Hierarchical Bayesian curve fitting and smoothing*

Jean-François ANGERS and Mohan DELAMPADY

Université de Montréal and *University of British Columbia*

Key words and phrases: Hierarchical Bayes, function estimation, curve fitting, smoothing.
AMS 1985 subject classifications: Primary 62F15, secondary 62H12, 62M99.

ABSTRACT

Estimation of a smooth function is considered when observations on this function added with Gaussian errors are observed. The problem is formulated as a general linear model, and a hierarchical Bayesian approach is then used to study it. Credible bands are also developed for the function. Sensitivity analysis is conducted to determine the influence of the choice of priors on hyperparameters. Finally, the methodology is illustrated using real and simulated examples where it is compared with classical cubic splines. It is also shown that our approach provides a Bayesian solution to some problems in discrete time series.

RÉSUMÉ

Nous étudierons le lissage d'une fonction lorsque les observations de cette fonction sont sujettes à des erreurs gaussiennes. Le problème sera formulé à l'aide d'un modèle linéaire et nous utiliserons l'approche bayesienne hiérarchique pour l'étudier. De plus nous développerons des bandes de crédibilité pour le lissage. Une analyse de sensibilité sera faite pour déterminer l'influence sur le lissage de la densité a priori sur les hyperparamètres. Pour conclure, nous illustrerons cette nouvelle méthodologie à l'aide de données réelles et d'une simulation; nous comparerons les résultats obtenus avec ceux fournis par les splines cubiques. Il sera aussi montré que cette approche fournit une solution bayesienne à quelques problèmes en séries chronologiques.

1. INTRODUCTION

Consider the model

$$y(t_i) = g(t_i) + \epsilon_i, \quad t_i \in \mathcal{T}, \qquad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)^{\mathrm{T}} \sim \mathrm{N}(0, \sigma^2 \mathrm{I})$ (σ^2 unknown) and $g(\cdot)$ is a smooth function defined on some index set $\mathcal{T}$. This model has a wide range of applications and can be used to represent any continuous phenomenon (in one or several dimensions) which is measured with independent errors (or with known covariance matrix). We were motivated for the present study by the National Surface Water Survey, which collected large amounts of data on acidic depositions in fresh-water lakes in the U.S. and Canada. [Details can be found in Linthurst *et al.* (1986).] Our ultimate objective is to construct two-dimensional contours of acidic depositions using these data.

In this paper, only the one-dimensional case, i.e. $\mathcal{T} \subset \mathbb{R}$, is studied. The problem of interest is to estimate the function $g(t)$ for all $t \in \mathcal{T}$ and to provide an error band

*This research was partially funded by NSERC grants A9250 and OGP105078, and a grant from SIMS.

("credible band") for it. Before developing the main results in Sections 2.2 and 2.3, we shall describe the prior model in a hierarchical Bayesian way and show that the prior model used in this paper is equivalent to a generalized linear model. Using the Bayesian approach for linear models, an estimator for $g(t)$ will be proposed in Section 2.2, and an error band for it will be derived in Section 2.3. In Section 2.4, a procedure for the choice of the smoothness parameter will be proposed.

In Section 3, the proposed methodology will be applied to a simulated data set. It will also include a sensitivity analysis of the estimator with respect to the choice of priors on hyperparameters and with respect to the degree of smoothness indicated by the prior model. There has been a lot of interest recently on discrete time series with continuous intensity functions (e.g. Kitagawa and Gersch 1984). We shall also illustrate here how our procedure can be used in such situations. Data on monthly precipitation in Vancouver will be used. In this case $g(t)$ is the underlying intensity of precipitation, which can be reasonably assumed to be a smooth function.

It should be noted that all the results described here have been obtained using the hierarchical Bayes approach. Several other approaches have been utilized in the literature for estimating the regression function $g(\cdot)$. Among them are polynomial splines (Wahba 1978), Bayesian curve fitting (O'Hagan 1978), locally regular smoothers (Weerahandi and Zidek 1988), point-process fitting (Delampady 1987), and Bayesian multiple regression (Nebebe and Stroud, 1986). Later, in Section 2.1, our approach will be compared with some of these works to illustrate the flexibility and usefulness of our work.

2. DESCRIPTION OF THE MODEL AND DEVELOPMENT OF THE ESTIMATOR

The prior on $g(\cdot)$ will be described in several stages using the hierarchical Bayesian approach. To facilitate the construction of the prior, the function $g(\cdot)$ will be expanded in a Taylor series, that is,

$$\begin{aligned} g(t) &= g(t_0) + (t - t_0)g'(t_0) + \frac{(t-t_0)^2}{2!} g''(t_0) + \cdots \\ &\quad + \frac{(t-t_0)^{m-1}}{(m-1)!} g^{(m-1)}(t_0) + R_m(t) \\ &= \mathbf{\Phi}(t)^{\mathrm{T}} \boldsymbol{\theta} + R_m(t), \end{aligned} \tag{2}$$

where $t_0 \in \mathcal{T}$ (t_0 chosen by the user) and

$$\mathbf{\Phi}(t) = \left(1, t - t_0, \frac{(t-t_0)^2}{2!}, \ldots, \frac{(t-t_0)^{m-1}}{(m-1)!}\right)^{\mathrm{T}},$$

$$\boldsymbol{\theta} = \left(g(t_0), g'(t_0), \ldots, g^{(m-1)}(t_0)\right)^{\mathrm{T}},$$

$R_m(t)$ is the remainder, and m represents the smoothness parameter, being the number of derivatives plus 1. The parameters of interest will be $\boldsymbol{\theta}$ and $R_m(\cdot)$.

For convenience, it will be assumed that the first-stage prior of $\boldsymbol{\theta}$ is an m-variate normal distribution with mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_m)^{\mathrm{T}}$ and covariance matrix of the form $\phi^2\mathbf{\Gamma}$, where $\mathbf{\Gamma}$ is a known matrix. In most of the applications it will be reasonable to approximate $\mathbf{\Gamma}$ by the identity matrix. Since $R_m(\cdot)$ is the remainder from a Taylor series expansion, it is assumed to be of the form $R_m(t) = (t-t_0)^{m-1}Z(t-t_0)/(m-1)!$, where Z is a second-order-stationary Gaussian process with mean $\delta(t)$ and covariance kernel given by $Cov\,(Z(t), Z(s)) = \phi^2\rho(|t-s|)$, with $\rho(d)$ a monotone decreasing function of $0 \le d \le \infty$. The more slowly $\rho(\cdot)$ decreases, the more stable $R_m(\cdot)$ becomes. Therefore, it also measures the influence of the $y(t_i)$'s in estimating $R_m(t)$. Other covariance kernels, such as the spline kernel of Wahba (1978), were also considered, but they did not provide interesting results.

