PREMIUM RATING BY GEOGRAPHIC AREA
USING SPATIAL MODELS

By M. Boskov and R. J. Verrall

Department of Actuarial Science and Statistics
The City University

ABSTRACT

This paper gives a method for premium rating by postcode area The method is based on spatial models in a Bayesian framework and uses the Gibbs sampler for estimation. A summary of the theory of Bayesian spatial methods is given and the data which was analysed by TAYLOR (1989) is reanalysed An indication is given of the wide range of models within this class which would be suitable for insurance data. The aim of the paper is to introduce the models and to show how they can be utilised in an insurance setting.

KEYWORDS

Gibbs sampler; Postcodes; Premium rating; Spatial statistics

1. INTRODUCTION

The problem of accessing risk as a function of geographical area occurs in a number of fields, including insurance rating and epidemiology. The aim of the statistical analysis of the data is to assess the underlying variation in risk by area, usually postcode area. Two approaches can be taken. Either the raw data can be smoothed in order to remove as much random variation as possible, or the data can be used to allocate each postcode area to a rating category, allowing for the inherent random variation The example in this paper uses the first approach, although the methods can also be used for the second approach. The authors believe that the second approach may be more satisfactory if the data are in a suitable form.

The only previous paper, of which the authors are aware, which uses mathematical and statistical techniques for premium rating by postcode area is TAYLOR (1989). That paper 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. The present paper uses an entirely different approach, although some of the preprocessing aspects of the analysis wil be the same as those used by TAYLOR (1989). The example in Section 4 of this paper uses the data from TAYLOR (1989) As will become clear, there are disadvantages in using the data in the form available from that paper. The example is valid in that it applies a suitable model to the particular data set given However, the present authors believe that a slightly different model based on data for claim numbers and claim amounts separately could provide more informative results.

The methods described here are based on statistical methods for spatial data These methods have been developed for image restoration, often using data from satellites. However, the techniques can also be used for risk assessment in an insurance setting The aims of the analysis and some of the assumptions underlying the models differ from those in other applications, but the statistical and mathematical techniques are similar. The basis of the method is the use of a spatial probabilistic model in a Bayesian context. The Gibbs sampler is used to derive the posterior distribution from which inferences about the spatial structure of the data can be made These inferences can be used to assess the risk due to the geographic area. The basic philosophy is that there is an underlying "true" risk pattern over the whole region, and the data are a version of this pattern contaminated by random noise. The aim of the model is to reconstruct the "true" picture as far as possible. The analogy with image restoration is clear

The literature on spatial methods is large, and we mention just a few references which are particulary relevant to the work in this paper. The book by CRESSIE (1991) provides a useful overview and summary of the field. BESAG et al (1991) gives a summary of the Bayesian models and describe applications in archeology and epidemiology. The use of the Gibbs sampler was the subject of a discussion meeting at the Royal Statistical Society recently The papers and discussion are contained in part 1 of the Journal of the Royal Statistical Society, 1993 We would mention particulary GILKS et al. (1993) and SMITH and ROBERTS (1993).

The paper is set out as follows. Section 2 contains a specification of the spatial model. Section 3 describes the Gibbs sampler and simulation techniques which are used to estimate the posterior densities. Section 4 contains an example using the data from TAYLOR (1989) and the final section has the conclusions.

2. A BAYESIAN MODEL FOR SPATIAL DATA

The basis for any model for spatial data is that areas which are close together are more likely to be similar (in some sense) than areas which are far apart. In the context of image restoration, this would mean that adjacent areas would be likely to be the same, or similar, colour. In an insurance context, it means that we expect adjacent areas to be similar from the point of view of the underlying risk.

It is important to remember that we are interested in the true, underlying risk, and the data is just a sample providing an estimate of this risk In addition, we are considering only the risk due to geographical area. We will assume that the other factors have already been analysed, using (for example) a generalised linear model.

