Credibility Modeling via Spline Nonparametric Regression

Ashis Gangopadhyay and Wu-Chyuan Gau

Credibility Modeling via Spline Nonparametric Regression

Dr. Ashis Gangopadhyay

Associate Professor,

Mathematics and Statistics,

Boston University,

111 Cummington Street,

MA 02215, U.S.A.

Phone: 617-353-2560.

Email: ag@math.bu.edu.

Dr. Wu-Chyuan Gau

Assistant Professor,

Statistics and Actuarial Science,

University of Central Florida,

Orlando, FL 32816,

U.S.A.

Phone: 407-823-5530.

Email: wgau@mail.ucf.edu.

Credibility Modeling via Spline Nonparametric Regression

Abstract

Credibility modeling is a rate making process which allows actuaries to adjust future premiums according to the past experience of a risk or group of risks. 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 that is a linear combination of current observations and past records. During the last decade, the existence of high speed computers and statistical software packages allowed the introduction of more sophisticated methodologies. Some of these techniques are based on Markov Chain Monte Carlo (MCMC) approach to Bayesian inference, which requires extensive computations. However, very few of these methods made use of the additional covariate information related to the risk, or group 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 modeling. The model incorporates past and present observations related to the risk, as well as relevant covariate information. The Bayesian modeling is carried out by sampling from a multivariate Gaussian prior, where the covariance structure is based on a thin-plate spline (Wahba, 1990). The model uses MCMC technique to compute the predictive distribution of the future claims based on the available data. Extensive data analysis is conducted to study the properties of the proposed estimator, and compare against the existing techniques.

Keywords: Credibility Modeling, Thin-plate Spline, MCMC, RKHS.

1 Introduction

The dictionary definition of a spline is "a thin strip of wood used in building construction." This in fact gives 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 points. A mathematical spline is the solution to a constrained optimization problem.

In the credibility context, suppose we wish 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 the nonparametric regression model

$$y_{ij} = g(y_{i,j-1}, y_{i,j-2}) + \epsilon_i, \qquad i = 1, ..., n, \tag{1}$$

where g is a smooth function of its arguments. Our objective is to model the dependency between the current observations y_{ij} for all policyholders $i = 1, ..., n$, and those past losses $Y_{i,j-1}$ and $Y_{i,j-2}$ through a nonparametric regression at occasion j. The concept is similar to multiple linear regressions. Here, the dependent variable happens to be Y_{ij} while the past losses $Y_{i,j-1}$ and $Y_{i,j-2}$ are treated as covariates. For notational convenience, we let Y_i stands for the dependent variable (current observation) and use s_i or t_i for covariates (past losses). The key problem is to find a good approximation $\hat{g}$ of g. This is a tractable problem, and there are many different solutions to this problem. The purpose of this paper is to develop a methodology to estimate the function g given the data. We use a nonparametric Bayesian approach to estimate the multivariate regression model with Gaussian errors. In this approach, very little is assumed regarding the underlying model (signal); and we allow the data to "speak for itself".

Reproducing kernel Hilbert space (RKHS) models have been in use for at least ninety-five years. The systematic development of reproducing kernel Hilbert space theory is given by Aronszajn (1950). For further background the reader may refer to Weinert (1982) and Wahba (1990). A recent paper given by Evgeniou (2000) contains an introduction to RKHS, which we found to be useful for readers interested in further reading.

A reproducing kernel Hilbert space is a Hilbert function space characterized by the fact that it contains a kernel that reproduces (through a inner product) every function in the space, or, equivalently, by the fact that every point evaluation functional is bounded. RKHS models are useful in estimation problems because every covariance function is also a reproducing kernel for some RKHS. As a consequence, there is a close connection between a random process and the RKHS determined by its covariance function. These estimation problems can then be solved by evaluating a certain RKHS inner product. Thus it is necessary to be able to determine the form of inner product corresponding to a given reproducing kernel.

In optimal curve and surface fitting problems, in which one is reconstructing an unknown function based on the sample data, it is inevitable that the point evaluation functionals be bounded. Therefore, one is forced to express the problem in a RKHS whose inner product is determined by the quadratic cost functional that needs to be minimized. To solve these problems, one must find a basis for the range of a certain projection operator. One way to do this is to determine the reproducing kernel corresponding to the given inner product.

Consider an univariate model

$$y_i = f(t_i) + \epsilon_i, \ i = 1, 2, ...n \tag{2}$$

where $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)' \sim N(0, \sigma^2 I)$ and f is only known to be smooth. If f has $m-1$

continuous derivatives and is m^{th} derivative is square integrable, an estimate of f can be found by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda \int_a^b (f^{(m)}(t))^2 dt, \tag{3}$$

for some $\lambda > 0$. The smoothing parameter λ controls the trade-off between smoothness and accuracy. A discrete version of problems such as (3) was considered in the actuarial literature by Whittaker (1923), who considered smoothing $y_1, ..., y_n$ discretely by finding $\mathbf{f} = (f_1, ..., f_n)$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f_i)^2 + \lambda \sum_{i=1}^{n-3}(f_{i+3} - 3f_{i+2} + 3f_{i+1} - f_i)^2. \tag{4}$$

If $m = 2$, (3) becomes the penalized residual sum of squares

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda \int_a^b (f''(t))^2 dx \tag{5}$$

where λ is a fixed constant, and $a \leqslant t_1 \leqslant \cdots \leqslant t_n \leqslant b$. If we consider all functions $f(t)$ with two continuous derivatives, it can be shown that (5) has an unique minimizer which is a nature cubic spline with knots at the unique values of t_i. As $\lambda \longrightarrow \infty$, it forces $f''(t) = 0$ everywhere, and the solution is the least-squares line. As $\lambda \longrightarrow 0$, the solution tends to an interpolating twice-differentiable function. The cubic spline can be generalized to two or higher dimensions. The thin-plate spline is one example. It derives from generalizing the second derivative penalty for smoothness to a two dimensional Laplacian penalty (Wendelberger, 1982).

