine* **19**, 2399-2408.

[21] Taylor, G.C. (1989). Use of spline functions for premium rating by geographical area. *ASTIN Bulletin* **19**, No 1, 91-122.

[22] Taylor, G.C. (1996). Geographic premium rating by Whittaker spatial smoothing. *Technical report, Centre for actuarial studies, University of Melbourne.*

[23] VERRALL, R.J. (1993). A state space formulation of Whittaker graduation, with extensions. *Insurance: Mathematics & Economics* **13**, 7-14.