We assume that the geographical areas are numbered from 1 to n. Usually, the areas will correspond to postcode areas. Define x_i to be the true risk in area i and $\underline{x}$ to be the vector of risks over the whole region $\{x_i . i = 1, \ldots, n\}$. The joint prior distribution of $\underline{x}$ is not specified explicitly Instead, it is more useful to define the conditional densities

(2 1) $$p_i(x_i | x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

$$i = 1, \ldots, n.$$

This conditional density is the density of the risk at one location, given the risk at all the other locations. In reality, this will not depend on the risk at most of the other locations. This means that we can replace (2 1) by the conditional density of x_i, given the risk values in the neighbourhood of area i.

$$(2\ 2) \qquad\qquad p_i(x_i \mid \delta_i)$$

δ_i is defined as areas in the neighbourhood of the ith area For example, if we had an evenly spaced lattice, the prior distribution might be defined so that δ_i consisted of adjacent points One possibility is illustrated in the following diagram
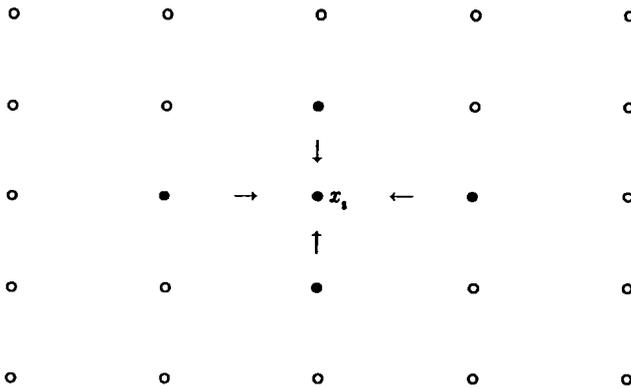
In the insurance setting, δ_i can be interpreted as postcode areas which are adjacent to, or close, to, the ith area.

Suppose that the data observed are denoted by $\underline{y}$ with components $\{y_i \cdot i = 1, \dots, n\}$. We use a simplified notation here, giving only the random variable y_i, and not the other (possibly non-random) information which may be in the data. The full likelihood may be found from

$$(2\ 3) \qquad\qquad f(\underline{y} \mid \underline{x}) = \prod_{i=1}^{n} f(y_i \mid x_i)$$

This assumes, as is reasonable, that the data are conditionally independent, given $\underline{x}$ The posterior density of $\underline{x}$, given $\underline{y}$, can be found using Bayes theorem:

$$(2\ 4) \qquad\qquad p(\underline{x} \mid \underline{y}) \propto f(\underline{y} \mid \underline{x}) \, p(\underline{x})$$

The usual Bayesian estimate of $\underline{x}$ is the value of $\underline{x}$ which maximises the posterior density, the maximum *a posteriori* estimate. Of course, the most diffficult part of this maximisation is to actually determine the posterior density $p(\underline{x} \mid \underline{y})$ Although we have the conditional prior distributions given by (2 2), it is not straightforward to find the unconditional prior distribution and the posterior distribution. Instead, we exploit the conditional densities to obtain realisations from the posterior density. After obtaining a sufficient number of realisations, we may use the empirical density generated to find maximum *a posteriori* estimates. In other words, the

estimation is based on a Monte Carlo method. The mechanics of this, which are based on a variant of the Metropolis algorithm called the Gibbs sampler (GEMAN and GEMAN (1984)) are given in Section 3.

x_i has been defined as the true risk in area i, and we now make the compounds of x_i more explicit. The risk level is assumed to be the sum of three components.

$$(2.5) \qquad\qquad x_i = t_i + u_i + v_i$$

t_i is based on known factors. It is measured through covariates using, for example, a generalised linear model We shall assume that this component of the risk has already been removed from the data. In effect, we assume that the data have already been "standardised" to remove all variation which can be explained by the usual covariates, other than geographic location In the rest of this paper, t_i is therefore dropped from the specification of the model.

u_i represents a component with significant spatial structure.

v_i represents unexplained variation

It is the component u_i that is of interest in an analysis of the spatial structure of the data

We must now formulate the conditional prior distribution of $x_i | \delta_i$, (2.2), in terms of u_i and v_i. Henceforth t_i is ignored since it has already been removed from the data. It is reasonable to assume that u_i and v_i are independent Also, since there are no reasons to use any other distribution, we shall use a normal prior distribution for $\{v_i : i = 1, \ldots, n\}$

$$(2.6) \qquad\qquad p(v_i) \propto \lambda^{-1/2} \exp\left(-\frac{1}{2\lambda} v_i^2\right)$$

