A.A.S. **2**, I, 91-114 (2007)

BAYESIAN NONPARAMETRIC APPROACH TO CREDIBILITY MODELLING

By A. Gangopadhyay and W.-C. Gau

ABSTRACT

Current methods in credibility theory often rely on parametric models, e.g., a linear function of past experience. During the last decade, the existence of high speed computers and statistical software packages allowed the introduction of more sophisticated and flexible modelling strategies. In recent years, some of these techniques, which made use of the Markov Chain Monte Carlo (MCMC) approach to modelling, have been incorporated in credibility theory. However, very few of these methods made use of additional covariate information related to risk, or collection of risks; and at the same time account for the correlated structure in the data. In this paper, we consider a Bayesian nonparametric approach to the problem of risk modelling. The model incorporates past and present observations related to risk, as well as relevant covariate information. This Bayesian modelling is carried out by sampling from a multivariate Gaussian prior, where the covariance structure is based on a thin-plate spline. The model uses the MCMC technique to compute the predictive distribution of future claims based on available data.

KEYWORDS

Thin-Plate Spline; Markov Chain Monte Carlo (MCMC); Loss Models; Credibility

CONTACT ADDRESSES

A. Gangopadhyay, Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA 02215, U.S.A. Tel: +1-617 353 2560; Fax: +1-617 353 8100; Email: ag@math.bu.edu
W.-C. Gau, Department of Statistics and Actuarial Science, University of Central Florida, Orlando, FL 32816, U.S.A. Tel: +1-407 823 5532; Fax: +1-407 823 3930; Email: wgau@mail.ucf.edu

1. Introduction

Many problems in actuarial science involve the prediction and assessment of future risks, particularly short-term risks. Over the last 40 years, various methodologies have been proposed, ranging from techniques based on linear models, to more complex models incorporating nonparametric and semiparametric structures. Current methods in credibility theory often rely on parametric models. Bühlmann (1967) developed an approach based on the best linear approximation, which leads to an estimator which is a linear combination of current observations and past records. In recent years,

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

several authors have introduced methods which use more flexible nonparametric and semiparametric statistical modelling techniques. Young (1997) used a decision theoretic approach to develop a credibility formula via minimisation of a loss function which is a linear combination of squared-error and second-derivative terms — i.e., a spline-based approach to modelling.

Bayesian methodology is particularly appealing in the context of credibility modelling. Most interesting data sets in credibility theory consist of repeated, but short, time series data. A Bayesian approach uses probability models to describe the inherent uncertainty of such data, and provides an ideal modelling framework. Mayerson (1964) gave a Bayesian view of credibility, and points the way towards the construction of a sound theoretical foundation in credibility theory. Jewell (1974) showed that credibility premiums are exact Bayesian premiums for all simple exponential families. Klugman (1992) provided a Bayesian analysis to credibility theory by carefully choosing a parametric conditional loss distribution for each risk and a parametric prior. Recent developments in Bayesian computational techniques, such as Gibbs sampling, have made it possible to consider complex and more realistic Bayesian models suitable for actuarial applications. Scollnik (2000) described several applications of the Markov Chain Monte Carlo (MCMC) technique to actuarial modelling using BUGS (Spiegelhalter *et al.*, 1996) software. Young (2000) considered a Robust Bayesian Credibility model, and showed its superiority to linear model-based approaches to data analysis.

Recently, there has been a fundamental shift in the application of Bayesian modelling. The traditional thinking in the Bayesian paradigm as an extension of parametric models (in the sense that the priors on parameters are used to diffuse the dependence on a specific parametric structure) has been replaced by nonparametric and semiparametric models, where priors are used in a novel way to facilitate identification and estimation of unknown functions. This approach incorporates the flexibility of nonparametric and semiparametric models, and at the same time takes advantage of Bayesian computation to facilitate the technical aspects of model fitting.

The key difficulty in incorporating the Bayesian framework in the context of nonparametric and semiparametric modelling has to do with the selection of prior distributions on functions. One possible solution to the problem was introduced by Wahba (1990, 2000). In this approach, a multivariate Gaussian process with a covariance structure is derived via a thin-plate spline, which can be used to define a prior distribution on an unknown regression function. The procedure has been used successfully in building flexible probability distributions which can be used in the context of Bayesian nonparametric modelling. This basic idea was used in modelling medical insurance data in a discussion paper by Gangopadhyay & Gau (2003). It should be noted that this powerful modelling framework is quite general, and

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

can be used successfully in actuarial loss modelling. In this paper, however, we focus on the specific application which relates to credibility modelling.

In Section 2 we introduce the general loss model problem, and the associated data structures. In Section 3 we discuss the methodology, and provide a detailed MCMC algorithm for fitting a flexible nonparametric Bayesian model to actuarial data. In Section 4 we illustrate the methodology in detail via a simulated example and a motivating example related to annual fire damages in various towns in eastern Massachusetts, United States of America.

2. Data Structure in Loss Models

The classical data types in this area involve realisations from present and past experiences of policyholders. Suppose that we have n different risks (or policyholders), with claim records over a certain number of years, say T, i.e.:

$$\begin{aligned} &Y_{11}, Y_{12}, \ldots, Y_{1T} \\ &Y_{21}, Y_{22}, \ldots, Y_{2T} \\ &\ldots\ldots\ldots\ldots\ldots \\ &Y_{n1}, Y_{n2}, \ldots, Y_{nT}. \end{aligned}$$

The data can be the amount of loss, the number of claims, or the loss ratio from insurance portfolios.

Our goal is to estimate the amount (or the number) of claims to be paid on a particular insurance policy in a future coverage period. The problem of interest is to model the relationship of $\mathbf{Y}_{T+1}$ to the past observed values of $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_T$, i.e. to establish the relationship:

$$y_{ij} = f(y_{i,j-1}, y_{i,j-2}, \ldots, y_{i,1}) + e_{ij} \quad \text{for } i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, T \tag{1}$$

where f is an unknown function and e_{ij} is a zero-mean random error term.

We can also have a more general form of data configuration. Let $\{y_{ij} : j = 1, 2, \ldots, T_i;\ i = 1, 2, \ldots, n\}$ be a time series of length T_i over which the measurements of the ith risk (or group of risks) were observed at time points t_{ij}, $j = 1, 2, \ldots, T_i$. In addition, let $\mathbf{X}_{ij}$ be the observed covariates, such as the gender of a policyholder, the industry type of insureds, the time signature, etc., for the ith risk (or group of risks) at time t_{ij}. In each risk (or group of risks), the data have the form:

$$(y_{ij}, \mathbf{X}_{ij}),\ j = 1, 2, \ldots, T_i \quad i = 1, 2, \ldots, n \tag{2}$$

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

where $\mathbf{X}_{ij} = (X_{ij1}, X_{ij2}, \ldots, X_{ijd})$ is the d-dimensional covariate vector measured at time t_{ij}. It may be of interest to study the association between the current response y_{ij} and the past responses $y^*_{i,j-1} = (y_{i,j-1}, y_{i,j-2}, \ldots, y_{i1})$, along with the covariates.