This paper is organized as follows. Section 2 introduces the credibility problem. Section 3 reviews the thin-plate spline and the Bayesian model behind the smoothing spline. Section 4 introduces the basic idea of bivariate regression with Gaussian errors. Section 5 generalizes the results in section 4 to the trivariate case. Section 6 introduces the applications of the results developed in section 4 and section 5. All computations are carried out using Gibbs sampler. All functions, functionals, random variables, and function spaces in this paper will be real valued unless specifically noted otherwise.

2 The Credibility Problem

The classical data type in this area involves realizations from present and past experience of individual policyholders. Suppose we have n different risks (or policyholders) with a claims record over a certain number of years, say T;

$$Y_{11}, Y_{12}, \dots, Y_{1T}$$
$$Y_{21}, Y_{22}, \dots, Y_{2T}$$
$$\dots\dots\dots\dots\dots\dots$$
$$Y_{n1}, Y_{n2}, \dots, Y_{nT}.$$

The data can be the amount of losses, the number of claims, or the loss ratio from insurance portfolios. Our goal is to estimate the amount or 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 time and the past observed values of $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_T$, i.e., to establish the relationship:

$$y_{ij} = f(t, y_{i1}, y_{i2}, \dots, y_{i,j-1}) + e_{ij} \text{ for } i = 1, 2, \dots, n;\ j = 1, 2, \dots, T \tag{6}$$

where f is an unknown function and e_{ij} is a 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 i^{th} risk (or group of risks) were observed at time points $\{t_{ij} : j = 1, 2, ..., T_i\}$. In addition, let $\mathbf{X}_{ij}$ be the observed covariates, such as gender of a policyholder, or industry type of insureds etc., for the i^{th} risk (or group of risks) at time t_{ij}. In each risk (or group of risks), the data has the form

$$(y_{ij}, \mathbf{X}_{ij}, t_{ij}),\ j = 1, 2, ..., T_i;\ i = 1, 2, ..., n, \tag{7}$$

where $\mathbf{X}_{ij} = (X_{ij1}, X_{ij2}, ..., X_{ijd})$ are the d covariate variables measured at time t_{ij}. In this case, of interest is to study the association between the current response y_{ij} and the past responses $y^*_{i,j-1} = (y_{i,j-1}, y_{i,j-2}, ..., y_{i1})$ as well as the covariates and to examine how the association varies with time. Table 1 provides the data lay-out assuming $j = 1, 2, ..., T$.

	Occasion		
Subject	1	$\cdots$	T
1	$y_{11}, x_{11,1}, x_{11,2}, ..., x_{11,d}, t_{11}$	$\cdots$	$y_{1T}, x_{1T,1}, x_{1T,2}, ..., x_{1T,d}, t_{1T}$
$\vdots$	$\vdots$	$\cdots$	$\vdots$
i	$y_{i1}, x_{i1,1}, x_{i1,2}, ..., x_{i1,d}, t_{i1}$	$\cdots$	$y_{iT}, x_{i1,T}, x_{i1,T}, ..., x_{iT,d}, t_{iT}$
$\vdots$	$\vdots$	$\cdots$	$\vdots$
n	$y_{n1}, x_{n1,1}, x_{n1,2}, ..., x_{n1,d}, t_{n1}$	$\cdots$	$y_{nT}, x_{nT,1}, x_{nT,2}, ..., x_{nT,d}, t_{nT}$

Table 1: Data Configuration.

Therefore, we propose the following modified model, which is more general than (7),

$$y_{ij} = f(t_{ij}, \mathbf{X}_{ij,} y^*_{i,j-1}) + e_{ij} \text{ for } i = 1, 2, ..., n;\ j = 1, 2, ..., T_i. \tag{8}$$

In this paper, a new Bayesian approach is presented for nonparametric multivariate regression with Gaussian errors. A smoothness 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 the computations are carried out using the Gibbs sampling schemes (Wood et. al., 2000). With a burn-in period, it is assumed that iterations have converged to draws from posterior distributions. A random sample from the convergence period are used to estimate characteristics of the posterior distribution. This model is used for estimation of function f and to predict for the future values. We analyze a real data from one Taiwan based insurance company. A comparison is being carried out between the proposed approach against other existing techniques.

3 The Thin-Plate Spline

RKHS methods have been successfully applied to a wide varieties of problems in the field of optimal approximation, which include interpolation and smoothing via spline function in one or more dimensions. The one dimensional case is generalized to the multidimensional case by Duchon (1977). Duchon's surface spline is called "thin plate" spline, because they approximate the equilibrium position of a thin plate deflected at scatter points. For an application of thin-plate splines to meteorological problems see Wahba and Wendelberger (1980).

3.1 The Thin-Plate Spline on E^d

The theoretical foundations for the thin-plate spline were from Duchon (1975, 1976, 1977) and Meinguet (1979), and some further results and applications to meteorological problems were given in Wahba and Wendelberger (1980) and Wood et. al. (2000).

Let us define the penalty functional

$$J_m(f) = \int_0^1 (f^{(m)}(u))^2 du.$$

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

$$y_i = f(\mathbf{X}_i) + \epsilon_i, \quad i = 1, 2, \cdots, n,$$

where $\mathbf{X}_i \in \Omega$ and Ω is a general index set. The function f is assumed to be a smooth function in a reproducing kernel Hilbert space H of a real-valued functions on Ω. The $\{\epsilon_i\}$ are independent zero mean errors with common unknown variance. It is desired to find an estimate of f given $\mathbf{y} = (y_1, \cdots, y_n)'$. The estimate f_λ of f will be taken as the minimizer in H of

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(\mathbf{X}_i))^2 + \lambda J_m(f), \tag{9}$$

where $J_m(f)$ is a seminorm on H with M-dimensional null space spanned by $\phi_1, \cdots, \phi_M$, $M < n$. The seminaorm on the vector space H is a mapping $p : H \to \mathbf{R}$ satisfying $\|a\| \geqq 0$, $\|\alpha a\| = |\alpha|\|a\|$, and $\|a+b\| \leqq \|a\| + \|b\|$. Here a and b are arbitrary vectors in H and α is any scalar.