We have assumed that the risk at the ith region depends only on regions which are in the neighbourhood of the ith region It is also assumed that the prior conditional density of the spatial component, u_i, can be factorized into components representing the dependencies on each of the neighbouring regions and hence can be written as

$$(2.7) \qquad p_i(u_i | u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n) \propto \exp\left(-\sum_{j \in \delta_i} \phi(u_i - u_j)\right)$$

for some function ϕ. Note that the summation in (2 7) is only over j in δ_i

The function ϕ must reflect the fact that the spatial dependence will reduce as the distance between the regions increases It must therefore favour similar values for regions which are adjacent, and can be any even function. It could be preceded by a factor to allow for the precise proximity of the regions i and j. In this case, (2 7) is replaced by

$$(2\ 8) \qquad p_i(u_i | u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n) \propto \exp\left(-\sum_{j \in \delta_i} w_{ij} \phi(u_i - u_j)\right)$$

Possible choices for ϕ include

$$\phi(z) = \frac{z^2}{2\varkappa} \qquad \text{and} \qquad \phi(z) = \frac{|z|}{\varkappa}.$$

In this paper, we use the first of these possibilities. Thus,

$$(2\,9) \quad p_i(u_i | u_1, u_2, \quad , u_{i-1}, u_{i+1}, \ldots, u_n) \propto \exp\left(-\frac{1}{2\varkappa} \sum_{j \in \delta_i} (u_i - u_j)^2\right)$$

The two scale parameters $\varkappa$ and λ, which determine the variances of u and v must also be given a prior distribution. A suitable choice for this prior distribution, which is close to the usual uninformative distribution but which avoids technical difficulties is

$$(2.10) \qquad \qquad \text{prior}(\varkappa, \lambda) \propto \exp\left(-\frac{\varepsilon}{2\varkappa} - \frac{\varepsilon}{2\lambda}\right)$$

where ε is a small positive constant, say $0\,01$. For a more detailed discussion of this choice, see BESAG et al (1991)

The conditional prior distribution for $x_i | \delta_i$, (2 2), can now be replaced by the prior distributions of $\underline{u}$, $\underline{v}$, $\varkappa$ and λ The posterior density of the parameters can be found as in (2 4), using Bayes theorem:

$$(2\,11) \qquad p(\underline{u}, \underline{v}, \varkappa, \lambda | \underline{y}) \propto \prod_{i=1}^{n} f(y_i | x_i) \varkappa^{-n_i/2} \times$$

$$\times \exp\left(-\frac{1}{2\varkappa} \sum_{j \in \delta_i} (u_i - u_j)^2\right) \lambda^{-1/2} \exp\left(-\frac{1}{2\lambda} v_i^2\right) \text{prior}(\varkappa, \lambda)$$

where n_i is the cardinality of δ_i.

Note that the joint prior distribution of $\underline{u}$ has been obtained from the conditional prior densities, (2.9), using the derivation given in Section 2 of BESAG (1974) Various forms for $f(y_i | x_i)$ are appropriate for insurance data. In the example in Section 4, we use a normal distribution For data on claim numbers a Poisson distribution would be appropriate In the case of Poisson data, it is usual to assume that the mean of this distribution is $c_i e^{x_i}$, where c_i is the expected number of claims in region i ignoring the spatial effect. Then

$$(2\,12) \qquad\qquad f(y_i | x_i) = \frac{\exp(-c_i e^{x_i})(c_i e^{x_i})^{y_i}}{y_i!}$$

A normal distribution for $f(y_i | x_i)$ is also useful in practice. The mean and variance of this distribution will depend on the application, and an example of this case is given in Section 4.

3. THE GIBBS SAMPLER

Having defined the Bayesian model, the remaing problem is to obtain maximum *a posteriori* estimates for the parameters The complexity, high dimensionality and multimodality of the problem rules out any normal optimization routines. However, it is possible to set up a Markov chain whose stationary distribution is consistent with the posterior distribution. One approach which produces such a Markov chain is called the Gibbs sampler. The principle of the Gibbs sampler is as follows

At each step in the chain the current value of each parameter is replaced by a new one which is chosen randomly from its distribution given all the other parameter values and the observed data Thus, in the terminology of Section 2, a value for x_i is sampled at random from the density

$$(3 1) \qquad\qquad p_i(x_i | \delta_i, \underline{y})$$

The values of the risk parameters in all regions other than i, including 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, $\varkappa$ and λ. Initial values of the parameters must be supplied.