Therefore, we propose the following modified model, a generalisation of (1):

$$y_{ij} = f(\mathbf{X}_{ij}, y^*_{i,j-1}) + e_{ij} \quad \text{for } i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, T_i. \tag{3}$$

The main contribution of the present work is to provide a flexible nonparametric environment for loss modelling. As opposed to traditional approaches, the modelling framework presented here does not assume any specific parametric structure for the unknown function f. In other words, the procedure lets the data speak for themselves.

One key question is how to estimate the unknown function f in an efficient manner. In the nonparametric and semiparametric literature, various estimation methods are available (e.g., kernel, nearest-neighbour, spline, wavelet, etc.). However, nonparametric estimation of multivariate functions can be difficult. The problem is particularly challenging when only a limited amount of data is available. In the context of loss modelling, usually only a few years of data for each risk are available. However, there may be a large number of risks from which to borrow strength. Thus, both computationally and conceptually, a Bayesian approach is most suited in this situation.

In this paper, a Bayesian approach is presented for nonparametric multivariate regression with Gaussian errors. A smooth prior based on thin-plate splines is assumed for each component of the model. We use the reproducing kernel for a thin-plate spline for an unknown multivariate function, as in Wahba (1990). All computations are carried out using the Gibbs sampling scheme, which follows the work of Wood *et al.* (2002). This model is used for the estimation of the function f, and the estimate is subsequently used to predict future values of Y_{i,T_i+1}. In the next section we will introduce the theoretical foundation behind the modelling strategy. The mathematical details are quite challenging. However, we provide a heuristic argument which is sufficient for understanding the methodology being introduced in this paper.

3. Methodology Based on Thin-Plate Splines

The dictionary definition of a spline is: "a thin strip of wood used in building construction." This, in fact, gives an insight into the mathematical definition of splines. Historically, engineering draftsmen used long thin strips of wood, called splines, to draw a smooth curve between specified

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

points. A mathematical spline is the solution to a constrained optimisation problem.

The theoretical foundations for thin-plate splines can be attributed to Duchon (1975, 1976, 1977) and Meinguet (1979). Duchon's surface spline is called a 'thin plate' spline, because it approximates the equilibrium position of a thin plate deflected at scatter points. For an application of thin-plate splines to meteorological problems see Wahba & Wendelberger (1980). Some computational advantages of using thin plate regression splines are addressed by Wood (2003).

3.1 *Connection Between Thin-Plate Splines and Bayesian Methods*

Following the work of Wahba (1990), we define a penalty functional:

$$J_m(f) = \int (f^{(m)}(v))^2 dv.$$

It is assumed that data $\mathbf{y} = (y_1, \ldots, y_n)'$ follow the model:

$$y_i = f(v_i) + \epsilon_i \quad i = 1, 2, \ldots, n \tag{4}$$

where v_i stands for the covariate information related to observation y_i. This model is a simplified version of the model introduced in (1) and (3). As before, f is assumed to be a smooth function, and $\{\epsilon_i\}$ are independent zero mean errors with a common unknown variance. It is desired to find an estimate of f given $\mathbf{y} = (y_1, \ldots, y_n)'$. The estimate f_λ of f will be taken as the minimiser of:

$$\sum_{i=1}^{n} (y_i - f(v_i))^2 + \lambda J_m(f). \tag{5}$$

Note that this criterion can be described as:

$$\text{Goodness of fit} + \lambda \text{ (Measure of smoothness)}.$$

Thus, the parameter λ provides a balance between the lack of fit and the smoothness of f estimated from the data. It can be shown that (5) can also be written as:

$$(\mathbf{y} - \mathbf{f})'(\mathbf{y} - \mathbf{f}) + \lambda(\mathbf{f}'\mathbf{K}\mathbf{f}) \tag{6}$$

where:

$$\mathbf{y} = (y_1, \ldots, y_n)'$$
$$\mathbf{f} = (f(v_1), \ldots, f(v_n))'$$

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

and $\mathbf{K}$ is a penalty matrix. It is easy to see that the solution to (6) is given by:

$$\widehat{\mathbf{f}} = \mathbf{K}(\mathbf{K} + \lambda \mathbf{I})^{-1}\mathbf{y}. \tag{7}$$

Now consider solution (7) from a different view point for the model given in (4), where we assume $\epsilon_i \overset{iid}{=} \sim N(0, \sigma^2)$. A Bayesian approach to the estimation of f can be formulated as follows.

Consider $\mathbf{f}$ as an n-dimensional vector of unknown parameters, and assume that $\mathbf{f}$ follows a multivariate normal prior distribution with zero mean and covariance matrix, with the (i, j)th entry given by:

$$\mathrm{E}(f(v_i)f(v_j)) = b \times \mathbf{K}(v_i, v_j)$$

for some constant b and function $\mathbf{K}$. Then it is easy to verify that the posterior mean (as the estimate of f under squared error loss) is given by:

$$\widehat{\mathbf{f}} = \mathrm{E}(\mathbf{f} \mid \mathbf{y}) = \mathbf{K}\left(\mathbf{K} + \frac{\sigma^2}{b}\mathbf{I}\right)^{-1}\mathbf{y} \tag{8}$$

where $\mathbf{K}$ is a positive definite matrix, with $\mathbf{K}(v_i, v_j)$ being the (i, j) entry of the matrix.

A comparison of (8) and (7) shows the underlying connection between the spline and the Bayesian solutions of the estimation problem under a Gaussian prior. This important observation was made by Wahba (2000), and has been extensively used in the Bayesian nonparametric literature. This result, along with the Bayesian Markov Chain Monte Carlo method, substantially simplifies the problem related to the estimation of f in high dimension.

At this point, we will briefly discuss the multivariate generalisation of the connection between the spline and the Bayesian methods of the estimation of the unknown function f. In the multivariate setup, we will assume $f \in \chi$, a space of functions whose partial derivatives of total order m exist and are continuous. The model is given by:

$$y_i = f(v_1(i), \ldots, v_d(i)) + \epsilon_i \quad i = 1, 2, \ldots, n \tag{9}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)' \sim N(0, \sigma^2 I)$ and $(v_1(i), \ldots, v_d(i))$ is the d-dimensional covariate information related to observation y_i. The penalty functional is given by:

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

$$J_m^d(f) = \sum_{\alpha_1+\cdots+\alpha_d=m} \frac{m!}{\alpha_1!\ldots\alpha_d!} \times \int\cdots\int \left(\frac{\partial^m f}{\partial v_1 \ldots \partial v_d}\right)^2 dv_1 \ldots dv_d. \tag{10}$$

A thin-plate smoothing spline is the solution to the following variational problem. Find $f \in \chi$ to minimise:

$$\sum_{i=1}^{n} (y_i - f(v_1(i), \ldots, v_d(i)))^2 + \lambda J_m^d(f). \tag{11}$$

Specifically, for $m = 2$, $d = 2$:

$$J_2^2(f) = \int\int \left[\left(\frac{\partial^2 f}{\partial v_1^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial v_1 \partial v_2}\right)^2 + \left(\frac{\partial^2 f}{\partial v_2^2}\right)^2\right] dv_1 dv_2 \tag{12}$$

and, in this special case, the solution of (11) will be given by a two-dimensional thin-plate spline. Arguing exactly as before, one can establish the underlying connection between the Bayesian approach to the estimation of f under a multivariate Gaussian prior and thin-plate splines. It is important to note that implementation of the procedure requires an explicit computation of the penalty matrix $\mathbf{K}$. However, this issue has been addressed by Wahba (2000), and we will be using this specific formulation in our computation.