In order to carry out the calculations analytically as much as possible, the following model will be chosen for the hyperparameters. However, note that (Berger 1985; Goel and DeGroot 1981) the effect of the second- and third-stage priors on the resulting estimator is quite limited. Consequently, the hyperparameters μ and $\delta(\cdot)$ will be assumed to have the following prior distributions: μ will be m-variate normal with mean 0 and covariance matrix $\tau^2\mathbf{\Gamma}$, and δ will be a second-order-stationary Gaussian process with mean 0 and covariance kernel given by $Cov\,(\delta(t), \delta(s)) = \tau^2\rho(|t-s|)$. Since the first-stage model assumes $\boldsymbol{\theta}$ and $R_m(\cdot)$ to be independent, it seems natural to assume their first-stage expectations, i.e., μ and $\delta(\cdot)$, to be independent also, given the hyperparameters ϕ^2 and τ^2. The nuisance parameter σ^2 and the hyperparameters ϕ^2 and τ^2 will be given a noninformative prior to be specified later.

In summary, the model can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{Y} = \left(y(t_1), y(t_2), \ldots, y(t_n)\right)^{\mathsf{T}}$, $\mathbf{X} = (\mathbf{T}, \mathbf{1})$, $\mathbf{T} = \left(\boldsymbol{\Phi}(t_1), \boldsymbol{\Phi}(t_2), \ldots, \boldsymbol{\Phi}(t_n)\right)^{\mathsf{T}}$, $\boldsymbol{\beta} = (\boldsymbol{\theta}^{\mathsf{T}}, \mathbf{R}^{\mathsf{T}})^{\mathsf{T}}$, $\mathbf{R} = \left(R_m(t_1), R_m(t_2), \ldots, R_m(t_n)\right)^{\mathsf{T}}$, $\mathbf{T}_* = \left((t_1-t_0)^{m-1}/(m-1)!, (t_2-t_0)^{m-1}/(m-1)!, \ldots, (t_n-t_0)^{m-1}/(m-1)!\right)^{\mathsf{T}}$. The first stage prior on $\boldsymbol{\beta}$ is $(m+n)$-variate normal with mean $(\boldsymbol{\mu}^{\mathsf{T}}, \boldsymbol{\delta}^{\mathsf{T}})^{\mathsf{T}}$ and covariance matrix $\phi^2\boldsymbol{\Sigma}$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{\Gamma} & 0 \\ 0 & \mathbf{Q}_n \end{pmatrix},$$

and $\mathbf{Q}_n$ is the $n \times n$ matrix whose (i, j) element is given by

$$(Q_n)_{(i,j)} = \frac{1}{\{(m-1)!\}^2}\left\{(t_i - t_0)(t_j - t_0)\right\}^{m-1}\rho(|t_i - t_j|).$$

2.1. Comparison with Other Models.

In this subsection, our approach to the problem of curve fitting and smoothing will be compared with those used in Wahba (1978) and O'Hagan (1978), and with a hierarchical Bayes approach to multiple regression proposed in Nebebe and Stroud (1986).

By modelling observations with Equations (1) and (2), we take an approach similar to Wahba (1978). However, our hierarchical prior approach for modelling the prior parameters, $\boldsymbol{\theta}$ and $R_m(\cdot)$, differs substantially from the one used by Wahba. Instead, Wahba considers a diffuse prior on $\boldsymbol{\theta}$ by assuming that $\boldsymbol{\theta} \sim \mathsf{N}_m(0, \xi\mathbf{I})$ with $\xi \to \infty$. We feel that often the prior information available on $\boldsymbol{\theta}$ is not vague, and hence a diffuse first-stage prior on $\boldsymbol{\theta}$ is not appropriate as a general prescription. Also, the prior distribution used there on $R_m(\cdot)$ corresponds to the distribution of an integrated Wiener process with the covariance kernel

$$\begin{aligned} Q(s, t) &= \int_0^1 \frac{(s-u)_+^{m-1}}{(m-1)!}\,\frac{(t-u)_+^{m-1}}{(m-1)!}\,du \\ &= \{\min(s, t)\}^m \sum_{i=0}^{m-1} \frac{1}{m+i}\binom{m-1}{i}\{\min(s, t)\}^i |t-s|^{m-1-i}. \end{aligned}$$

We feel that our choice of the covariance kernel which arises from the representation $R_m(t) = t^{m-1}Z(t)/(m-1)!$ is more appropriate for a remainder term. The spline estimator proposed in Wahba (1978) corresponds to the limit, when ξ goes to ∞, of the posterior mode. Note that Wahba does not use the prior model to derive the proposed estimator, but instead uses it as a Bayesian justification of generalized splines.

O'Hagan (1978) assumes the model

$$y(t) = \mathbf{f}(t)^{\mathsf{T}}\boldsymbol{\beta}(t) + \epsilon,$$

where $\mathbf{f}(t)$ is a vector of known functions of t, and $\boldsymbol{\beta}(t)$ is a vector of unknown parameters which may depend on t. It is assumed that the errors have an i.i.d. normal distribution with known variance. The prior assumptions made are that $\boldsymbol{\beta}(\cdot)$ is a second-order-stationary Gaussian process with

$$\mathcal{E}[\boldsymbol{\beta}(t)|b_0] = \mathbf{b}_0,$$
$$\mathcal{E}[\{\boldsymbol{\beta}(t) - \mathbf{b}_0\}\{\boldsymbol{\beta}(t^*) - \mathbf{b}_0\}^{\mathsf{T}}|b_0] = \rho(|t - t^*|)B_0,$$

where $\mathbf{b}_0$ and B_0 are assumed to be known and to be independent of t. Also considered is the case where $\mathbf{b}_0$ is assigned a second-stage normal prior with known mean $\mathbf{b}^*$ and covariance $k\mathbf{B}^*$, with the multiplicative constant k allowed to increase to ∞, providing a diffuse prior. Our approach is very similar, but we prefer (2), since often substantial prior information is available about the derivatives of a regression function. However, exactly specifying $\boldsymbol{\mu}$ ($\mathbf{b}_0$) and $\boldsymbol{\Gamma}$ (B_0) is almost always an impossible task, so that a natural solution is to assign a second-stage prior on these quantities as we suggest. Also, we consider the assumption that the error variance σ^2 is known to be unrealistic and hence assign a noninformative prior distribution to this nuisance parameter. The major difference between the two approaches is our choice of hyperparameters using a careful sensitivity analysis instead of adopting a diffuse prior in the last stage. Clearly, our approach will provide an estimator of the regression function which is more robust with respect to the choice of priors. It is important to note also that careful choice of hyperparameters and some of our key simplifications will help keep the additional computations to a minimum.