Typically, the chain must be allowed to run for 1,000 steps before it will have converged to its stationary distribution, which can be used to find the maximum *a posteriori* estimates for the parameters. Once convergence has been obtained, a sample of every 10th step over the next 10,000 steps usually provides a reasonable estimate of the stationary distribution. This can be treated as a an empirical distribution from which the required estimates can be obtained in the usual way.

Note that the conditional posterior distributions which are required by the Gibbs sampler can be obtained in a straightforward manner. For example,

$$(3.2) \qquad\qquad p(u_i | u_{-i}, \underline{v}, \varkappa, \lambda, \underline{y}) \propto f(y_i | x_i) p(u_i | u_{-i}, \varkappa)$$

where u_{-i} denotes all values in $\underline{u}$ except u_i.

For example, when the data have Poisson distributions and the posterior density is given by (2.11) and (2.12), then the marginal posterior of u_i is given by

$$(3.3) \qquad p(u_i | u_{-i}, \underline{v}, \varkappa, \lambda, \underline{y}) \propto \exp\left(-c_i e^{u_i + v_i} + u_i y_i - \frac{n_i}{2\varkappa}(u_i - \bar{u}_i)^2 \right)$$

where $\bar{u}_i$ is the mean value of u_i over δ_i Details of the marginal distributions of the other parameters in the case of Poisson data can be found in BESAG et al. (1991)

Once the marginal densities have been found and initial values of all the parameters supplied, the Gibbs sampler can be used to generate values of the parameters from the required posterior distribution. In effect, the procedure exploits the simpler conditional distributions to simulate the posterior distribution.

In some cases the random sampling does not present any problems. For example, when the data are normally distributed, the posterior distributions are also normal and the sampling procedure described above is fairly straightforward In other cases, the posterior distributions are more complicated and samples cannot be obtained by

a direct method. Instead, a method such as adaptive rejection sampling must be used. It is very important that the sampling procedure and the computational approach are highly efficient in order to produce results reasonably fast. A particularly efficient from of rejection sampling is described by GILKS and WILD (1992). This form of sampling has to be used, for example, in the case of Poisson data. We now summarise the sampling process as described in greater detail in GILKS and WILD (1992).

Suppose a sample is required from the distribution whose density function is $f(x)$. For example, this density might take the form given in (3.3). The density, $f(x)$, need only be known up to a constant of integration. i.e. instead of knowing $f(x)$, we may only know $g(x)$ where

(3.4) $$g(x) = cf(x)$$

and c is an unknown constant.

It is necessary to define an envelope function $g_u(x)$ such that $g_u(x) \geq g(x)$ for all x, and a squeezing function $g_l(x)$ such that $g_l(x) \leq g(x)$ for all x. The procedure to obtain a sample from $f(x)$ is then as follows

Take a sample x^* from $g_u(x)$ and a sample w from $U(0, 1)$. Now use the squeezing function to test the value

if $w \leq \dfrac{g_l(x^*)}{g_u(x^*)}$ then accept x^*; if not then test

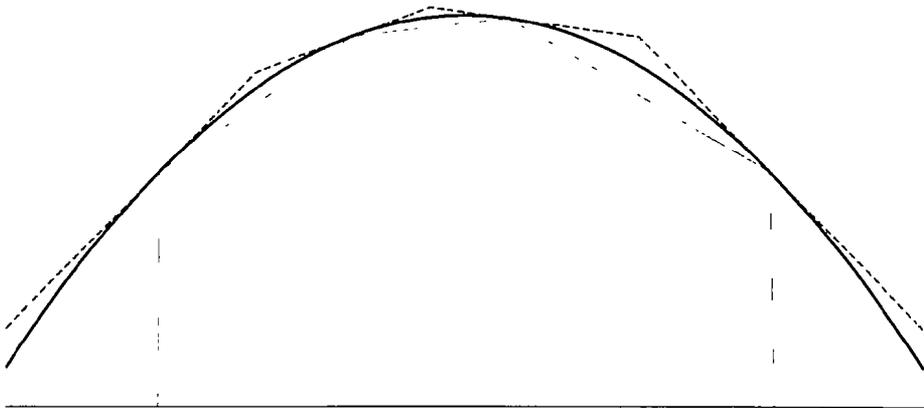if $w \leq \dfrac{g(x^*)}{g_u(x^*)}$ then accept x^*; otherwise reject x^*