In the thin-plate spline case, we will assume $f \in \chi$, a space of functions whose partial derivatives of total order m are in $L_2(E^d)$. The data model is given by

$$y_i = f(x_1(i), \cdots, x_d(i)) + \epsilon_i, \quad i = 1, 2, \cdots, n, \tag{10}$$

where $f \in \chi$ and $\epsilon = (\epsilon_1, \cdots, \epsilon_n)' \sim N(0, \sigma^2 I)$. And $J(f) = J_m^d(f)$ is given by

$$J_m^d(f) = \sum_{\alpha_1+\cdots+\alpha_d=m} \frac{m!}{\alpha_1! \cdots \alpha_d!} \times \int \cdots \int \left(\frac{\partial^m f}{\partial x_1 \cdots \partial x_d}\right)^2 dx_1 \cdots dx_d. \tag{11}$$

We want χ endowed with the seminorm $J_m^d(f)$ to be an RKHS (that is, for the evaluation functionals in χ to be bounded with respect to $J_m^d(f)$). Then, a thin-plate smoothing spline is the solution to the following variational problem. Find $f \in \chi$ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_1(i), \cdots, x_d(i)))^2 + \lambda J_m^d(f). \tag{12}$$

Let us use the notation $t = (x_1, \cdots, x_d)'$ and $t_i = (x_1(i), \cdots, x_d(i))'$. The null space of the penalty functional $J_m^d(f)$ is the M-dimensional space spanned by the polynomials in d variables of total degree $\leqslant m-1$, where

$$M = \binom{d+m-1}{d}. \tag{13}$$

In the space $H = \{f : J_m^d(f) < \infty\}$ with $J_m^d(f)$ as a square semi norm, it is necessary that $2m - d > 0$ for the evaluation functional $L_t f = f(t)$ to be continuous; see Duchon (1977), Meinguet (1979), and Wahba and Wendelberger (1980). For $m = 2$, $d = 2$,

$$J_2^2(f) = \int\int [(\frac{\partial^2 f}{\partial x_1^2})^2 + 2(\frac{\partial^2 f}{\partial x_1 \partial x_2})^2 + (\frac{\partial^2 f}{\partial x_2^2})^2] dx_1 dx_2, \tag{14}$$

with $M = 3$, and the null space is spanned by ϕ_1, ϕ_2, ϕ_3 given by

$$\phi_1(x_1, x_2) = 1, \phi_2(x_1, x_2) = x_1, \phi_3(x_1, x_2) = x_2.$$

Before we go further, we need additional notations. Let $s, t \in E^d$, $s = (s_1, \cdots, s_d)'$ and $t = (t_1, \cdots, t_d)'$, then

$$| s - t |= (\sum_{i=1}^{d}(s_i - t_i)^2)^{1/2}.$$

We can define

$$\begin{aligned} E(r) &= \theta_m^d r^{2m-d} \log r \quad \text{if } d \text{ even} \\ &= \theta_m^d r^{2m-d} \qquad\quad \text{if } d \text{ odd,} \end{aligned} \tag{15}$$

where

$$\begin{aligned} \theta_m^d &= \frac{(-1)^{d/2+1}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} \quad \text{if } d \text{ even} \\ &= \frac{(-1)^m \Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} \qquad\quad \text{if } d \text{ odd.} \end{aligned} \tag{16}$$

We can also define

$$E_m(s,t) = E(| s - t |). \tag{17}$$

Duchon (1977) showed that, if $t_1, \cdots, t_n$ are such that least squares regression on $\phi_1, \cdots, \phi_M$ is unique, then (12) has an unique minimizer f_λ with representation

$$f_\lambda(t) = \sum_{\nu=1}^{M} d_\nu \phi_\nu + \sum_{i=1}^{n} c_i E_m(t, t_i). \tag{18}$$

Note that θ_m^d can be absorbed into c_i in (18). Let $u_1, \cdots, u_M$ be any fixed points in E^d such that least squares regression on the M-dimensional space of polynomials of total degree less than m at the points $u_1, \cdots, u_M$ is unique. Let $p_1, \cdots, p_M$

be the polynomials of total degree less than m satisfying $p_i(u_j) = 1, i = j$ and $p_i(u_j) = 0, i \neq j$. And let

$$K^1(s,t) = E_m(s,t) - \sum_{i=1}^{M} p_i(t)E_m(u_i,s) \tag{19}$$
$$- \sum_{j=1}^{M} p_j(s)E_m(t,u_j)$$
$$+ \sum_{i=1}^{M}\sum_{j=1}^{M} p_i(t)p_j(s)E_m(u_i,u_j).$$

It can be shown that K^1 is positive semidefinite and is a reproducing kernel for H_K and f_λ has a representation (Wahba, 1990)

$$f_\lambda = \sum_{\nu=1}^{M} d_\nu \phi_\nu + \sum_{i=1}^{n} c_i K^1_{t_i}(t), \tag{20}$$

where

$$K^1_t(\cdot) = K^1(t,\cdot).$$

The result from (20) can be shown to be the same as (18).

3.2 Bayes Model Behind The Thin-Plate Spline

Let us now take a look at the Bayes estimates behind the thin-plate spline. It is known that certain Bayes estimates are solutions to variational problems, and vice versa. Consider the random effect model

$$F(t) = \sum_{\nu=1}^{M} \theta_\nu \phi_\nu(t) + b^{1/2} X(t), \quad t \in [0, 1], \tag{21}$$

$$Y_i = F(t_i) + \epsilon_i, \quad i = 1, \cdots, n.$$

Let $\{\phi_1, \cdots, \phi_M\}$ span H_0, the space of polynomials of total degree less than m, and H_1 be a RKHS with the reproducing kernel defined by

$$EX(s)X(t) = K^1(s,t),$$

where $K^1(s,t)$ given by (19). Then, the model in (21) will result in the thin-plate spline. To understand this result, let

$$\widetilde{y_i} = y_i - \sum_{\nu=1}^{M} \theta_\nu \phi_\nu(t_i),$$

and set $f(t_i) = b^{1/2} X(t_i)$. Then (21) becomes

$$\widetilde{y_i} = f(t_i) + \epsilon_i, \quad i = 1, \cdots, n,$$

with $\epsilon = (\epsilon_1, \cdots, \epsilon_n)' \sim N(0, \sigma^2 I)$ and

$$\begin{aligned} Ef(s)f(t) &= E[b^{1/2} X(s) b^{1/2} X(t)] \\ &= bE[X(s)X(t)] \\ &= bK^1(s,t). \end{aligned}$$

Then, from Wahba (2000),

$$E\left(\begin{bmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_n) \end{bmatrix} \mid \tilde{\mathbf{y}}\right) = K^1\left(K^1 + \frac{\sigma^2}{b}\mathbf{I}\right)^{-1}\tilde{\mathbf{y}} \tag{22}$$

$$\equiv A(\lambda)\tilde{\mathbf{y}}, \text{ with } \lambda = \frac{\sigma^2}{b}.$$

$A(\lambda)$ is known as the influence matrix and we will use the result later.