3.2 *Application to Credibility Modelling*

Traditionally, linear models have played an important role in loss models, and, in particular, in credibility models. Given the observed data $\mathbf{y}$, one suggestion is to approximate $\mu_{T+1}(\theta) \equiv \mathrm{E}(Y_{T+1} \mid \boldsymbol{\Theta} = \theta)$ by a linear function of past data. The idea is to restrict the solution to estimators of the form $\alpha_0 + \sum_{t=1}^{T} \alpha_t Y_t$, where $\alpha_0, \alpha_1, \ldots, \alpha_T$ need to be chosen. We will choose the α_i's to minimise:

$$Q = \mathrm{E}\left\{[\mu_{T+1}(\theta) - \alpha_0 - \sum_{t=1}^{T} \alpha_t Y_t]^2\right\}.$$

The resulting credibility premium formula is given by $[Z \times \overline{Y} + (1 - Z) \times \mu]$, where $\overline{Y}$ represents the individual sample average, μ represents the overall population mean, and Z is called the credibility factor. Thus, the linear model approach provides an estimator which is a weighted average of the individual's past experience and the overall population experience. One major criticism of this approach (and hence of credibility as a premium setting tool) is that it tends to give lower premiums to bad risks, and higher premiums to good risks. It is definitely desirable to avoid such outcomes.

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

This motivates the approach which we developed in the present work. This nonlinear model-based technique provides a more intricate understanding of future losses. The loss model which we propose is significantly more general than models currently available in the literature. The proposed model borrows strength from similar risks, and provides a more precise prediction of future losses, along with estimates of the associated prediction error.

To determine how the current claim loss y_{ij} depends on the past losses, say $y_{i,j-1}$ and $y_{i,j-2}$, our approach is to consider a nonparametric regression model:

$$y_{ij} = g(y_{i,j-1}, y_{i,j-2}) + \epsilon_i \quad i = 1, \ldots, n$$

where g is a smooth function of its arguments. What we wish to accomplish is a description of the dependency between the current observations y_{ij} for all policyholders $i = 1, \ldots, n$ and past losses $y_{i,j-1}$ and $y_{i,j-2}$ at occasion j, through the nonparametric regression function g, which is assumed to be common to all risks in a portfolio of risks. Once the model is identified (estimated), we can perform a one-step ahead prediction of y_{i,T_i+1} by using y_{iT_i} and y_{i,T_i-1} as covariates. For notational convenience, let y_i stand for the dependent variable (current observation) and s_i and t_i (or, equivalently, covariates v_1 and v_2 in (9)) for covariates (two consecutive past losses). We develop a methodology to estimate the function g using a Bayesian nonparametric approach, which assigns a Gaussian prior to the unknown function g. As discussed in Section 3.1, the choice of the covariance matrix of the Gaussian prior must be very specific. In particular, the covariance is generated as the reproducing kernel of a thin-plate spline (Wahba, 1990, p30; Wahba, 2000).

3.3 *Model and Prior*

Without loss of generality, we assume that s_i and t_i lie in the interval [0, 1]. Consider the bivariate regression model:

$$y_i = g(s_i, t_i) + \epsilon_i \quad i = 1, \ldots, n \tag{13}$$

where g is a smooth regression of s and t, and errors ϵ_i are independent $N(0, \sigma^2)$. For convenience, we write (13) as:

$$y_i = \alpha_0 + \alpha_1 s_i + \alpha_2 t_i + f(s_i, t_i) + \epsilon_i \quad i = 1, \ldots, n \tag{14}$$

with f having zero initial conditions:

$$f(0, 0) = 0 \quad \frac{\partial f}{\partial s}(0, 0) = 0 \quad \frac{\partial f}{\partial t}(0, 0) = 0. \tag{15}$$

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

The prior for $f(s,t)$ is a reproducing kernel for the thin-plate spline. This means that $f(s,t)$ are zero-mean Gaussian random variables with the covariance function given by:

$$\text{cov}\{f(s_i,t_i),f(s_j,t_j)\} = \tau^2\mathbf{K}\{(s_i,t_i),(s_j,t_j)\}. \tag{16}$$

The exact form of the positive definite matrix $\mathbf{K}$ is given by Wahba (1990), and is stated in the Appendix.

3.4 *Model Implementation*

In this section we discuss the implementation of the model in (14). Our approach follows closely the work of Wood *et al.* (2002), where the basic computational tools were developed in a model selection context. To make this model computationally feasible, we consider factoring the covariance matrix $\mathbf{K}$ as $\mathbf{QDQ}'$, where $\mathbf{Q}$ is an orthonormal matrix and $\mathbf{D}$ is a diagonal matrix with diagonal elements d_i, which are the eigenvalues of $\mathbf{K}$.

To simplify the notation, we rewrite the model in (14) as:

$$y_i = \alpha_0 + \alpha_1 s_i + \alpha_2 t_i + f_i + \epsilon_i \quad i = 1,\ldots,n \tag{17}$$

where $f_i = f(s_i,t_i)$, and $\mathbf{f} = (f_1,\ldots,f_n)'$ is multivariate Gaussian with zero mean and covariance $\tau^2\mathbf{K}$.

The model (14) can be written in matrix form as:

$$\mathbf{y} = \mathbf{Z}\alpha + \mathbf{f} + \varepsilon \tag{18}$$

where:

$$\begin{aligned} \alpha &= (\alpha_0,\alpha_1,\alpha_2)' \\ \mathbf{y} &= (y_1,\ldots,y_n)' \\ \epsilon &= (\epsilon_1,\ldots,\epsilon_n)' \end{aligned} \tag{19}$$

and

$$\mathbf{Z} = \begin{bmatrix} 1 & s_1 & t_1 \\ 1 & s_2 & t_2 \\ \vdots & \vdots & \vdots \\ 1 & s_n & t_n \end{bmatrix}. \tag{20}$$

Thus, the prior specification is completed by assigning the following priors to all unknown components of the model as follows:

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

$$\begin{aligned} \alpha &\sim \mathbf{N}(0, c\mathbf{I}) \\ \mathbf{f} &\sim \mathbf{N}(0, \tau^2\mathbf{K}) \\ \epsilon &\sim \mathbf{N}(0, \sigma^2\mathbf{I}) \\ \tau^2 &\sim unif[0, 10^{10}] \\ \sigma^2 &\sim IG(10^{-8}, 10^{-10}) \end{aligned} \tag{21}$$

where τ^2 is uniformly distributed on the interval $[0, 10^{10}]$ and σ^2 has an inverse Gamma distribution with parameters 10^{-8} and 10^{-10}. It should be noted that, for computational simplicity, these priors are chosen as the conjugate to the family of distribution; thus the posterior distribution belongs to the same family as the prior, and the resulting parameters of the posterior can be readily obtained.