FIGURE 1

After each rejection of a sample value, the envelope and squeezing functions are redefined so as to reduce the probability of further rejection If the log density, $h(x) = \log(g(x))$ is considered, it can be seen that for the density (3.3), and for many others, $h''(x) < 0, \forall x$. It is therefore possible to define an envelope $h_u(x) = \log(g_u(x))$ where $h_u(x)$ is a piecewise linear function such that each line segment is a tangent to $h(x)$. Similarly, a piecewise linear function

$h_l(x) = \log(g_l(x))$ can be defined by chords meeting $h(x)$ at the same points as $h_u(x)$.

After each rejection of a value of x^*, this value is added to the set of points at which $h_l(x)$ and $h_u(x)$ meet $h(x)$ GILKS and WILD (1992) show that this provides an efficient method of generating samples for the Gibbs sampler

4. EXAMPLE

In order to illustrate the methods and to give an indication of the nature of the results, the data from TAYLOR (1989) are reanalysed in this section We would emphasise that this is really an illustration and does not represent a definitive rating conclusion. In particular, we would prefer to analyse claim numbers and claim amounts separately · see Section 5 for a more detailed discussion However, this example does enable the results to be compared with the method used by Taylor, which imposed a much greater degree of smoothness onto the results

The data relates to Household Contents Insurance in and around Sydney, Australia This region is divided into approximately 200 postcode areas The data have already been processed to remove the effects of all factors which can be modelled using generalised linear modelling techniques. All factors corresponding to t_i in (2.5) have been controlled out in order to make the data suitable for investigating the spatial effects Taylor also included a "rough fit of the 'geographic area effect'" in order to improve the fit of the other factors but this effect was, of course, not controlled out The final data used in this example consists of adjusted loss ratios.

The adjusted loss ratios are assumed to be normally distributed:

$$y_i \mid x_i \sim N\left(x_i, \frac{\alpha}{e_i}\right)$$

where e_i is the earned exposure in postcode area i,

and α is a constant, chosen as indicated below.

As noted in TALOR (1989), this normal approximation may be poor where e_i is small. However, in the model considered here, areas with low values of e_i will have a limited effect on the overall results. The constant α controls the amount of smoothing applied, as can be seen from the following maps. The maps show the values of the adjusted loss ratios divided into six bandes as follows ·

A	Less than 0.5
B	0.5 to 0 7
C	0.7 to 0 9
D	0.9 to 1 1
E	1 1 to 1.3
F	Greater than 1.3

Map 1 shows the adjusted loss ratios of the raw data before the fitting of the spatial model. Maps 2 to 5 show those of the fits for various values of α. A value of 100 appears to be produce a similar level of smoothing to that achieved by Taylor

and the overall pattern is very similar with areas of low risk in the south-east and north-east corners and a band of high risk just south of the river.

A referee has pointed out that a value of α of around 100 can be justified as follows

$$\alpha = \text{variance of loss ratio for a single risk}$$

If it is assumed that losses occur according to a Poisson process with rate θ and that the first and second moments of the distribution of the size of individual losses are μ_1 and μ_2, then
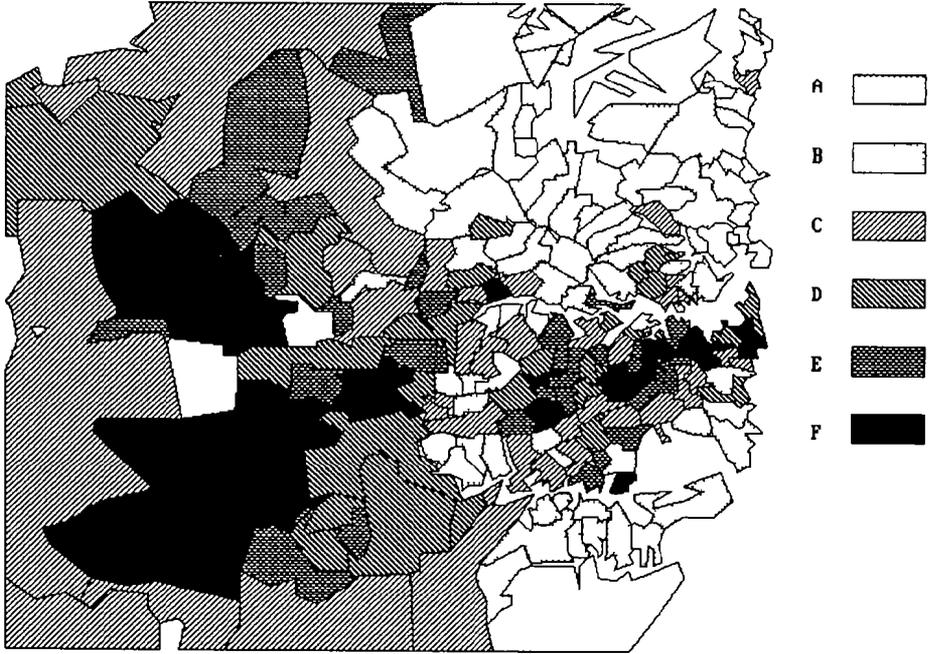
$$\alpha \simeq \frac{\theta\mu_2}{(\theta\mu_1)^2}$$