Now, consider the variational problem in H_1, we want to find f_λ to minimize

$$\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - f(t_i))^2 + \lambda \parallel f \parallel_{H_1}^2,$$

where $\parallel f \parallel_{H_1}^2$ is the squared norm in H_1. It can be shown that

$$E\left(\begin{bmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_n) \end{bmatrix} \mid \tilde{\mathbf{y}}\right) = K^1(K^1 + \lambda\mathbf{I})^{-1}\tilde{\mathbf{y}}$$

$$\equiv A(\lambda)\tilde{\mathbf{y}}.$$

In summary, given the prior $\mathbf{f} \sim N(0, bK^1)$, a zero-mean Gaussian stochastic process with $\epsilon \doteq (\epsilon_1, \cdots, \epsilon_n)' \sim N(0, \sigma^2\mathbf{I})$, the posterior mean for $\mathbf{f}$ given $\tilde{\mathbf{y}}$ is the solution to a variational problem in an RKHS.

4 Bivariate Regression

Let us now return to credibility problems. Recall that, in the Bühlmann-Straub model, we wish to use the conditional distribution $f_{Y_{T+1}|\Theta}(y_{T+1}|\theta)$ or the hypothetical mean $E(Y_{T+1}|\Theta = \theta) \equiv \mu_{T+1}(\theta)$ for estimation of next year's claims. Since we have observed $\mathbf{y}$, one suggestion is to approximate $\mu_{T+1}(\theta)$ by a linear function of the past data. It turns out that the resulting credibility premium formula $Z\bar{Y} + (1 - Z)\mu$ is of this form. The idea is to restrict estimators of the form $\alpha_0 + \sum_{t=1}^{T} \alpha_t Y_t$, where $\alpha_0, \alpha_1, ..., \alpha_T$ need to be chosen. We will choose the $\alpha' s$ to minimize square error loss, that is,

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

We denote the result by $\widetilde{\alpha}_0, \widetilde{\alpha}_1, ..., \widetilde{\alpha}_T$ for the values of $\alpha_0, \alpha_1, ..., \alpha_T$ which minimize Q. Then the credibility premium can be written as:

$$\widetilde{\alpha}_0 + \sum_{t=1}^{T} \widetilde{\alpha}_t Y_t.$$

Meanwhile, the resulting $\widetilde{\alpha}_0, \widetilde{\alpha}_1, ..., \widetilde{\alpha}_T$ also minimize

$$Q_1 = E\left\{[E(Y_{T+1} \mid \mathbf{Y} = \mathbf{y}) - \alpha_0 - \sum_{t=1}^{T} \alpha_t Y_t]^2\right\}$$

and

$$Q_2 = E\left\{[Y_{T+1} - \alpha_0 - \sum_{t=1}^{T} \alpha_t Y_t]^2\right\}.$$

Hence, the credibility premium $\widetilde{\alpha}_0 + \sum_{t=1}^{T} \widetilde{\alpha}_t Y_t$ is the best linear estimator of each of the hypothetical mean $E(Y_{T+1}|\Theta = \theta)$, the Bayesian premium $E(Y_{T+1} \mid \mathbf{Y} = \mathbf{y})$, and Y_{T+1} in the sense of square error loss.

Now, we want to extend the standard credibility techniques into nonparametric regression models. In the credibility context, suppose we wish 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 establish a model as the nonparametric regression $y_{ij} = g(y_{i,j-1}, y_{i,j-2}) + \epsilon_i$, $i = 1, ..., n$, where g is a smooth function of its arguments. What we want to accomplish is to model the dependency between the current observations y_{ij} for all policyholders $i = 1, ..., n$, and those past losses $y_{i,j-1}$ and $y_{i,j-2}$ through a nonparametric regression at occasion j. Once the model is established, we can perform one-step ahead prediction on $y_{i,j+1}$ by using y_{ij} and $y_{i,j-1}$ as covariates. For notational convenience, we let y_i stands for the dependent variable (current observation) and use s_i or t_i for covariates (past losses). We develop a methodology to estimate the function g, given the data, from a nonparametric regression perspective. We will be using a Bayesian approach to fit the proposed model using a Gaussian prior on the unknown function g, which uses the reproducing kernel of a thin-plate spline as the covariance of the prior distribution (Wahba, 1990, p.30).

4.1 Model and Prior

Without loss of generality, we assume variables s_i, t_i lie in the interval $[0, 1]$. Consider the model from the bivariate regression model

$$y_i = g(s_i, t_i) + \epsilon_i, \quad i = 1, \cdots, n, \tag{23}$$

where g is a smooth regression of the variables s and t, and the errors ϵ_i are independent $N(0, \sigma^2)$. It is convenient to write (23) as

$$y_i = \alpha_0 + \alpha_1 s_i + \alpha_2 t_i + f(s_i, t_i) + \epsilon_i, \quad i = 1, \cdots, n, \tag{24}$$

with f having the zero initial conditions:

$$\begin{aligned} f(0,0) &= 0, \\ \frac{\partial f}{\partial s}(0,0) &= 0, \\ \frac{\partial f}{\partial t}(0,0) &= 0, \end{aligned} \tag{25}$$

which means that

$$\begin{aligned} \alpha_0 &= g(0,0), \\ \alpha_1 &= \frac{\partial g}{\partial s}(0,0), \\ \alpha_2 &= \frac{\partial g}{\partial t}(0,0). \end{aligned}$$

Model (24) has the same form as (21) with

$$\phi_1(s_i, t_i) = 1, \phi_2(s_i, t_i) = s_i, \phi_3(s_i, t_i) = t_i.$$

Now we can specify the prior on (24).