Since $\mathbf{K}$ is positive definite, we can factor $\mathbf{K}$ as $\mathbf{QDQ}'$ such that:

$$\mathbf{QQ}' = \mathbf{I}. \tag{22}$$

If we pre-multiply equations (18) by $\mathbf{Q}'$, then, with $\mathbf{y}^* = \mathbf{Q}'\mathbf{y}$, the model becomes:

$$\mathbf{y}^* = \mathbf{Z}^*\alpha + \mathbf{f}^* + \epsilon^* \tag{23}$$

where:

$$\begin{aligned} \mathbf{Z}^* &= \mathbf{Q}'\mathbf{Z} \\ \mathbf{f}^* &= \mathbf{Q}'\mathbf{f} \\ \epsilon^* &= \mathbf{Q}'\epsilon. \end{aligned} \tag{24}$$

The priors for α, τ^2, σ^2 will remain the same as in (21). Meanwhile, ϵ^* has the same distribution as $\epsilon \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$ because of (22). However, the prior distribution of $\mathbf{f}^*$ becomes:

$$\mathbf{f}^* \sim \mathbf{N}(\mathbf{0}, \tau^2\mathbf{D}) \tag{25}$$

because:

$$\begin{aligned} \mathrm{Var}(\mathbf{Q}'\mathbf{f}) &= \mathbf{Q}'\mathrm{Var}(\mathbf{f})\mathbf{Q} \\ &= \mathbf{Q}'\tau^2\mathbf{KQ} \\ &= \tau^2\mathbf{Q}'\mathbf{QDQ}'\mathbf{Q} \\ &= \tau^2\mathbf{D} \end{aligned}$$

where τ^2 and $\mathbf{D}$ are as defined before.

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

3.5 *Modification for the Bühlmann-Straub Model*

Consider the Bühlmann-Straub (Bühlmann & Straub, 1970) model, which allows different numbers of exposure units or different distributions of claim size across past policy years. This can be handled in model (17) by assuming:

$$\epsilon_i \sim N\left(0, \frac{\sigma^2}{w_i}\right) \quad i = 1, \ldots, n \tag{26}$$

where w_i is the corresponding weight for data value y_i. We can also have the same matrix form as (18)

$$\mathbf{y} = \mathbf{Z}\alpha + \mathbf{f} + \epsilon \tag{27}$$

but with priors:

$$\begin{aligned} \alpha &\sim \mathbf{N}(\mathbf{0}, c\mathbf{I}) \\ \mathbf{f} &\sim \mathbf{N}(\mathbf{0}, \tau^2\mathbf{K}) \\ \epsilon &\sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{W}^{-1}) \\ \tau^2 &\sim unif[0, 10^{10}] \\ \sigma^2 &\sim IG(10^{-8}, 10^{-10}). \end{aligned} \tag{28}$$

Here $\mathbf{W}$ is a diagonal matrix with diagonal elements w_i, the corresponding weight for data values y_i.

This model can be transformed easily into a similar model as in (18) and (21), and can be implemented analogously as in Section 3.4. Let:

$$\begin{aligned} \mathbf{y}' &= \sqrt{\mathbf{W}}\mathbf{y} \\ \mathbf{Z}' &= \sqrt{\mathbf{W}}\mathbf{Z} \\ \mathbf{f}' &= \sqrt{\mathbf{W}}\mathbf{f} \\ \epsilon' &= \sqrt{\mathbf{W}}\epsilon. \end{aligned} \tag{29}$$

An interim model is given by:

$$\mathbf{y}' = \mathbf{Z}'\alpha + \mathbf{f}' + \epsilon' \tag{30}$$

with priors:

$$\begin{aligned} \alpha &\sim \mathbf{N}(\mathbf{0}, c\mathbf{I}) \\ \mathbf{f}' &\sim \mathbf{N}(\mathbf{0}, \tau^2\sqrt{\mathbf{W}}\mathbf{K}\sqrt{\mathbf{W}}) \\ \epsilon' &\sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I}) \\ \tau^2 &\sim unif[0, 10^{10}] \\ \sigma^2 &\sim IG(10^{-8}, 10^{-10}). \end{aligned} \tag{31}$$

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

With $\mathbf{K}' = \sqrt{\mathbf{W}}\mathbf{K}\sqrt{\mathbf{W}}$, we have $f'(s,t)$ distributed as zero-mean Gaussian random variables, with the covariance function given by:

$$\text{cov}\{f'(s_i,t_i), f'(s_j,t_j)\} = \mathbf{K}'\{(s_i,t_i),(s_j,t_j)\}.$$

This is simply a different choice of prior for the full bivariate surface $f'(s,t)$. We can then use results from (17) to (25) based on (30) and (31).

We use a Gibbs sampling scheme, where the bivariate regression surface is modelled by a thin-plate spline prior, as described earlier in the paper. A good introduction to the Gibbs sampler is given by Gelfand & Smith (1990). One of the advantages of Gibbs sampling is that it can take advantage of any additive structure in the model, as explained in Wong & Kohn (1996).

3.6 *Implementation of Sampling Scheme*

We present a Gibbs sampling scheme where the bivariate regression surface is modelled by a thin-plate spline prior. The basic algorithm was developed by Wood *et al.* (2002). The estimates of α, τ^2, and σ^2 are obtained by generating iterations $\alpha^{[j]}$, $\tau^{2[j]}$, and $\sigma^{2[j]}$ from this sampling scheme using noninformative priors for α, τ^2 and σ^2, described in (21) with $c = 10^{20}$. We use the following sequential sampling scheme to implement our model.

(1) Initialise τ^2 and λ as $\tau^{2[0]}$ and $\lambda^{[0]}$ respectively. Then set $\sigma^{2[0]} = \lambda^{[0]} \times \tau^{2[0]}$.
(2) Let $\mathbf{\Psi} = \text{var}(\mathbf{f}^* + \epsilon^*) = \tau^2\mathbf{D} + \sigma^2\mathbf{I}$, so that $\mathbf{\Psi}$ is a diagonal matrix. Then the model is simply a linear model with known covariance matrix $\mathbf{\Psi}$. Therefore, generate α from the conditional pdf $p(\alpha \mid \mathbf{y}^*, \tau^2, \sigma^2)$, which is a Gaussian distribution with mean $(\mathbf{Z}^{*\prime}\mathbf{\Psi}^{-1}\mathbf{Z}^*)^{-1}\mathbf{Z}^{*\prime}\mathbf{\Psi}^{-1}\mathbf{y}^*$ and covariance $(\mathbf{Z}^{*\prime}\mathbf{\Psi}^{-1}\mathbf{Z}^*)^{-1}$.
(3) Let $\tilde{\mathbf{y}}^* = \mathbf{y}^* - \mathbf{Z}^*\alpha$. Then generate $\mathbf{f}^*$ from the conditional pdf $p(\mathbf{f}^* \mid \tilde{\mathbf{y}}^*, \tau^2, \sigma^2)$, which is again a Gaussian distribution, with mean $(\tau^2/\sigma)^2\mathbf{D}[\mathbf{I} + (\tau^2/\sigma^2)\mathbf{D}]^{-1}\tilde{\mathbf{y}}^*$ and covariance:

$$\sigma^2\left(\frac{\tau^2}{\sigma^2}\mathbf{D} - \frac{\tau^2}{\sigma^2}\mathbf{D}\left(\mathbf{I} + \frac{\tau^2}{\sigma^2}\mathbf{D}\right)^{-1}\frac{\tau^2}{\sigma^2}\mathbf{D}\right).$$

(4) Generate τ^2 from the conditional pdf $p(\tau^2 \mid \mathbf{y}^*, \alpha, \mathbf{f}^*)$, which is an inverse gamma distribution with parameters $(n-3)/2$ and $\sum_{i=1}^{n}(f_i^{*2}/d_i)$.
(5) Let $\epsilon^* = \mathbf{y}^* - \mathbf{Z}^*\alpha - \mathbf{f}^*$. Generate σ^2 from the conditional pdf $p(\sigma^2 \mid \mathbf{y}^*, \alpha, \mathbf{f}^*)$, which is again an inverse gamma with parameters $[(n/2) - 1 + 10^{-8}]$ and $\left[\left(\sum_{i=1}^{n}\epsilon_i^{*2}/2\right) + 10^{-10}\right]$.