Without the $R_m(t)$ term, the model described by Equations (1) and (2) can also be viewed as a polynomial regression model (Nebebe and Stroud 1986) with dependent errors. Also, our use of a second-stage prior on the hyperparameters of the first-stage prior is similar to their analysis. However, the representation (2) that we use in our model can be considered an appealing alternative to Bayesian multiple linear regression with polynomials. The remainder $R_m(\cdot)$ acts as an insurance against the possibility that the function $g(\cdot)$ may not be a polynomial of order $m-1$.

2.2. *Development of* $\hat{g}(t)$.

Let $\boldsymbol{\beta}_0 = (\boldsymbol{\mu}^{\mathsf{T}}, \boldsymbol{\delta}^{\mathsf{T}})^{\mathsf{T}}$. Then, using standard hierarchical Bayes techniques for the linear model (Lindley and Smith 1972), one can show that the first-stage posterior of $\boldsymbol{\beta}$ is a $(m+n)$-variate normal distribution with mean

$$(\phi^{-2}\boldsymbol{\Sigma}^{-1} + \sigma^{-2}\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}(\phi^{-2}\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_0 + \sigma^{-2}\mathbf{X}^{\mathsf{T}}\mathbf{Y}), \tag{3}$$

and covariance matrix

$$(\phi^{-2}\boldsymbol{\Sigma}^{-1} + \sigma^{-2}\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}.$$

If one uses the posterior mode as decision rule, the first-stage Bayes estimator is given by (3).

Note that the marginal of $\mathbf{Y}$ given $\boldsymbol{\beta}_0$, σ^2, and ϕ^2 is an n-variate normal with mean $\mathbf{X}\boldsymbol{\beta}_0$ and covariance matrix $\boldsymbol{\Sigma}_* = \sigma^2\mathbf{I} + \phi^2\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}}$.

The Bayes estimator of $\boldsymbol{\beta}$ is obtained by taking the posterior expectation of (3) with respect to the hyper- and nuisance parameters (cf. Berger 1985). The main reason for conducting hierarchical Bayesian analysis is to reduce the dimension of the numerical integration required to compute this posterior expectation. Since, as shown in Section 3, the estimator does not depend crucially on the second- and higher-stage hyperpriors, one can choose these hyperpriors in ways that will allow most of the calculations to be done analytically.

As mentioned previously, the prior on $\boldsymbol{\beta}_0$ is a $(m+n)$-variate normal centered at 0 with covariance matrix $\tau^2\boldsymbol{\Sigma}$. Therefore the posterior distribution of $\boldsymbol{\beta}_0$ given $\mathbf{Y}$, σ^2, ϕ^2, and τ^2 is $(m+n)$-variate normal with mean

$$(\tau^{-2}\boldsymbol{\Sigma}^{-1} + \mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}_*^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}_*^{-1}\mathbf{Y}$$

and covariance matrix

$$(\tau^{-2}\boldsymbol{\Sigma}^{-1} + \mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}_*^{-1}\mathbf{X})^{-1}.$$

Replacing $\boldsymbol{\beta}_0$ by its posterior expectation in (3), one obtains

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{|\sigma^2,\phi^2,\tau^2} &= (\phi^{-2}\boldsymbol{\Sigma}^{-1} + \sigma^{-2}\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1} \\ &\quad \times \{\phi^{-2}\boldsymbol{\Sigma}^{-1}(\tau^{-2}\boldsymbol{\Sigma}^{-1} + \mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}_*^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}_*^{-1}\mathbf{Y} + \sigma^{-2}\mathbf{X}^{\mathsf{T}}\mathbf{Y}\}. \qquad (4)\end{aligned}$$

Using the matrix identities (cf. Searle 1982)

$$(\tau^{-2}\boldsymbol{\Sigma}^{-1} + \mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}_*^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}_*^{-1} = \tau^2\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}}(\boldsymbol{\Sigma}_* + \tau^2\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}})^{-1}, \qquad (5)$$

$$\sigma^{-2}(\phi^{-2}\boldsymbol{\Sigma}^{-1} + \sigma^{-2}\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}} = \phi^2\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}_*^{-1}, \qquad (6)$$

one can show that Equation (4) becomes

$$\hat{\boldsymbol{\beta}}_{|\sigma^2,\phi^2,\tau^2} = \boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}_*^{-1}\{\sigma^2\tau^2(\boldsymbol{\Sigma}_* + \tau^2\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}})^{-1} + \phi^2\mathbf{I}\}\mathbf{Y}. \qquad (7)$$

Noting that $\boldsymbol{\Sigma}_* = \sigma^2\mathbf{I} + \phi^2\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}}$, Equation (7) can be simplified to

$$\hat{\boldsymbol{\beta}}_{|\sigma^2,\phi^2,\tau^2} = (\phi^2 + \tau^2)\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}}(\boldsymbol{\Sigma}_* + \tau^2\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}})^{-1}\mathbf{Y}.$$

It can also be shown that the marginal of $\mathbf{Y}$ given σ^2, ϕ^2, and τ^2 is an n-variate normal with mean 0 and covariance matrix $\boldsymbol{\Sigma}_* + \tau^2\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}}$.

Before we can proceed to the third-stage calculations, some algebraic simplifications are needed. Spectral decomposition yields $\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}} = \mathbf{H}\mathbf{D}\mathbf{H}^{\mathsf{T}}$, where $\mathbf{D} = \operatorname{diag}(d_1, d_2, \ldots, d_n)$ is the matrix of eigenvalues and $\mathbf{H}$ is the orthogonal matrix of eigenvectors. Thus

$$\begin{aligned}\boldsymbol{\Sigma}_* + \tau^2\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}} &= \mathbf{H}\{\sigma^2\mathbf{I} + (\phi^2 + \tau^2)\mathbf{D}\}\mathbf{H}^{\mathsf{T}} \\ &= u\mathbf{H}(v\mathbf{I} + \mathbf{D})\mathbf{H}^{\mathsf{T}}, \qquad (8)\end{aligned}$$

where $u = \phi^2 + \tau^2$ and $v = \sigma^2/u$.