The prior for $f(s,t)$ is the reproducing kernel for the thin-plate spline in (19). This means that $f(s,t)$ is of zero-mean Gaussian random variables with the covariance function between $f(s_i, t_i)$ and $f(s_j, t_j)$ given by

$$cov\{f(s_i,t_i), f(s_j,t_j)\} = \tau^2 \mathbf{K}\{(s_i,t_i),(s_j,t_j)\}. \tag{26}$$

Using results from (19), the kernel $\mathbf{K}$ is given by

$$\begin{aligned}\mathbf{K}\{(s_i,t_i),(s_j,t_j)\} = {} & E\{(s_i,t_i),(s_j,t_j)\} \\ & - \sum_{k=1}^{3} p_k(s_j,t_j) E\{u_k,(s_i,t_i)\} \\ & - \sum_{k=1}^{3} p_k(s_i,t_i) E\{(s_j,t_j),u_k\} \\ & + \sum_{k=1}^{3}\sum_{l=1}^{3} p_k(s_i,t_i) p_l(s_j,t_j) E\{u_k,u_l\},\end{aligned} \tag{27}$$

where

$$E\{(s_i,t_i),(s_j,t_j)\} = r^2\log(r), \; r = \sqrt{(s_i-s_j)^2+(t_i-t_j)^2}. \tag{28}$$

This is because that, with $d=2$ and $m=2$, we have $J_2^2(f)$ given by (14). Obviously, $d/2+1$ is even in (16), so $E(r)$ is proportional to $r^2\log(r)$. Note that θ_m^d can be absorbed into c_i in (18). Furthermore, let

$$p_1(s_i,t_i) = -1+2s_i+2t_i, \; p_2(s_i,t_i) = 1-2s_i, \; p_3(s_i,t_i) = 1-2t_i. \tag{29}$$

By choosing

$$u_1 = (\frac{1}{2},\frac{1}{2}), u_2 = (0,\frac{1}{2}), u_3 = (\frac{1}{2},0),$$

we have p_1,p_2,p_3 be the polynomials of total degree less than 2 satisfying $p_i(u_j) =$

$1, i = j$ and $p_i(u_j) = 0, i \neq j$. Then we are now ready to apply the random effect model in (21).

To complete the prior specification for model (24), we take uninformative priors for all unknown parameters (Wood et. al., 2000). We take uniform independent prior on $[0, 10^{10}]$ for the smoothing parameter τ^2. The prior for $\alpha = (\alpha_0, \alpha_1, \alpha_2)'$ is

$$\alpha \sim \mathbf{N}(\mathbf{0}, c\mathbf{I}),$$

with $c \longrightarrow \infty$. The prior for σ^2 is

$$p(\sigma^2) \propto (\sigma^2)^{-1-10^{-8}} \exp(-10^{-10}/\sigma^2).$$

The resulting Bayes estimate will be the solution to the variational problem in (12) with $d = 2$ and $m = 2$.

4.2 Model Implementation

In this subsection, we will discuss the implementation of the model in (24). To make this model computationally feasible, we will consider a transformed model. As in Wood et. al. (2000), the sampling scheme requires factoring the covariance matrix $\mathbf{K}$ as $\mathbf{QDQ}'$, where $\mathbf{Q}$ is an orthonormal matrix and $\mathbf{D}$ is the diagonal matrix with diagonal elements, d_i, that are the eigenvalues of $\mathbf{K}$.

To ease the notation, we rewrite model in (24) as

$$y_i = \alpha_0 + \alpha_1 s_i + \alpha_2 t_i + f_i + \epsilon_i, \quad i = 1, \cdots, n, \tag{30}$$

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

$$\alpha = (\alpha_0, \alpha_1, \alpha_2)', \tag{31}$$

$$\mathbf{y} = (y_1, \cdots, y_n)',$$

$$\epsilon = (\epsilon_1, \cdots, \epsilon_n)',$$

and

$$\mathbf{Z} = \begin{bmatrix} 1 & s_1 & t_1 \\ 1 & s_2 & t_2 \\ \vdots & \cdots & \vdots \\ 1 & s_n & t_n \end{bmatrix}, \tag{32}$$

then we can write (30) in the matrix form

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & s_1 & t_1 \\ 1 & s_2 & t_2 \\ \vdots & \cdots & \vdots \\ 1 & s_n & t_n \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \tag{33}$$

that is,

$$\mathbf{y} = \mathbf{Z}\alpha + \mathbf{f} + \epsilon, \tag{34}$$

with the priors,

$$
\begin{aligned}
&\alpha \sim \mathbf{N}(\mathbf{0}, c\mathbf{I}), \\
&\mathbf{f} \sim \mathbf{N}(\mathbf{0}, \tau^2\mathbf{K}), \\
&\epsilon \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I}), \\
&\tau^2 \sim unif[0, 10^{10}], \\
&\sigma^2 \sim IG(10^{-8}, 10^{-10}),
\end{aligned} \tag{35}
$$

where τ^2 is uniformly distributed in the interval $[0, 10^{10}]$ and σ^2 has a inverse Gamma distribution with parameters 10^{-8} and 10^{-10}. Let us return to the covariance matrix $\mathbf{K}$. Since $\mathbf{K}$ is positive definite, we can factor $\mathbf{K}$ as $\mathbf{QDQ}'$ such that

$$
\mathbf{QQ}' = \mathbf{I}. \tag{36}
$$

We can pre-multiply $\mathbf{Q}'$ to (34), so we have $\mathbf{y}^* = \mathbf{Q}'\mathbf{y}$. And the model becomes

$$
\mathbf{y}^* = \mathbf{Z}^*\alpha + \mathbf{f}^* + \epsilon^*, \tag{37}
$$

where

$$
\begin{aligned}
\mathbf{Z}^* &= \mathbf{Q}'\mathbf{Z}, \\
\mathbf{f}^* &= \mathbf{Q}'\mathbf{f}, \\
\epsilon^* &= \mathbf{Q}'\epsilon.
\end{aligned} \tag{38}
$$