In the sampling scheme, step 2 to step 5 are repeated. Gibbs sampling is carried out in two stages. The first stage is a burn-in period. At the end of this stage, it is assumed that the iterations have converged to draws from the

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

posterior distribution. The second stage, called the sampling stage, is used to estimate the posterior means.

3.7 *One-Step Ahead Prediction*

Many problems in actuarial science involve the building of a mathematical model which can be used to predict insurance costs, or to forecast losses in the future, particularly the short-term future. Suppose that we are interested in a one-step ahead prediction of $Y_{i,T+1}$. We take the posterior mean $\mathrm{E}(Y_{i,T+1} \mid \mathbf{y})$ as the best predictor of $Y_{i,j+1}$, and use the posterior variance $\mathrm{var}(Y_{i,T+1} \mid \mathbf{y})$ to estimate the associated uncertainty of the prediction. We estimate the posterior mean and variance of $Y_{i,j+1}$ using empirical estimates based on the values of $y_{i,j+1}$ generated during the sampling period by using the model (17); that is:

$$y_{ij} = \alpha_0 + \alpha_1 y_{i,j-1} + \alpha_2 y_{i,j-2} + f_i + \epsilon_i \quad i = 1, \ldots, n.$$

To generate $y_{i,j+1}$, we plug in y_{ij} and $y_{i,j-1}$ as covariates. For each iteration, with the generated values of α and $\mathbf{f}$ from the sampling scheme, we have:

$$y_{i,j+1} = \alpha_0 + \alpha_1 y_{i,j} + \alpha_2 y_{i,j-1} + f_i + \epsilon_i \quad i = 1, \ldots, n.$$

Therefore, for each j, the one-step ahead prediction is given by

$$\hat{y}_{i,j+1} = \hat{\alpha}_0 + \hat{\alpha}_1 y_{i,j} + \hat{\alpha}_2 y_{i,j-1} + \hat{f}_i.$$

3.8 *Implementation in Higher Dimensions*

The model can be implemented easily in a higher dimensional setup. For example, one can impose an additive structure to our proposed model, and repeat the sampling scheme via a backfitting algorithm, as described by Hastie & Tibshirani (1990). Some illustrations of this idea can also be found in Wood (2003). In this section, we illustrate how to handle functions involving three covariates without resorting to the assumption of additivity. As before, we treat s_i and t_i as past losses, and incorporate additional relevant covariate information u_i (or, equivalently, covariates v_1, v_2 and v_3 in (9)). For example, u_i can be the number of years for which a policyholder remains in the same policy with the same insurer, or represents a driving age group in auto insurance. Then the regression model is given by:

$$y_i = g(s_i, t_i, u_i) + \epsilon_i \quad i = 1, \ldots, n$$

with ϵ_i independent $\mathrm{N}(0, \sigma^2)$ and $g(\cdot)$ of the form:

$$y_i = \alpha_0 + \alpha_1 s_i + \alpha_2 t_i + \alpha_3 u_i + f(s_i, t_i, u_i) + \epsilon_i \quad i = 1, \ldots, n.$$

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

The prior on f is specified similarly to (16), that is:

$$\text{cov}\{f(s_i, t_i, u_i), f(s_j, t_j, u_j)\} = \tau^2 \mathbf{K}\{(s_i, t_i, u_i), (s_j, t_j, u_j)\}. \tag{32}$$

As before, the form of the kernel $\mathbf{K}$ is given in the Appendix.

4. Examples

In this section we discuss the implementation of our model in data analysis. Note that the model incorporates past and present observations related to risk, as well as relevant covariate information. As stated earlier, all computations are carried out using a Gibbs sampling scheme. The model uses a MCMC technique to compute the predictive distribution of future claims based on available data. With a burn-in period, it is assumed that iterations have converged to draws from posterior distributions. A large number of such draws can be used to identify the posterior distributions of all unknown components of the model; and, in turn, can be used for the purpose of short-term prediction. In the next sections, we discuss a simulated example, and an application of the model in describing fire related losses in Massachusetts counties. The following example illustrates the superiority of the proposed technique when compared against existing methods.

4.1 *Simulation Results*

In this example we illustrate the accuracy and performance of a Bayesian nonparametric (BN) technique, proposed in this paper, against a standard Bühlmann-Straub (BS) credibility model. In this simulation study, four different regression functions are considered, along with zero mean Gaussian errors with variances $(1/8)$, $(1/4)$ and $(1/2)$:

— *Model a.* $y = 7[(s-0.5)^2 + (t-0.5)^2] \times [0.75 - (s-0.5)^2 - (t-0.5)^2]$.
— *Model b.* $y = \sin(4\pi s)$.
— *Model c.* $y = 0.5$.
— *Model d.* $y = 0.5 + 0.5s + 0.25t$.

Note that two of the regression models are linear, while the other two are nonlinear. Two covariates s and t are generated as independent *uniform*$(0, 1)$. One hundred replicates are created with each regression model and for each error variance, with one exposure per cell; and the model is estimated from the data using both BN and BS techniques. To examine the performances of the competing estimation methods, the average squared error (ASE) was calculated for the estimates of the regression functions, where ASE is given by:

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

$$ASE = \frac{1}{n}\sum_{i=1}^{n}(\hat{g}(s_i, t_i) - g(s_i, t_i))^2. \tag{33}$$

Table 1 shows the ASE for BN and BS estimates of each data generating model. It is easy to see the superiority of the BN method compared to the BS approach. Even in the cases where the regression model is actually linear,

Table 1. Bivariate regression; the ASE is calculated using (33); 100 replications were generated for each combination of the function and the noise

Model	σ^2					
	1/8		1/4		1/2	
	BN	BS	BN	BS	BN	BS
(a)	0.00195	0.11716	0.00719	0.18087	0.02896	0.40740
(b)	0.00042	0.75619	0.00461	0.80690	0.02922	1.00744
(c)	0.00134	0.02004	0.00586	0.05307	0.02160	0.18875
(d)	0.00130	0.14913	0.00586	0.18006	0.02179	0.30963