or

$$\alpha \simeq \frac{(1 + r)^2}{\theta}$$

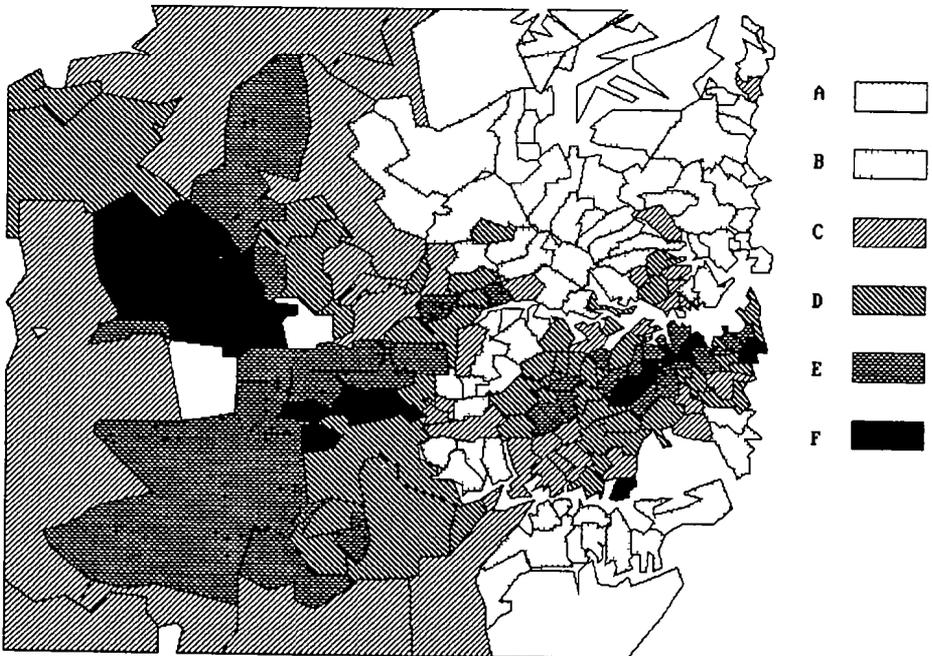where $r = $ coefficient of variance of claim size.