The priors for α, τ^2, σ^2 will remain the same as in (35). Meanwhile, ϵ^* has the same distribution as $\epsilon \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{I})$ because of (36) in $\mathbf{N}(\mathbf{0}, \sigma^2\mathbf{Q}'\mathbf{IQ})$. However, the prior for $\mathbf{f}^*$ becomes

$$\mathbf{f}^* \sim \mathrm{N}(\mathbf{0}, \tau^2\mathbf{D}), \tag{39}$$

because of

$$\begin{aligned} Var(\mathbf{Q}'\mathbf{f}) &= \mathbf{Q}' Var(\mathbf{f})\mathbf{Q}. \\ &= \mathbf{Q}'\tau^2\mathbf{K}\mathbf{Q} \\ &= \tau^2\mathbf{Q}'\mathbf{Q}\mathbf{D}\mathbf{Q}'\mathbf{Q} \\ &= \tau^2\mathbf{D}, \end{aligned}$$

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

4.3 Bivariate Regression for the Bühlmann-Straub Model

Consider data in Bühlmann-Straub Model, it allows different number of exposure units or different distribution of claim size across past policy years. This can be handled in model (30) by assuming

$$\epsilon_i \sim N(0, \frac{\sigma^2}{w_i}), \quad i = 1, \cdots, n, \tag{40}$$

where w_i is the corresponding weight for data value y_i. We can also have the same matrix form as (34),

$$\mathbf{y} = \mathbf{Z}\alpha + \mathbf{f} + \epsilon, \tag{41}$$

but with the priors

$$
\begin{aligned}
\alpha &\sim \mathbf{N}(\mathbf{0}, c\mathbf{I}), \\
\mathbf{f} &\sim \mathbf{N}(\mathbf{0}, \tau^2\mathbf{K}), \\
\epsilon &\sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{W}^{-1}), \\
\tau^2 &\sim unif[0, 10^{10}], \\
\sigma^2 &\sim IG(10^{-8}, 10^{-10}).
\end{aligned} \tag{42}
$$

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

This model can be easily transformed to a similar model as in (34) and (35). Then we can implement the modified model analogously as in Section 4.2. Now let

$$
\begin{aligned}
\mathbf{y}' &= \sqrt{\mathbf{W}}\mathbf{y} \\
\mathbf{Z}' &= \sqrt{\mathbf{W}}\mathbf{Z}, \\
\mathbf{f}' &= \sqrt{\mathbf{W}}\mathbf{f}, \\
\epsilon' &= \sqrt{\mathbf{W}}\epsilon.
\end{aligned} \tag{43}
$$

An interim model is given by

$$
\mathbf{y}' = \mathbf{Z}'\alpha + \mathbf{f}' + \epsilon', \tag{44}
$$

with the priors

$$\begin{aligned} &\alpha \sim \mathbf{N}(\mathbf{0}, c\mathbf{I}), \\ &\mathbf{f}' \sim \mathbf{N}(\mathbf{0}, \tau^2 \sqrt{\mathbf{W}}\mathbf{K}\sqrt{\mathbf{W}}), \\ &\epsilon' \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \\ &\tau^2 \sim unif[0, 10^{10}], \\ &\sigma^2 \sim IG(10^{-8}, 10^{-10}). \end{aligned} \tag{45}$$

We can then set $\mathbf{K}' = \sqrt{\mathbf{W}}\mathbf{K}\sqrt{\mathbf{W}}$. This means that $f'(s,t)$ is of zero-mean Gaussian random variables with the covariance function between $f'(s_i,t_i)$ and $f'(s_j,t_j)$ given by

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

This is just a different choice of the prior for the full bivariate surface $f'(s,t)$. We can then use results from (30) to (39) based on (44) to (45).

We use the Gibbs sampling scheme where the bivariate regression surface is modeled by the thin-plate spline prior as described earlier in the paper. A good introduction to the Gibbs sampler is given by Gelfand and 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 and Kohn (1996). The sampling scheme is similar to the one used by Wood et. al. (2000) in a model selection context. In our case, the estimates of α, $\mathbf{f}'$, τ^2, and σ^2 are obtained by generating the iterations $\alpha^{[j]}$, $\mathbf{f}'^{[j]}$, $\tau^{2[j]}$, and $\sigma^{2[j]}$ from the sampling scheme described in (45). The constant c is chosen to be a large number ($c = 10^{20}$) so as to ensure that the prior for α is essentially a noninformative flat prior.

4.4 One-Step Ahead Prediction

Many problems in actuarial science involve the building of a mathematical model that can be used to predict insurance costs or to forecast losses in the future, particularly the short-term future. Our approach is to establish a nonparametric regression model $y_{ij} = g(y_{i,j-1}, y_{i,j-2}) + \epsilon_i$, $i = 1, ..., n$, where g is a smooth function of its arguments. This model allows us to describe the dependency between the current observations y_{ij} for all policyholders $i = 1, ..., n$, and those past losses $y_{i,j-1}$ and $y_{i,j-2}$ through a nonparametric regression function at occasion j.

Suppose that we are interested in one step ahead prediction of $Y_{i,j+1}$. We take the posterior mean $E(Y_{i,j+1} \mid \mathbf{y})$ as the best predictor of $Y_{i,j+1}$ and use the posterior $var(Y_{i,j+1} \mid \mathbf{y})$ to obtain the posterior pointwise prediction interval. For convenience, 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 in (30), that is,

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

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

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

Therefore, the prediction is

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

where $\hat{f}_i$ is the expected noise on y_i given observed data from (22). After a burn-in period, it is assumed the iterations have converged to draws from the posterior distribution. We estimate the posterior mean and the posterior variance of $Y_{i,j+1}$ based on the values of $Y_{i,j+1}$ generated during the sampling period.

5 Nonparametric Regression with Higher Dimensions

Suppose the model is now extended to handle three variables. Similarly, we can treat s_i and t_i as the past losses and incorporate other relevant information as v_i. For example, v_i can be the number of years a policyholder remain in the same policy with the same insurer, or represents different driving age group in auto insurance. Then, the regression model is given by

$$y_i = g(s_i, t_i, v_i) + \epsilon_i, \qquad i = 1, \cdots, n,$$

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

$$y_i = \alpha_0 + \alpha_1 s_i + \alpha_2 t_i + \alpha_3 v_i + f(s_i, t_i, v_i) + \epsilon_i, \qquad i = 1, \cdots, n.$$

The prior on f is specified similarly to (26) and (27), that is

$$cov\{f(s_i, t_i, v_i), f(s_j, t_j, v_j)\} = \tau^2 \mathbf{K}\{(s_i, t_i, v_i), (s_j, t_j, v_j)\}$$

where

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

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

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

and

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

This is because of (13), where $m = 2$, $d = 3$, and

$$
\begin{aligned}
M &= \binom{d+m-1}{d} \\
&= \binom{4}{3} \\
&= 4.
\end{aligned}
$$

Without loss of generality, we assume that the variables s, t, and v all lie in the interval $[0,1]$.