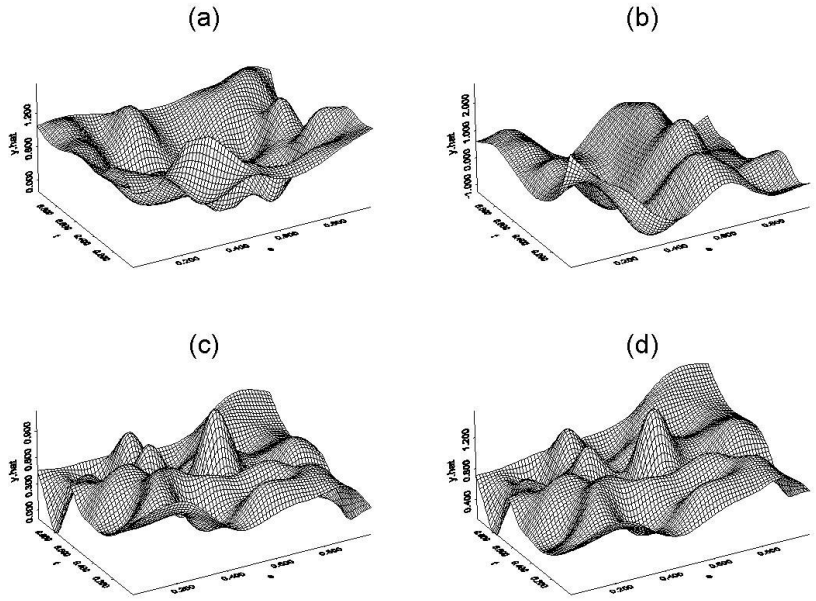


Figure 1. Plots of fitted surfaces using the Bayesian nonparametric method for the four regression models with the noise variance 1/4

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

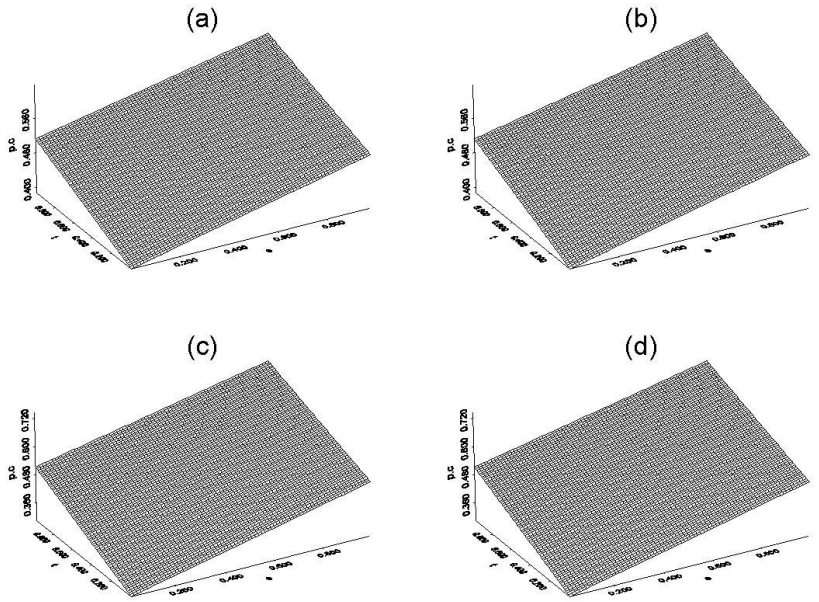


Figure 2. Plots of fitted surfaces using the Bühlmann-Straub method for the four regression models with the noise variance 1/4

the BN approach still performs at a superior level, due to the fact that the BN model makes more efficient use of the information in the error distribution.

Figure 1 and Figure 2 show the fitted surfaces using the Bayesian regression model and the Bühlmann-Straub model for the simulated data, with error variance $\sigma^2 = 1/4$. These figures show the flexibility of the proposed regression model when compared against the linear BS model.

4.2 *Application to Massachusetts Fire Data*

In this example, we analyse data collected from the Massachusetts Fire Incident Reporting System. By law, fire departments must report any fire or explosion resulting in a dollar loss or human casualty to the State Fire Marshal's Office. These annual aggregate data are reported by the Commonwealth of Massachusetts. For example, in the year 2000, a total of 24,391 fires were reported in Massachusetts; resulting in 78 civilian deaths, 619 civilian injuries, 837 fire service injuries, and approximately 189 million dollars in property damage. For this study, we chose four counties in eastern Massachusetts; namely, Essex County, Middlesex County, Norfolk County,

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

and Plymouth County. All four counties are located close to the City of Boston. In the context of loss ratios from insurance portfolios, the fire ratio (number of fires per 100 population) of each town is of interest.

Essex County is ranked sixth out of 14 Massachusetts counties in total reported fires. In the year 2000, Essex County reported 2,168 fires to the Massachusetts Fire Incident Reporting System, including 848 structural fires, 658 vehicle fires and 662 other fires. Middlesex County ranked third out of 14 Massachusetts counties in total reported fires. In the year 2000, the county reported 3,007 fires, including 1,695 structural fires, 838 vehicle fires, and 474 other fires. Norfolk County ranked seventh out of 14 Massachusetts counties in total reported fires. The county reported 1,769 fires, including 628 structural fires, 430 vehicle fires and 711 other fires. Plymouth County ranked eighth out of 14 Massachusetts counties in total reported fires. The county reported 1,721 fires, including 673 structural fires, 424 vehicle fires and 624 other fires. The fire ratios are 0.30, 0.21, 0.27 and 0.36 per 100 population in Essex County, Middlesex County, Norfolk County and Plymouth County respectively.

Our interest is in the fire ratios for each town of these four counties. Based on available data, we used the proposed Bayesian nonparametric regression model to predict the future fire ratio in each of these towns in our data set. The analysis is expected to provide valuable information on many insurance policies, such as auto insurance and structure fire insurance. Table 2 shows a small part of the data layout. It is reasonable to assume that the population remained unchanged during this short time period. In our analysis, we included 111 towns, each with populations of more than 7,000. The fire ratio, in the units of per 100 population, associated with town j $(= 1, \cdots, 111)$ in year t $(= 1996, 1997, 1998, 1999, 2000)$, is represented by the random variable Y_{jt}.