Using the representation in Equation (8), the marginal density of $\mathbf{Y}$ given u and v can be written as

$$\begin{aligned} m(\mathbf{Y}|u, v) &= (2\pi)^{-n/2}u^{-n/2}\det(v\mathbf{I}+\mathbf{D})^{-\frac{1}{2}}\exp\left(-\frac{1}{2u}\mathbf{Y}^{\mathsf{T}}\mathbf{H}(v\mathbf{I}+\mathbf{D})^{-1}\mathbf{H}^{\mathsf{T}}\mathbf{Y}\right) \\ &= (2\pi)^{-n/2}u^{-n/2}\prod_{i=1}^{n}(v+d_i)^{-\frac{1}{2}}\exp\left(-\frac{1}{2u}\sum_{i=1}^{n}\frac{s_i^2}{v+d_i}\right), \end{aligned}$$

where $\mathbf{s} = (s_1, s_2, \ldots, s_n)^{\mathsf{T}} = \mathbf{H}^{\mathsf{T}}\mathbf{Y}$. Using the same representation, the second-stage Bayes estimator of $\boldsymbol{\beta}$ given the hyperparameters can be reexpressed as

$$\hat{\boldsymbol{\beta}}_{|\sigma^2,\phi^2,\tau^2} = \boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}}\mathbf{H}(v\mathbf{I}+\mathbf{D})^{-1}\mathbf{s}.$$

Therefore, the only numerical integration required to obtain the third-stage estimator is the posterior expectation of $1/(v+d_i)$ for $i = 1, 2, \ldots, n$ with respect to u and v.

Since Equation (2) corresponds to a polynomial regression, one may be tempted to use orthogonal polynomials to compute the Bayes estimator. Doing so, Equation (2) can be written as

$$g(t_i) = \sum_{j=0}^{m-1}\gamma_j\psi_j(t_i) + R_m(t_i), \tag{9}$$

where $\psi_j(t_i)$ is a jth-degree polynomial such that $\sum_{i=1}^{n}\psi_j(t_i)\psi_l(t_i) = 0$ for all $j \neq l$. Using the same prior model as previously defined, the hierarchical Bayes estimators of $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_{m-1})^{\mathsf{T}}$ and $\mathbf{R}_m(t) = \left(R_m(t_1), R_m(t_2), \ldots, R_m(t_n)\right)^{\mathsf{T}}$ are given by

$$\hat{\boldsymbol{\gamma}} = \mathcal{E}[v\mathbf{F}\mathbf{E}^{-1}]\mathbf{Y}, \tag{10}$$

$$\hat{\mathbf{R}}_m(t) = \mathbf{Y} - \mathcal{E}[v\mathbf{E}^{-1}]\mathbf{Q}_n^{-1}\mathbf{Y}, \tag{11}$$

where the expectation is taken with respect to the posterior of v, and where

$$\begin{aligned} \mathbf{F} &= (v\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Psi}^{\mathsf{T}}\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^{\mathsf{T}}, \\ \mathbf{E} &= v\mathbf{Q}_n^{-1} + \mathbf{I} - F^{\mathsf{T}}\boldsymbol{\Psi}^{\mathsf{T}}, \end{aligned}$$

and $\boldsymbol{\Psi}$ is an $n \times m$ matrix whose (i,j)th element corresponds to $\psi_{j-1}(t_i)$. Hence to compute the $\hat{\boldsymbol{\gamma}}$ and $\hat{\mathbf{R}}_m(t)$, an $m \times m$ and an $n \times n$ matrix have to be inverted, in contrast with the diagonalization of a $n \times n$ matrix needed in the previous setup. However, since those inversions have to be done for all v's (which is not the case for the diagonalization), it is less computationally intensive to use the original formulation given by Equation (2).

At the next step, a noninformative prior on σ^2, ϕ^2, and τ^2 of the form $\pi^*(\sigma^2/(\phi^2 + \tau^2), \phi^2 + \tau^2)$ will be used here. Recalling that $u = \phi^2 + \tau^2$ and $v = \sigma^2/u$, the prior π^* implies a prior

$$\pi(u, v) = |\mathbf{J}(u, v)|\pi^*(v, u)$$

on u and v, where $\mathbf{J}(u, v)$ is the Jacobian of the transformation. Therefore, u and v being hyperparameters, the choice of the prior on them is not crucial, since its effect on the estimator is limited (cf. Berger 1985; Goel and DeGroot 1981). Consequently, a prior of the form

$$\pi(u, v) = u^{-a}(v + v_0)^{-b}, \qquad v_0 \text{ a constant},$$

has been chosen in order to reduce the dimension of the numerical integration.

Theorem 1. *The hierarchical Bayes estimator of $\boldsymbol{\beta}$ is given by*

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\Sigma}\mathbf{X}^{\mathsf{T}}\mathbf{H}\mathcal{E}[(v\mathbf{I}+\mathbf{D})^{-1}]\mathbf{s}, \tag{12}$$

where the expectation is taken with respect to

$$\pi(v|\mathbf{Y}) \propto (v+v_0)^{-b}\left(\prod_{i=1}^{n}(v+d_i)\right)^{-\frac{1}{2}}\left(\sum_{i=1}^{n}\frac{s_i^2}{v+d_i}\right)^{-\{(n/2)+a\}}. \tag{13}$$

Proof. To compute $\hat{\boldsymbol{\beta}}$, the expectation of $(v+d_j)^{-1}$ for $j = 1, \ldots, n$ with respect to $\pi(u, v|\mathbf{Y})$ is required. Note that

$$\begin{aligned}\mathcal{E}[(v+d_j)^{-1}] &= \int_0^\infty\int_0^\infty (v+d_j)^{-1}\pi(u, v|\mathbf{Y})\,du\,dv\\ &= \int_0^\infty (v+d_j)^{-1}\left(\int_0^\infty \pi(u, v|\mathbf{Y})\,du\right)dv.\end{aligned}$$

It can be shown that

$$\begin{aligned}\int_0^\infty \pi(u, v|\mathbf{Y})\,du &\propto \int_0^\infty u^{-\{(n/2)+a+1\}}(v+v_0)^{-b}\left[\prod_{i=1}^{n}(v+d_i)\right]^{-\frac{1}{2}}\\ &\qquad\times\exp\left(-\frac{1}{2u}\sum_{i=1}^{n}\frac{s_i^2}{v+d_i}\right)du\\ &= (v+v_0)^{-b}\left(\prod_{i=1}^{n}(v+d_i)\right)^{-\frac{1}{2}}\\ &\qquad\times\int_0^\infty u^{-\{(n/2)+a+1\}}\exp\left(-\frac{1}{2u}\sum_{i=1}^{n}\frac{s_i^2}{v+d_i}\right)du\\ &\propto (v+v_0)^{-b}\left(\prod_{i=1}^{n}(v+d_i)\right)^{-\frac{1}{2}}\left(\sum_{i=1}^{n}\frac{s_i^2}{v+d_i}\right)^{-\{(n/2)+a\}},\end{aligned}$$

which completes the proof. Q.E.D.

Remark. All the above integrals remain finite provided that $b > a+1$.