Let $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)'$ be the vector of linear regression parameters, and let

$$
\mathbf{Z} = \begin{bmatrix} 1 & s_1 & t_1 & v_1 \\ 1 & s_2 & t_2 & v_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & s_n & t_n & v_n \end{bmatrix}
$$

The priors for α, the smoothing parameter τ^2, and σ^2 are the same as in (35) which gives us

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

The model implementation and the sampling scheme will be exactly the same as in the bivariate model.

6 Application to medical insurance data

In this section, the results of Bayesian nonparametric regression model for the Bühlmann-Straub type data with unequal exposure units will be illustrated by an application to a collective medical health insurance data from an insurance company in Taiwan. We consider a portfolio consisting of thirty-five group policyholders that has been observed for a period of three years. The claim associated with group j $(= 1, \cdots, 35)$ in year of observation t $(= 1, 2, 3)$ is represented by the random variable Y_{jt}, which is an average taken over w_{jt} employee. We choose groups with moderate group size (23 to 80 individuals), and assume that the number of employee does not change over the periods. Therefore, we have $w_{jt} = w_j$ for all t, and the claim Y_{jt} with weights w_j fulfill the Bühlmann-Straub assumptions. Table 2 shows some observed realizations of the Y_{jt}, and the numbers of employee w_j. We want to determine the estimated premium to be charged to each group in year 4. The data is in new Taiwan dollar (NTD). The exchange rate is about 1 US dollar to 35.16 new Taiwan dollar in March, 2002.

We consider the semiparametric regression model discussed in section 4.3,

$$y_{j,t} = \alpha_0 + \alpha_1 y_{j,t-1} + \alpha_2 y_{j,t-2} + f(y_{j,t-1}, y_{j,t-2}) + \epsilon_i, \quad j = 1, \cdots, n, \tag{46}$$

where $\epsilon_i \sim N(0, \frac{\sigma^2}{w_i})$, $i = 1, \cdots, n$. The function $f(y_{i,t-1}, y_{i,t-2})$ has zero initial conditions and is estimated nonparametrically using the priors in (42). The scatter plot, fitted surface by (46), and fitted surface by the Bühlmann-Straub (BS) model are shown in Figure 1. The contour plots shown in Figure 2 provide a better look of different levels of surfaces.

To examine the performance of the regression function, the average squared error (ASE) was calculated for the estimates of the regression functions. The ASE is calculated as follows

$$ASE = \frac{1}{\sum_{j=1}^{n} w_j} \sum_{j=1}^{n} w_j (\hat{g}(s_j, t_j) - g(s_j, t_j))^2. \tag{47}$$

The ASE of the bivariate spline nonparametric regression model is about 214.26, while that of the Bühlmann-Straub model is 2208.07. Clearly, the bivariate spline nonparametric regression model outperforms the Bühlmann-Straub model.

Our goal is to determine the estimated premium to be charged to each group in year 4. We perform one-step ahead prediction discussed in section 4.5. We estimate the posterior mean and variance of $Y_{j,4}$ using empirical estimates based on the values of $y_{j,4}$ generated during the sampling period. Figure 3 shows some of the posterior distributions of $Y_{j,4}$. Some of the estimated premiums together with 95 percent posterior pointwise prediction intervals (in parenthesis) are shown in Table 3. For example, for group 3, the estimated premium is 5965.36 NTD for each employee in this group, and the total estimated premium is $49 \times (5965.36) = 292302.64$ NTD. Similar calculations can be done for other groups.

7 Conclusions

Many problems in actuarial science involve the building mathematical models that can be used to predict insurance cost in the future, particularly the short-term future. A Bayesian nonparametric approach is proposed to the problem of risk modeling. The model incorporates past and present observations related to the risk, as well as relevant covariate information, and uses MCMC technique to compute the predictive distribution of the future claims based on the available data, where the covariance structure 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 feasible approach to the computation of posterior distributions. Combined with assumed thin-plate spline structure of the regression surface and the computational availability of the 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 applications, the technique can be easily extended to higher dimensional problems. Our investigation shows that this method performs at a superior level compared to the existing techniques in the credibility literature.

In this paper, we have outlined a new approach to modeling actuarial and financial data. The model uses a Bayesian nonparametric procedure in a novel manner by incorporating a Gaussian prior on function space. We believe that this procedure provides a flexible approach to function estimation and can be used successfully in the statistical analyses of a wide range of important problems.

	Policyholder	Year 1	Year 2	Year 3	Year 4
Average claim	1	5419.09	1691.38	5984.65	?
No. in group		74	74	74	74
Average claim	2	5603.50	4150.12	5797.48	?
No. in group		52	52	52	52
⋮	⋮	⋮	…	⋮	⋮
Average claim	35	4554.38	4646.96	5059.80	?
No. in group		80	80	80	80

Table 2: Average claims in group policyholders during three years.

	Policyholder	Year 4
Average claim	3	5965.36 (3080.30, 8877.31)
No. in group		49
Average claim	5	5485.61 (1429.58, 9398.640)
No. in group		45
Average claim	9	5024.05 (−595.66, 10495.44)
No. in group		42
Average claim	14	5000.68 (2173.55, 7648.52)
No. in group		31
Average claim	20	6437.74 (2558.72, 7116.71)
No. in group		36
Average claim	24	5217.51 (−361.08, 10462.63)
No. in group		35
Average claim	26	4959.50 (308.02, 9345.16)
No in group		42
Average claim	35	5074.70 (4620.35, 5497.48)
No. in group		80

Table 3: Estimated average claim for year 4 with 95 percent posterior pointwise prediction interval in parenthesis.