From Figure 3, it appears that there is a strong association between the current observation and observations from the previous three years. Some departure from the standard linear model is also evident in these plots. We

Table 2. Fire ratio per 100 population from 1996 to 2000

	Town	1996	1997	1998	1999	2000
Fire ratio	Amesbury	0.49	0.54	0.50	0.43	0.06
Population	16,450	16,450	16,450	16,450	16,450	16,450
Fire ratio	Andover	0.51	0.56	0.47	0.58	0.20
Population	31,247	31,247	31,247	31,247	31,247	31,247
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Fire ratio	Whitman	0.43	0.53	0.63	0.48	0.43
Population	13,882	13,882	13,882	13,882	13,882	13,882

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

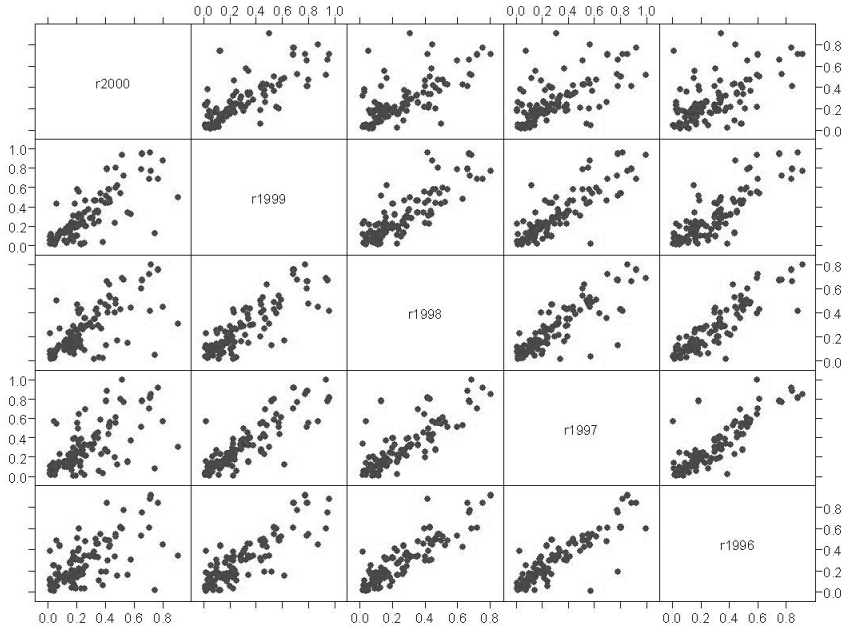


Figure 3. Scatter matrix plot; plots of fire ratios from years 1996 to 2000

consider a bivariate regression model of the fire ratio in 1999 on the fire ratios in 1998 and 1997. In addition, we also fit a trivariate regression model of the fire ratio of 1999 on the fire ratios in 1998, 1997 and 1996. To examine the relative performances of these two proposed models, the averaged squared error (ASE) was calculated for estimates of the regression functions. The ASE is computed as:

$$\text{ASE} = \frac{1}{\sum_{j=1}^{n} w_j} \sum_{j=1}^{n} w_j (Y_{j,1999} - \widehat{Y}_{j,1999})^2 \tag{34}$$

where w_j is the size of the population in town j in year 1999, and $\widehat{Y}_{j,1999}$ is the fitted value of the fire ratio for town j in 1999. The ASE of the bivariate model is 0.005, while that of the trivariate model is equal to 0.00047. In addition, we use data in year 2000 to validate the predictive performance of the proposed model. The average squared prediction error (ASPE) was calculated for year 2000 predictions using the proposed model, where the ASPE is given by:

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

$$\text{ASPE} = \frac{1}{\sum_{j=1}^{n} w_j} \sum_{j=1}^{n} w_j (Y_{j,2000} - \hat{Y}_{j,2000})^2. \tag{35}$$

The results show that the ASPE is reduced by about 32% from the bivariate model (ASPE = 0.246) to the trivariate model (ASPE = 0.167). This indicates that the trivariate model has a better predictive performance than the bivariate model. Therefore, we use the trivariate regression model to predict loss ratios associated with year 2001 for various Massachusetts towns. It should be noted that, in datasets with a larger number of observations within each exposure unit (compared to only five years of observations which we have here), it may be useful to consider a higher-order additive model to improve the predictive power. In this particular case, however, the small magnitudes of ASE and ASPE suggest that a higher-order model is not needed for this dataset.

Table 3 gives fire ratio predictions for year 2001 for a sample of towns in Massachusetts counties. As a matter of fact, this procedure provides the entire predictive distribution of $Y_{j,2001}$. The predictive distributions for the year 2001 for a sample of towns are shown in Figure 4. We can easily construct credible intervals for predictions based on these distributions. Moreover, we can also provide answers to certain specific questions related to the predictions which are deemed important from an insurance stand point. For example, for each town, suppose that we wish to estimate the probability of a fire ratio more than 0.3 in the year 2001; a threshold which can be considered as an indicator of high casualty loss. The last column of Table 3 provides estimates of the probability for a sample of Massachusetts towns. With this information, we expect that higher premiums would be charged in towns with a higher probability of getting relatively high fire ratios.

Table 3. Predicted fire ratio and probability in year 2001

Town	Population	$Y_{j,2001}$	$P(Y_{j,2001} > 0.3)$
Belmont	24,194	0.27	0.385
Brockton	94,304	0.30	0.476
Brookline	57,107	0.04	0.000
Cambridge	101,355	0.37	0.913
Framingham	66,910	0.13	0.001
Lexington	30,355	0.12	0.001
Lowell	105,167	0.21	0.000
Newton	83,829	0.24	0.007
Plymouth	51,701	0.44	0.937
Sudbury	16,841	0.07	0.000
Waltham	59,226	0.18	0.009
Wellesley	26,613	0.38	0.906

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

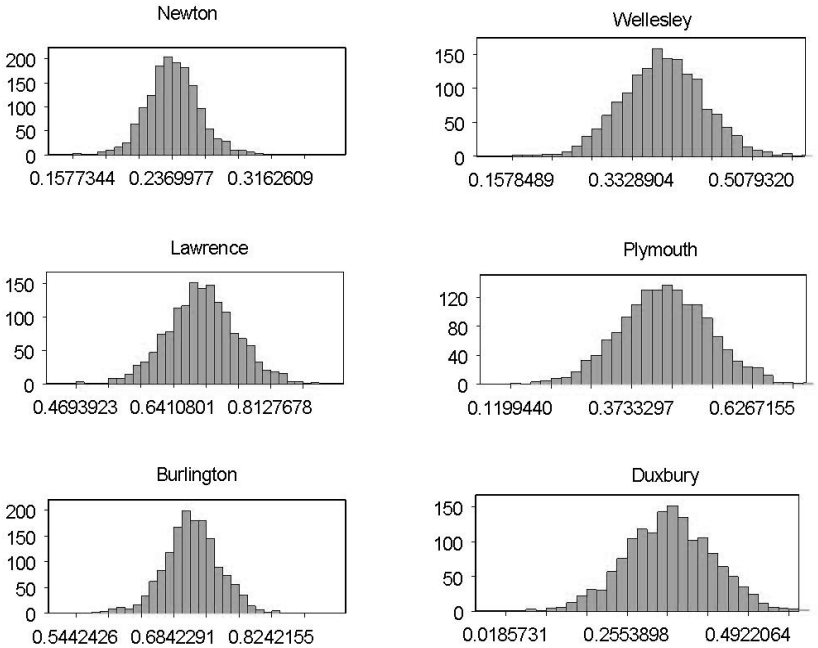


Figure 4. Predicted fire ratio in year 2001

5. Conclusions

Many problems in actuarial science involve the building of mathematical models which can be used to predict the insurance cost in the future, particularly the short-term future. A Bayesian nonparametric approach is proposed to the problem of risk modelling. The model incorporates past and present observations related to risk, as well as relevant covariate information; and uses a MCMC technique to compute the predictive distribution of the future claims based on available data, where the covariance structure of a Gaussian prior is based on a thin-plate spline (Wahba, 1990).