From the data the observed value of θ is approximately 0 1, so that $\alpha = 100$ corresponds to a value of r of 3 which seems reasonable. However, the choice of value for α should be a pragmatic one based on the level of smoothing which is thought to be appropriate
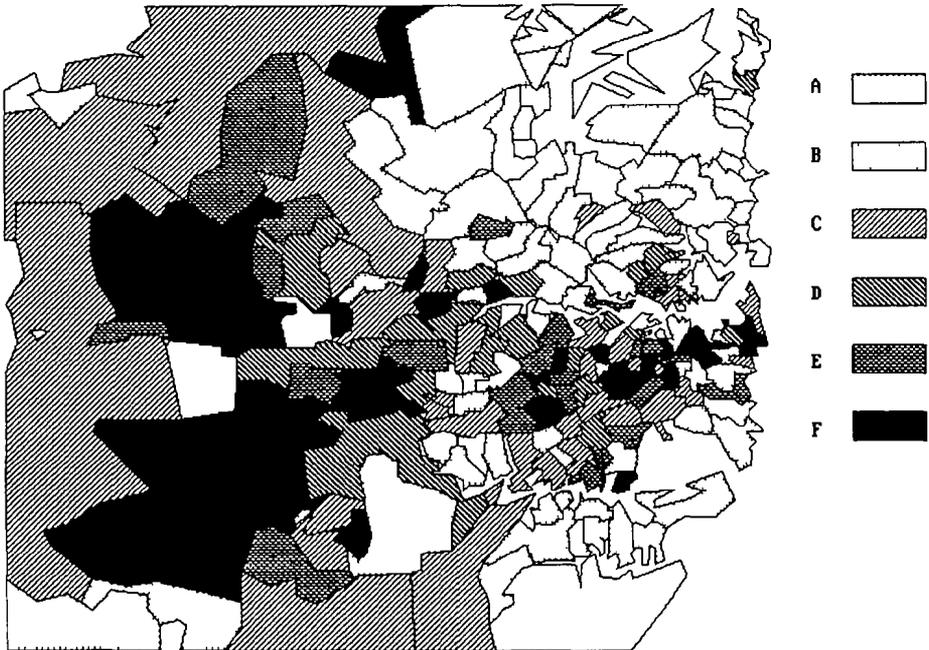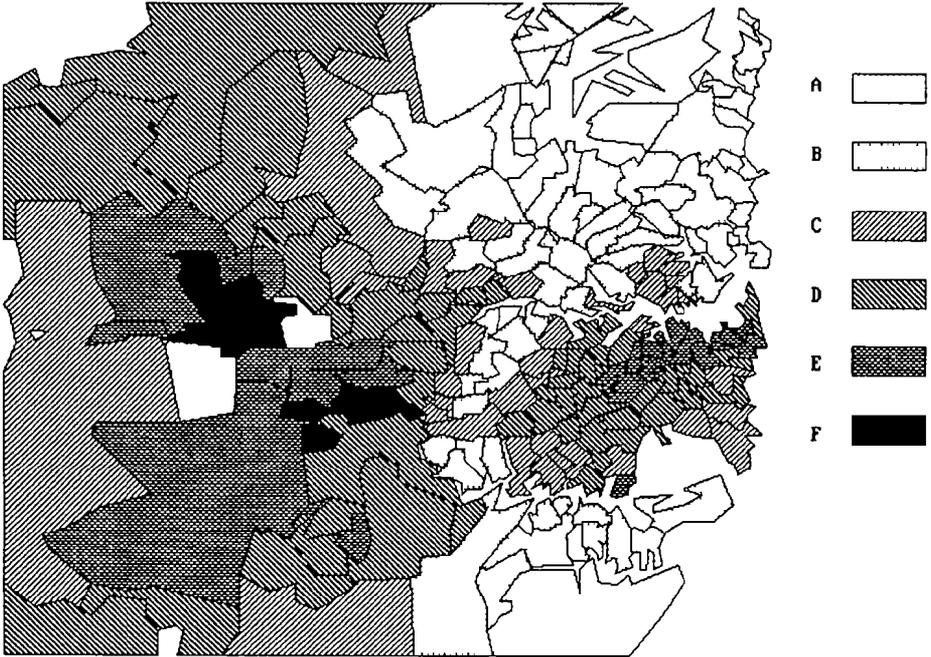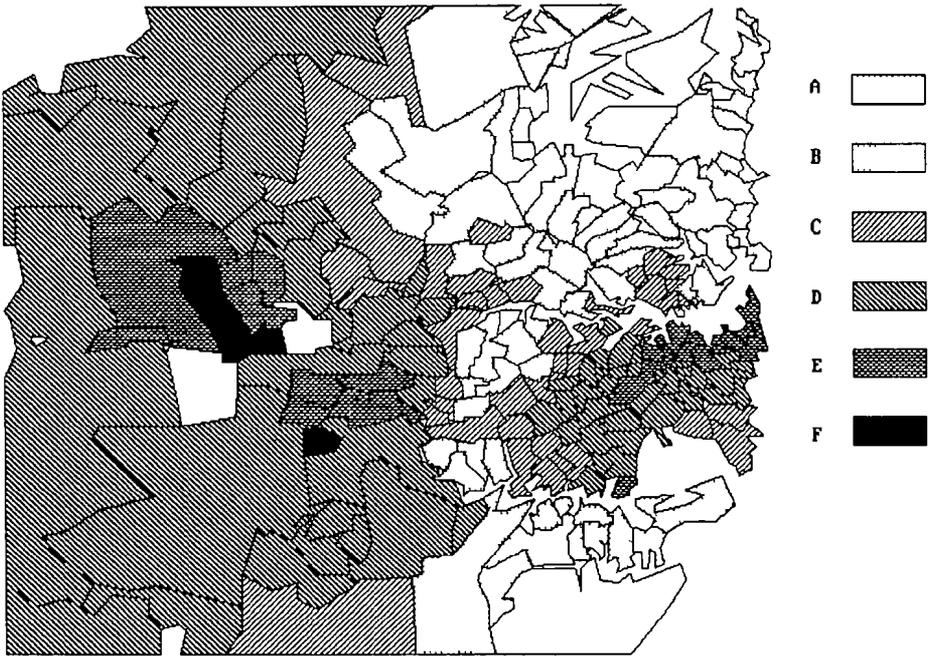