Corollary 1. *The hierarchical Bayes estimator of* $\left(g(t_1), g(t_2), \ldots, g(t_n)\right)^{\mathsf{T}} = \mathbf{X}\boldsymbol{\beta}$ *is* $\mathbf{X}\hat{\boldsymbol{\beta}}$.

Corollary 2. *The hierarchical Bayes estimator of $g(t)$ for any $t \in \mathcal{T}$ is*

$$\hat{g}(t) = \left(\boldsymbol{\Phi}(t)^{\mathsf{T}}, \mathbf{q}(t)^{\mathsf{T}}\mathbf{Q}_n^{-1}\right)\hat{\boldsymbol{\beta}}, \tag{14}$$

where

$$\begin{aligned}\mathbf{q}(t) = \frac{1}{\{(m-1)!\}^2}\Big(&\{(t-t_0)(t_1-t_0)\}^{m-1}\rho(|t-t_1|),\\ &\{(t-t_0)(t_2-t_0)\}^{m-1}\rho(|t-t_2|), \ldots, \{(t-t_0)(t_n-t_0)\}^{m-1}\rho(|t-t_n|)\Big)^{\mathsf{T}}.\end{aligned}$$

Proof. From Equation (2),

$$\hat{g}(t) = \mathbf{\Phi}(t)^{\mathsf{T}}\boldsymbol{\theta} + \hat{R}_m(t). \tag{15}$$

Since $\hat{\boldsymbol{\theta}}$ is known (first m coordinates of $\hat{\boldsymbol{\beta}}$), all that is left to be done is to compute $\hat{R}_m(t) = \mathcal{E}[R_m(t)|\mathbf{Y}]$. The joint distribution of $R_m(t)$ and $\mathbf{Y}$ given u and v is an $(n+1)$-variate normal with mean 0 and covariance matrix

$$u\begin{pmatrix} 1 & \mathbf{q}(t)^{\mathsf{T}} \\ \mathbf{q}(t) & \mathbf{H}(v\mathbf{I}+\mathbf{D})\mathbf{H}^{\mathsf{T}} \end{pmatrix}.$$

Consequently the expectation of $R_m(t)$ given $\mathbf{Y}$, u, and v is

$$\hat{R}_m(t)_{u,v} = \mathbf{q}(t)^{\mathsf{T}}\mathbf{H}(v\mathbf{I}+\mathbf{D})^{-1}\mathbf{H}^{\mathsf{T}}\mathbf{Y}.$$

Therefore,

$$\hat{R}_m(t) = \mathbf{q}(t)^{\mathsf{T}}\mathbf{H}\mathcal{E}[(v\mathbf{I}+\mathbf{D})^{-1}]\mathbf{s}.$$

Thus replacing $\hat{R}_m(t)$ in Equation (15) by the last expression, we get

$$\begin{aligned}\hat{g}(t) &= \mathbf{\Phi}(t)^{\mathsf{T}}\hat{\boldsymbol{\theta}} + \mathbf{q}(t)^{\mathsf{T}}\mathbf{H}\mathcal{E}[(v\mathbf{I}+\mathbf{D})^{-1}]\mathbf{s} \\ &= (\mathbf{\Phi}(t)^{\mathsf{T}}, \mathbf{0}^{\mathsf{T}})\boldsymbol{\beta} + (\mathbf{0}^{\mathsf{T}}, \mathbf{q}(t)^{\mathsf{T}}\mathbf{Q}_n^{-1})\begin{pmatrix} \mathbf{\Gamma}\,\mathbf{T}^{\mathsf{T}} \\ \mathbf{Q}_n \end{pmatrix}\mathbf{H}\mathcal{E}[(v\mathbf{I}+\mathbf{D})^{-1}]\mathbf{s}.\end{aligned}$$

Noting that $(\mathbf{\Gamma}\,\mathbf{T}^{\mathsf{T}}, \mathbf{Q}_n^{\mathsf{T}}) = \mathbf{X}\mathbf{\Sigma}$ and using Equation (12), we have

$$\begin{aligned}\hat{g}(t) &= (\mathbf{\Phi}(t)^{\mathsf{T}}, \mathbf{0}^{\mathsf{T}})\hat{\boldsymbol{\beta}} + (\mathbf{0}^{\mathsf{T}}, \mathbf{q}(t)^{\mathsf{T}}\mathbf{Q}_n^{-1})\hat{\boldsymbol{\beta}} \\ &= (\mathbf{\Phi}(t)^{\mathsf{T}}, \mathbf{q}(t)^{\mathsf{T}}\mathbf{Q}_n^{-1})\hat{\boldsymbol{\beta}},\end{aligned}$$

which is the desired result. Q.E.D.

2.3. Construction of the Error Band.

The error band that we propose for $g(t)$ is of the form

$$\hat{g}(t) \pm 2\sqrt{v(t)}, \tag{16}$$

where $v(t)$ is the posterior variance of $g(t)$. As it will be shown in Corollary 4, $v(t)$ depends on the posterior covariance matrix of $\boldsymbol{\beta}$, which is given in the following theorem.

THEOREM 2. *The posterior covariance matrix of* $\boldsymbol{\beta}$ *is*

$$\begin{aligned}\mathcal{D}(\boldsymbol{\beta}) = \mathcal{E}[u]\mathbf{\Sigma} - \mathbf{\Sigma}\mathbf{X}^{\mathsf{T}}\mathbf{H}\mathcal{E}[u(v\mathbf{I}+\mathbf{D})^{-1}]\mathbf{H}^{\mathsf{T}}\mathbf{X}\mathbf{\Sigma} \\ + \mathcal{E}[\hat{\boldsymbol{\beta}}_{|\sigma^2,\phi^2,\tau^2}\hat{\boldsymbol{\beta}}^{\mathsf{T}}_{|\sigma^2,\phi^2,\tau^2}] - \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{\mathsf{T}},\end{aligned} \tag{17}$$

where the expectation is taken with respect to the posterior distribution of u and v.

Proof. The proof of Theorem 2 is immediate from the identity

$$\mathcal{E}[(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^{\mathsf{T}}] = \mathcal{E}^{\sigma^2,\phi^2,\tau^2|\mathbf{Y}}[\mathcal{E}^{\boldsymbol{\beta}_0|\mathbf{Y},\sigma^2,\phi^2,\tau^2}[\mathcal{E}^{\boldsymbol{\beta}|\mathbf{Y},\sigma^2,\phi^2,\tau^2,\boldsymbol{\beta}_0}[(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^{\mathsf{T}}]]],$$

and from the matrix identities (5) and (6). Q.E.D.