We have illustrated applications of Gibbs sampling within the context of nonparametric regression and smoothing. Gibbs sampling provides a feasible approach to the computation of posterior distributions. Combined with the assumed thin-plate spline structure of the regression surface and the computational availability of bivariate or trivariate surface estimation, this methodology opens up a new dimension to credibility literature. Although our discussion concentrates primarily on two and three-dimensional

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

applications, the technique can be extended easily to higher dimensional problems. Our investigations show that this method performs at a superior level compared to the existing linear model-based techniques commonly used in modelling credibility.

Acknowledgements

The authors gratefully acknowledge the two referees for their extensive comments and suggestions which led to a significant improvement of the paper, and would like to thank the editor for her help throughout the review process.

References

Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin*, **4**, 199-207.

Bühlmann, H. & Straub, E. (1970). Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker*, **70**, 111-133.

Duchon, J. (1975). Functions splines et vecteurs aleatoires. Technical Report 213, Seminaire d'Analyse Numeriaue.

Duchon, J. (1976). Functions-spline et esperances conditionnelles de champs Gaussians *Annal Science University Clermont Ferrand II Mathematics*, **14**, 19-27.

Duchon, J. (1977). *Spline minimising rotation-invariant semi-norms in Sobolev spaces in constructive theory of functions of several variables.* Springer-Verlag, Berlin.

Gangopadhyay, A. & Gau, W. (2003). Credibility modeling via spline nonparametric regression. *CAS Forum*, Winter, 215-252.

Gelfand, A.E. & Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.

Hastie, T.J. & Tibshirani, R.J. (1990). *Generalized additive models*. Chapman & Hall.

Jewell, W.S. (1974). Credible means are exact Bayesian for exponential families. *ASTIN Bulletin*, **8**, 336-341.

Klugman, S.A. (1992). *Bayesian statistics in actuarial science with emphasis on credibility*. Kluwer, Boston.

Mayerson, A.L. (1964). A Bayesian view of credibility. *Proceedings of the Casualty Actuarial Society*, **51**, 85-104.

Meinguet, J. (1979). Multivariate interpolation at arbitrary points made simple. *Journal of Applied Mathematical Physics (ZAMP)*, **30**, 292-304.

Scollnik, D.P.M. (2000). Actuarial modeling with MCMC and BUGS. Actuarial Education and Research.

Spiegelhalter, D.J., Thomas, A. & Best, N.G. (1996). *BUGS 0.5:Bayesian inference using Gibbs sampling manual (version II)*. MRC Biostatistics Unit, Cambridge.

Wahba, G. & Wendelberger, J. (1980). Some new mathematical methods for variational objective analysis using splines and cross-validation. *Monthly Weather Review*, **108**, 1122-1145.

Wahba, G. (1990). *Spline models for observational data*. SIAM, Philadelphia, 69.

Wahba, G. (2000). An introduction to model building with reproducing kernel Hilbert space. Technical report No. 1020, Department of Statistics, University of Wisconsin, Madison, WI.

Wong, C. & Kohn, R. (1996). A Bayesian approach to additive semiparametric regression. *Journal of Econometrics*, **7**, 209-223.

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

WOOD, S., KOHN, R., SHIVELY, T. & JIANG, W. (2002). Model selection in spline nonparametric regression. *Journal of the Royal Statistical Society B*, **64**, 119-139.

WOOD, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society B*, **65**, Part 1, 95-114.

YOUNG, V.R. (1997). Credibility using a loss function from spline theory. *Scandinavian Actuarial Journal*, **2**, 160-185.

YOUNG, V.R. (2000). Robust Bayesian credibility using semiparametric models. Preprint.

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

APPENDIX

Wahba (1990) identified a general formulation for computing kernel K related to a d-dimensional thin-plate spline. Here we provide the specific forms of the kernel matrix in two and three dimensions.

For a two-dimensional function $f(s, t)$, with a covariance structure as described in (16), the kernel $\mathbf{K}$ is given by:

$$\begin{aligned}\mathbf{K}\{(s_i, t_i), (s_j, t_j)\} &= \mathrm{E}\{(s_i, t_i), (s_j, t_j)\} \\ &\quad - \sum_{k=1}^{3} p_k(s_j, t_j)\mathrm{E}\{u_k, (s_i, t_i)\} \\ &\quad - \sum_{k=1}^{3} p_k(s_i, t_i)\mathrm{E}\{(s_j, t_j), u_k\} \\ &\quad + \sum_{k=1}^{3}\sum_{l=1}^{3} p_k(s_i, t_i)p_l(s_j, t_j)\mathrm{E}\{u_k, u_l\}\end{aligned} \tag{36}$$

where:

$$\mathrm{E}\{(s_i, t_i), (s_j, t_j)\} = r^2 \log(r) \quad r = \sqrt{(s_i - s_j)^2 + (t_i - t_j)^2} \tag{37}$$

$$p_1(s_i, t_i) = -1 + 2s_i + 2t_i \quad p_2(s_i, t_i) = 1 - 2s_i \quad p_3(s_i, t_i) = 1 - 2t_i \tag{38}$$

and

$$u_1 = \left(\frac{1}{2}, \frac{1}{2}\right) \quad u_2 = \left(0, \frac{1}{2}\right) \quad u_3 = \left(\frac{1}{2}, 0\right).$$

Similarly for a three-dimensional function $f(s, t, v)$, with the covariance structure (32), the kernel $\mathbf{K}$ is given by:

$$\begin{aligned}\mathbf{K}\{(s_i, t_i, v_i), (s_j, t_j, v_j)\} &= \mathrm{E}\{(s_i, t_i, v_i), (s_j, t_j, v_j)\} \\ &\quad - \sum_{k=1}^{4} p_k(s_j, t_j, v_j)\mathrm{E}\{u_k, (s_i, t_i, v_i)\} \\ &\quad - \sum_{k=1}^{4} p_k(s_i, t_i, v_i)\mathrm{E}\{(s_j, t_j, v_j), u_k\} \\ &\quad + \sum_{k=1}^{4}\sum_{l=1}^{4} p_k(s_i, t_i, v_i)p_l(s_j, t_j, v_j)\mathrm{E}\{u_k, u_l\}\end{aligned} \tag{39}$$

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press

where:

$$\begin{aligned} \mathrm{E}\{(s_i, t_i, v_i), (s_j, t_j, v_j)\} &= \tau^2 \log(\tau) \\ \tau &= \sqrt{(s_i - s_j)^2 + (t_i - t_j)^2 + (v_i - v_j)^2} \end{aligned} \tag{40}$$

$$\begin{aligned} p_1(s_i, t_i, v_i) &= -1 + 2s_i + 2t_i + 2v_i \\ p_2(s_i, t_i, v_i) &= 1 - 2s_i \\ p_3(s_i, t_i, v_i) &= 1 - 2t_i \\ p_4(s_i, t_i, v_i) &= 1 - 2v_i \end{aligned} \tag{41}$$

and

$$u_1 = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right) \quad u_2 = \left(0, \frac{1}{2}, \frac{1}{2}\right) \quad u_3 = \left(\frac{1}{2}, 0, \frac{1}{2}\right) \quad u_4 = \left(\frac{1}{2}, \frac{1}{2}, 0\right).$$

https://doi.org/10.1017/S1748499500000270 Published online by Cambridge University Press