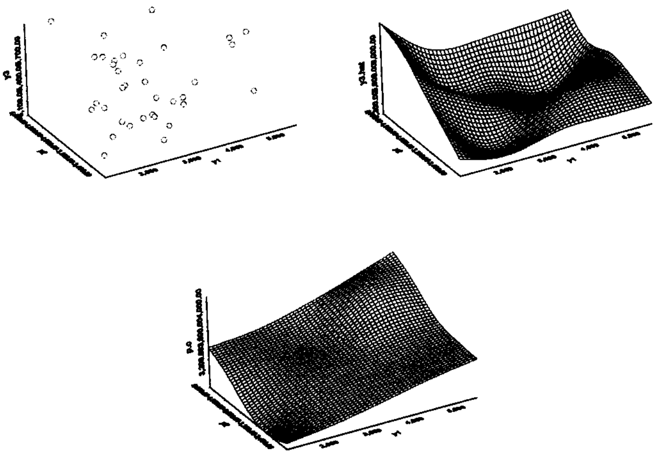

Figure 1: Surface Plots. (a) Scatter plot. (b) Plot of regression surface as a function of y_{t-1} and y_{t-2}. (c) Plot of BS model surface.

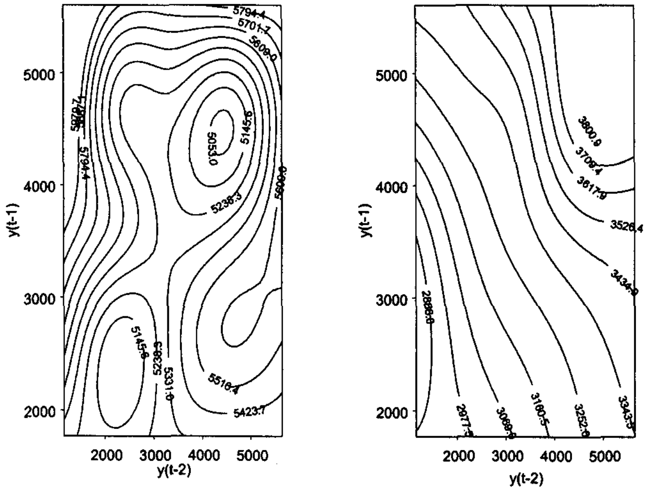

Figure 2: Contour Plots. (a) Plot of regression function as a function of y_{t-1} and y_{t-2}. (b) Plot of BS model.

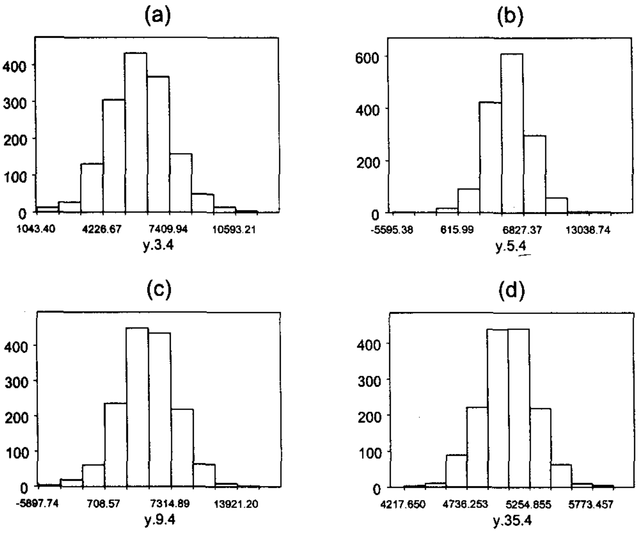

Figure 3: Posterior Plots. (a) $Y_{3,4}$. (b) $Y_{5,4}$. (c) $Y_{9,4}$. (d) $Y_{35,4}$.

References

1. Aronszajn, N. Theory of Reproducing Kernels. Transactions of the American Mathematical Society, 68:337-404, 1950.

2. Bühlmann, H. Experience Rating anf Credibility. ASTIN Bulletin, 4:199-207, 1967.

3. Duchon, J. Functions Splines et Vecteurs Aleatoires. Technical Report 213, Seminaire d'Analyse Numeriaue, 1975.

4. Duchon, J. Functions-spline et Esperances Conditionnelles de Champs Gaussians, Annal Science University Clermont Ferrand II Mathematics, 14:19-27. 1976.

5. Duchon, J. Spline Minimizing Rotation-Invariant Semi-Norms in Sobolev Spaces. in Constructive Theory of Functions of Several Variables, Springer-Verlag, Berlin, 1977.

6. Evgeniou, T., Pontil, M., and Poggio, T. Regularization Networks and Support Vector Machines. Advances in Computational Mathematics, 13:1-50, 2000.

7. Gelfand, A.E. and Smith, A.F.M. Sampling-based Approaches to Calculating Marginal Densities. Journal of the American Statistical Association, 85:398-409, 1990.

8. Meinguet, J. Multivariate Interpolation at Arbitrary Points Made Simple. Journal of Applied Mathematical Physics (ZAMP), 30:292-304, 1979.

9. Wahba, G. Spline Models for Observational Data. SIAM, Philadephia, 69, 1990.

10. Wahba, G. An Introduction to Model Building with Reproducing Kernel Hilbert Space. Technical Report NO. 1020, Department of Statistics, University of Wisconsin, Madison, WI, 2000.

11. Wahba, G. and Wendelberger, J. Some New Mathematical Methods for Variational Objective Analysis using Splines and Cross-Validation. Monthly Weather Review, 108:1122-1145, 1980.

12. Weinert, H.L (Editor). Reproducing Kernel Hilbert Spaces: applications in statistical signal processing. Hutchinson Ross Pub. Co., 1982.

13. Wendelberger, J. Smothing Noisy Data with Multidimensional Splines and Generalied Cross Validation. Ph.D. Thesis. Department of Statistics, University of Wisconsin, Madison, WI, 1982.

14. Whittaker, E.T. On A New Method of Graduation. Proceedings of Edinburgh Mathematical Society, 41:63-75, 1923.

15. Wang, C. and Kohn, R. A Bayesian Approach to Additive Semiparametric Regression. Journal of Econometrics, 7:209-223, 1996.

16. Wood, S., Kohn, R., Shively, T., and Jiang, W. Model Selection in Spline Nonparametric Regression. Preprint, 2000.