Corollary 3. *The posterior covariance matrix of* $(g(t_1), g(t_2), \ldots, g(t_n))^{\mathsf T} = \mathbf{X}\boldsymbol{\beta}$ *is*

$$\mathcal{D}\begin{pmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{pmatrix} = \mathcal{E}[u]\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf T} - \mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf T}\mathbf{H}\mathcal{E}[u(v\mathbf{I}+\mathbf{D})^{-1}]\mathbf{H}^{\mathsf T}\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf T}$$

$$+\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf T}\mathbf{H}\{\mathcal{E}[(v\mathbf{I}+\mathbf{D})^{-1}\mathbf{s}\mathbf{s}^{\mathsf T}(v\mathbf{I}+\mathbf{D})^{-1}]$$

$$-\mathcal{E}[(v\mathbf{I}+\mathbf{D})^{-1}]\mathbf{s}\mathbf{s}^{\mathsf T}\mathcal{E}[(v\mathbf{I}+\mathbf{D})^{-1}]\}\mathbf{H}^{\mathsf T}\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf T},$$

where the expectation is taken with respect to the posterior of u and v.

Proof. The proof is straightforward from Theorem 2, since $(g(t_1), g(t_2), \ldots, g(t_n))^{\mathsf T} = \mathbf{X}\boldsymbol{\beta}$. Q.E.D.

Corollary 4. *The posterior variance of* $\hat{g}(t)$ *is*

$$v(t) = (\boldsymbol{\Phi}(t)^{\mathsf T}\boldsymbol{\Gamma}, \mathbf{q}(t)^{\mathsf T})\{\mathcal{E}[u]\boldsymbol{\Sigma}^{-1} - \mathbf{X}^{\mathsf T}\mathbf{H}\mathcal{E}[u(v\mathbf{I}+\mathbf{D})^{-1}]\mathbf{H}^{\mathsf T}\mathbf{X}$$

$$+\boldsymbol{\Sigma}^{-1}(\mathcal{E}[\hat{\boldsymbol{\beta}}_{\sigma^2,\phi^2,\tau^2}\hat{\boldsymbol{\beta}}^{\mathsf T}_{|\sigma^2,\phi^2,\tau^2}] - \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{\mathsf T})\boldsymbol{\Sigma}^{-1}\}\begin{pmatrix} \boldsymbol{\Gamma}\boldsymbol{\Phi}(t) \\ q(t) \end{pmatrix}. \quad (18)$$

Proof. From Corollary 2, we have

$$\hat{g}(t) = (\boldsymbol{\Phi}(t)^{\mathsf T}, \mathbf{q}(t)^{\mathsf T}\mathbf{Q}_n^{-1})\hat{\boldsymbol{\beta}}$$

$$= (\boldsymbol{\Phi}(t)^{\mathsf T}\boldsymbol{\Gamma}, \mathbf{q}(t)^{\mathsf T})\begin{pmatrix} \boldsymbol{\Gamma}^{-1} & 0 \\ \mathbf{0} & \mathbf{Q}_n^{-1} \end{pmatrix}\hat{\boldsymbol{\beta}}$$

$$= (\boldsymbol{\Phi}(t)^{\mathsf T}\boldsymbol{\Gamma}, \mathbf{q}(t)^{\mathsf T})\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\beta}}. \quad (19)$$

The calculation of $v(t)$ is now immediate from Theorem 2 and Equation (19). Q.E.D.

To obtain an error band for $g(t)$ for $t \in \mathcal{T}$, one computes $\hat{g}(t)$ [Equation (14)] and $v(t)$ [Equation (18)] and substitutes these values in Equation (16). Note that the numerical integrations required to evaluate Equations (14) and (18) do not depend on t but depend only on $t_1, t_2, \ldots, t_n$ through the calculation of $\hat{\boldsymbol{\beta}}$ [Equation (12)] and $\mathcal{D}(\boldsymbol{\beta})$ [Equation (17)].

2.4. Choice of the Smoothness Parameter m.

The smoothness parameter m corresponds to the number of unknown regression parameters in Equation (2). Thus m itself should be considered a hyperparameter. Therefore, the classical Bayesian approach to this problem is to assign a prior on m and to conduct the full Bayesian analysis on the complete model. This is equivalent to computing the posterior expectation of $\hat{\boldsymbol{\beta}}$ [refer to Equation (12)] with respect to m.

Although this involves only another stage of numerical integration, the amount of calculation required will be enormous. This is because the first-stage covariance matrix $\boldsymbol{\Sigma}$ and the design matrix $\mathbf{X} = (\mathbf{T}, \mathbf{I})$ depend on m through $\mathbf{Q}_n$ and $\mathbf{T}$ respectively. This implies that the spectral decomposition of $\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^{\mathsf T} = \mathbf{H}\mathbf{D}\mathbf{H}^{\mathsf T}$ will have to be done for each m separately. Therefore, we shall use a simpler alternative approach.

Our approach involves Bayesian testing to choose the most appropriate value of m. One way of doing Bayesian testing is to choose that m which maximizes the posterior probability of m given $\mathbf{Y}$. (Equivalently, one could compute the Bayes factor of the hypothesis $\mathcal{H}_0 : m = m_0$ vs. $\mathcal{H}_1 : m \neq m_0$, for each m_0.) Note that comparing the posterior probability of m given y is equivalent to comparing $\text{marg}(\mathbf{Y}|m) \times \pi(m)$, where $\text{marg}(\mathbf{Y}|m)$ represents the marginal density of $\mathbf{Y}$ given m, and $\pi(m)$ is the prior mass function of m on the set of positive integers. This approach is equivalent to finding the ML-II estimate of m, which is a standard method of choosing some of the hyperparameters in hierarchical Bayes analysis.

The problem of choosing m can also be viewed as a model selection problem (cf. Rao and Wu 1989). If one knows the largest value m can take, denoted M, then choosing the optimal value of m is equivalent to deciding among the hypotheses

$$\mathcal{H}_j : \{\theta_i \neq 0,\ i = 1, \ldots, j;\ \theta_i = 0,\ i = j+1, \ldots, M.\}$$

(Note that accepting the hypothesis $\mathcal{H}_j$ corresponds to decide that $m = j + 1$.) Since it is assumed that M is known, one can, for example, use a uniform prior on m $[\pi(m) = (1/M)\mathbf{I}_{(1\leq m\leq M)}(m)]$ and carry out a proper Bayesian analysis of the problem.

2.5. Case of Non-I.I.D. Errors.

In this subsection, ϵ will be assumed an n-variate normal with mean 0 and covariance matrix $\sigma^2\mathbf{R}$, where $\mathbf{R}$ is a known positive definite matrix.