MAP 1   Raw Data

MAP 2 Smoothed values $\alpha = 20$

MAP 3 Smoothed values $\alpha = 50$

MAP 4 Smoothed values $\alpha = 100$

MAP 5 Smoothed values $\alpha = 200$

5 CONCLUSIONS

This paper has described how spatial statistical models can be used to analyse the geographic area effect in insurance data. The methods are applicable to all data in which there is a geographic area effect. The authors believe that the potential for these methods in insurance applications is great, and that they represent the best way of premium rating by postcode area.

The example has been approached from the point of view of smoothing the data over the postcode areas, using a continuous scale for the rating results These smoothed results have then been divided into bands for rating purposes An alternative approach would be to use a discrete scale for the results, corresponding to the required number of rating classes The spatial model would then be required to allocate each postcode area to one of the rating classes The use of this type of model is at present under investigation.

Unlike the method used in TAYLOR (1989) this method could easily be extended to an entire country rather than just one metropolitan area

It would be preferable to analyse the data for claim numbers and claim amounts separately This approach is already used to model claims experience with respect to other factors c.f RENSHAW (1993) Such a separation is particularly important in cases where claim severity has a long tailed distribution (e g. liability) where one large claim could dominate the loss ratio of a small area It may also prove to be the case that a simpler model using only a few of the factors is appropriate for claim severity while a more complicated model including spatial data can be used for the frequency. This involves more complex computations since the data would no longer be normally distributed. Again, this is under investigation and will be the subject of a subsequent paper.

REFERENCES

BESAG, J (1974) Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion) *J Royal Statist Soc , Series B*, **Vol. 36**, No 2

BESAG, J , YORK, J and MOLLIE, A (1991) Bayesian Image Restoration, with Applications in Spatial Statistics (with discussion) *Ann Inst Statist Math* , **Vol. 43**, 1–59

CRESSIE, N (1991) *Statistics for Spatial Data* John Wiley and Sons, New York

GEMAN, S and GEMAN, D (1984) Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images *I E E E Trans Pattn Anal Mach Intell* , **Vol. 6**, 721–741

GILKS, W R , CLAYTON, D G , SPIEGELHALTER, D J , BEST, N G McNEIL, A J , SHARPLES, L D and KIRBY, A J (1993) Modelling Complexity Applications of Gibbs Sampling in Medicine *J Royal Statist Soc , Series B*, **Vol. 55**, No 1

GILKS, W R and WILD, P (1992) Adaptive Rejection Sampling for Gibbs Sampling *Appl Statist* , **Vol. 41**, 337–348

RENSHAW, A E (1993) *Modelling the Claims Process in the Presence of Covariates* ASTIN Colloquium

SMITH, A F M and ROBERTS, G O (1993) Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods *J Royal Statist Soc , Series B,* **Vol. 55**, No 1
TAYLOR, G C (1989) Use of Spline Functions for Premium Rating by Geographic Area *ASTIN Bulletin* **Vol. 19**, No 1, 91–122

MICHAEL BOSKOV
Department of Actuarial Science and Statistics, The City University,
Northampton Square, London EC1V 0HB, England.