Straightforward calculation leads to the first-stage estimator given by

$$\hat{\beta}_{|\beta_0,\sigma^2,\phi^2,\tau^2} = (\phi^2\boldsymbol{\Sigma}^{\ 1} + \sigma^{\ 2}\mathbf{X}^\mathsf{T}\mathbf{R}^{-1}\mathbf{X})^{-1}(\phi^2\boldsymbol{\Sigma}^{-1}\beta_0 + \sigma^{-2}\mathbf{X}^\mathsf{T}\mathbf{R}\mathbf{Y}),$$

and shows that the first-stage marginal is normal with mean $\mathbf{X}\beta_0$ and covariance matrix $\boldsymbol{\Delta} = \sigma^2\mathbf{R}+\phi^2\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^\mathsf{T}$. Using matrix identities as in Equations (5) and (6), the second-stage estimator can be written as

$$\begin{aligned}\hat{\beta}_{|\sigma^2,\phi^2,\tau^2} &= (\phi^2+\tau^2)\boldsymbol{\Sigma}\mathbf{X}^\mathsf{T}\{\sigma^2\mathbf{R} + (\phi^2+\tau^2)\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^\mathsf{T}\}^{\ 1}\mathbf{Y}\\ &= \boldsymbol{\Sigma}\mathbf{X}^\mathsf{T}\mathbf{R}^{-\frac{1}{2}}(v\mathbf{I} + \mathbf{R}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^\mathsf{T}\mathbf{R}^{-\frac{1}{2}})^{\ 1}\mathbf{R}^{-\frac{1}{2}}\mathbf{Y}.\end{aligned}$$

The marginal of $\mathbf{Y}$ given σ^2, ϕ^2, and τ^2 will be n-variate normal with mean $\mathbf{0}$ and covariance matrix $u[v\mathbf{R} + \mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^\mathsf{T}]$.

Consequently the hierarchical Bayes estimator will be given by

$$\hat{\beta} = \boldsymbol{\Sigma}\mathbf{X}^\mathsf{T}\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathcal{E}[(v\mathbf{I} + \mathbf{D})^{\ 1}]\mathbf{H}^\mathsf{T}\mathbf{R}^{-\frac{1}{2}}\mathbf{Y}, \tag{20}$$

where $\mathbf{H}\mathbf{D}\mathbf{H}^\mathsf{T} = \mathbf{R}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\Sigma}\mathbf{X}^\mathsf{T}\mathbf{R}^{-\frac{1}{2}}$. Note that Equation (20) is identical to Equation (12), so that Theorem 1 and Corollaries 2 and 3 apply directly.

3. ILLUSTRATIVE EXAMPLES AND SENSITIVITY STUDY

In this section, our result will be illustrated by means of examples. We have chosen the covariance kernel $\rho(x) = e^{-cx}$. First, we shall apply our procedure to a data set consisting of observations from a discrete time series. To be specific, the data used in this example comes from Ma and Zidek (1988), and they represent the monthly precipitation (rain plus one-tenth snow) in inches from March 1965 to December 1966 in Vancouver (Canada). As indicated in the introduction, here $g(t)$ represents the underlying intensity of precipitation.

and not just the true rainfall (as in a measurement-error model). Therefore a model such as (2) is meaningful and leads to our methodology. Also note that this is, therefore, an appealing alternative to the computationally intensive approach of Kitagawa and Gersch (1984). Figure 1 displays the graph of $\hat{g}(t)$ $(a = 0, c = 1, m = 2, \nu_0 = 1)$ along with the error bands and the cross-validated smoothing spline (cf. Wahba 1983). The hyperparameters were chosen after sensitivity analyses were conducted as described below for the simulated data set. Note that $\hat{g}(\cdot)$ is very close to the smoothing spline. Being a cubic spline, the smoothing spline is smoother, but $\hat{g}(\cdot)$ is not as smooth, since it includes the error $Z(t)$ in the model. It can also be seen that the smoothing spline lies within the error band.

We shall now describe sensitivity analyses and the proper choice of hyperparameters using a simulated data set. We generated $n = 20$ observations using the model $y_i = g(t_i) + \epsilon_i$, where $g(t) = 10t^3 \log(1 + t)$, $t_i = i/n$, $1 < i \leq n$, and ϵ_i are independent observations from $N(0, 1)$. All the calculations have been done with $c = 1$, unless specified otherwise. Figure 2 shows the regression function g, the generated data, and the estimate $\hat{g}$ that we obtained with $a = 0$, $m = 3$, and $\nu_0 = 1$.

A sensitivity analysis has been conducted for m, and the results are displayed in Figure 3. Note that a and ν_0 are fixed at $a = 0$, $\nu_0 = 1$. As expected, large values of m imply larger smoothness for $\hat{g}(\cdot)$. When $m = 1$, the procedure tries to fit a constant plus a Gaussian process; $m = 2$ also does not provide enough smoothing; $\text{marg}(Y|m)$ is maximized at $m = 3$; but the graph corresponding to $m = 4$ is only slightly more smooth.

Figure 4 shows graphs of $\hat{g}(t)$ for values of $a = 0, \frac{1}{2}$, and 1, for fixed values of m $(m = 3)$ and ν_0 $(\nu_0 = 1)$. Note that b has been chosen to be $a + 1.05$. Generally the smoothness of $\hat{g}(t)$ increases with a, but in this example the sensitivity to the choice of a is insignificant.

Figure 5 shows the behaviour of $\hat{g}(\cdot)$ when ν_0 varies from 0.5 to 5 [vbar = $(\text{tr}\, D)/n$, which is approximately 1.5] while a and m are fixed $(a = 0, m = 2)$. The behaviour of $\hat{g}(\cdot)$ is similar to that in Figure 4, that is, a large value of ν_0 generally implies more smoothness.

It can be seen from Figure 6 that the value of c is important. The shape of the curves is similar for all values of c, but the larger c is, the closer $\hat{g}(\cdot)$ is to the $y(t_i)$'s. The choice of c, therefore, depends on whether one is interested in curve fitting or curve smoothing.

Computer programs written in S-language and FORTRAN which generate all the figures except Figure 1 are available from the authors upon request.

ACKNOWLEDGEMENT

We would like to thank Alan Douglas for helping us with some numerical computations. We are very grateful to the referees and an associate editor for some very valuable suggestions and references.

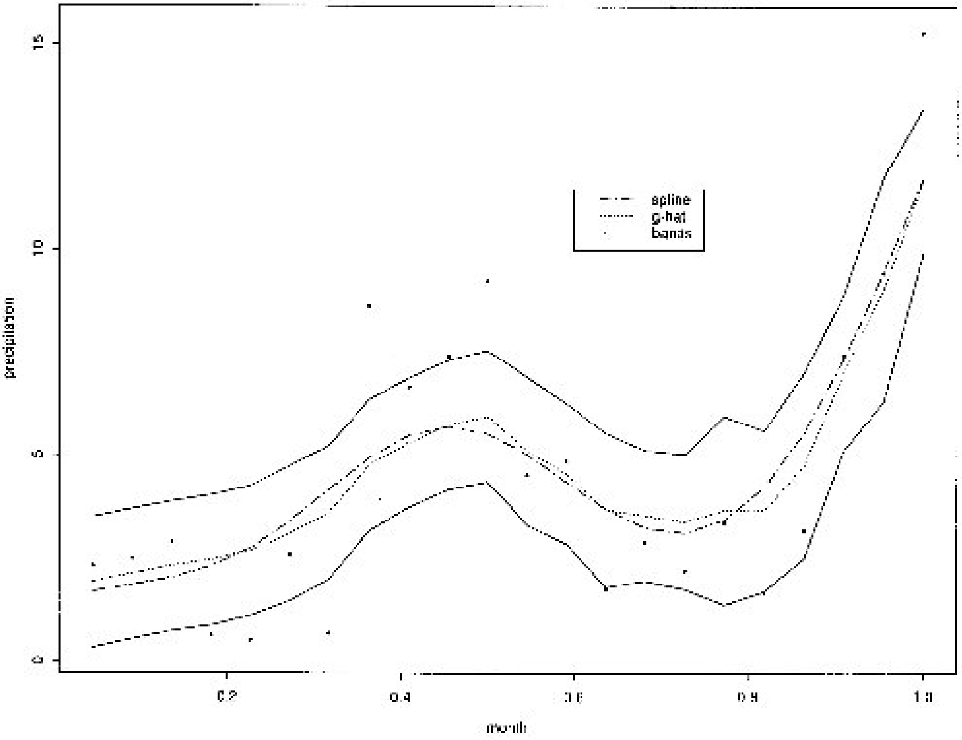

FIGURE 1: Classical spline, $\hat{g}$, and error bands for $\hat{g}$.

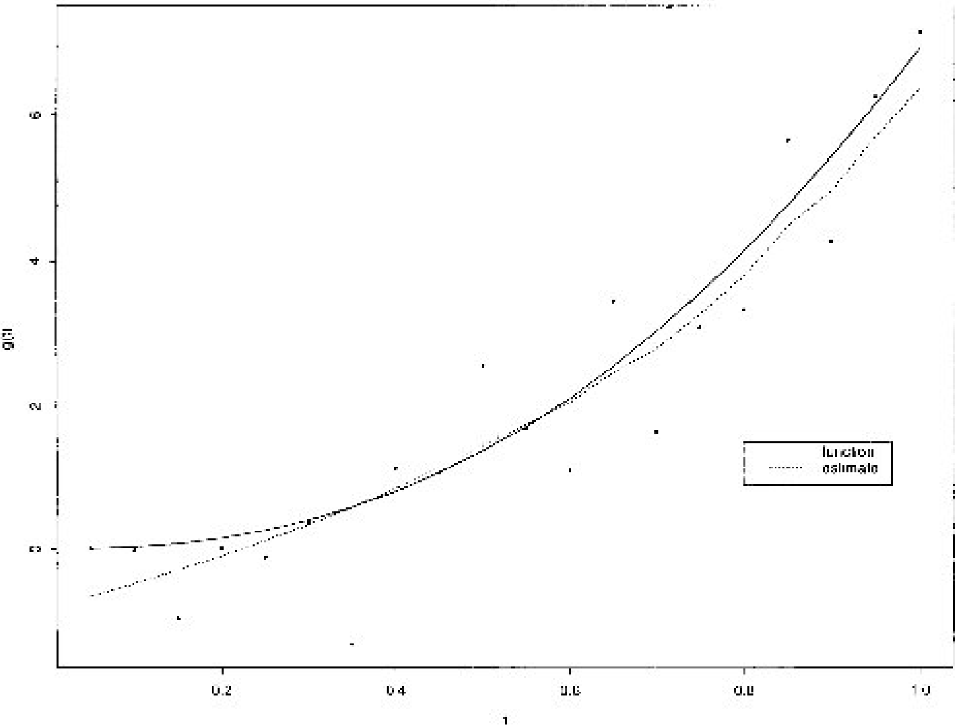

FIGURE 2: Regression function, data, and estimate.

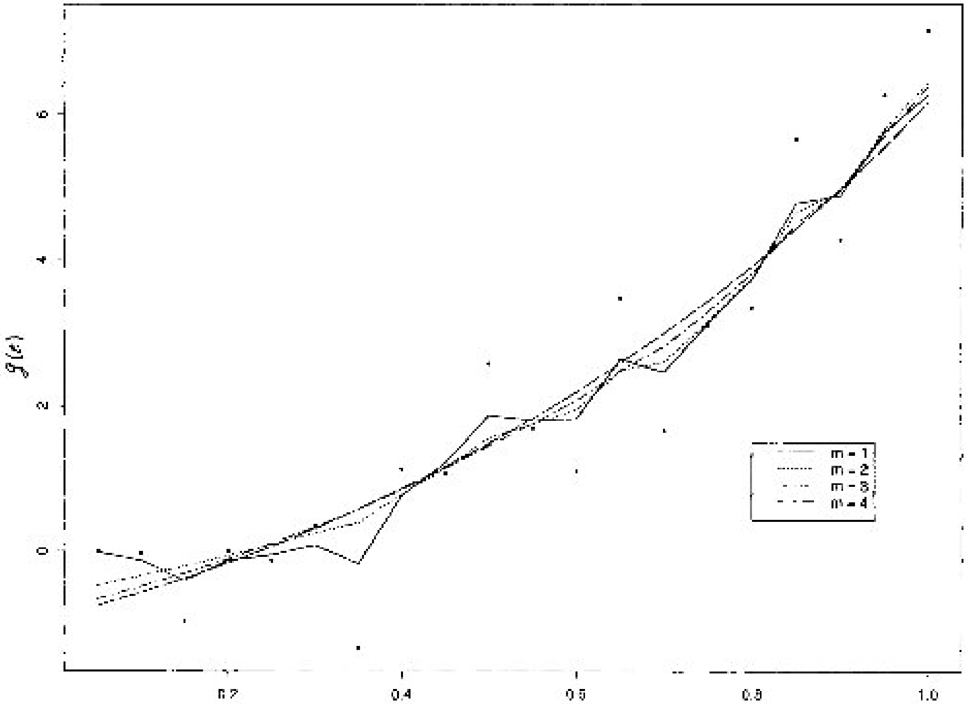

FIGURE 3: Sensitivity to the choice of *m*.

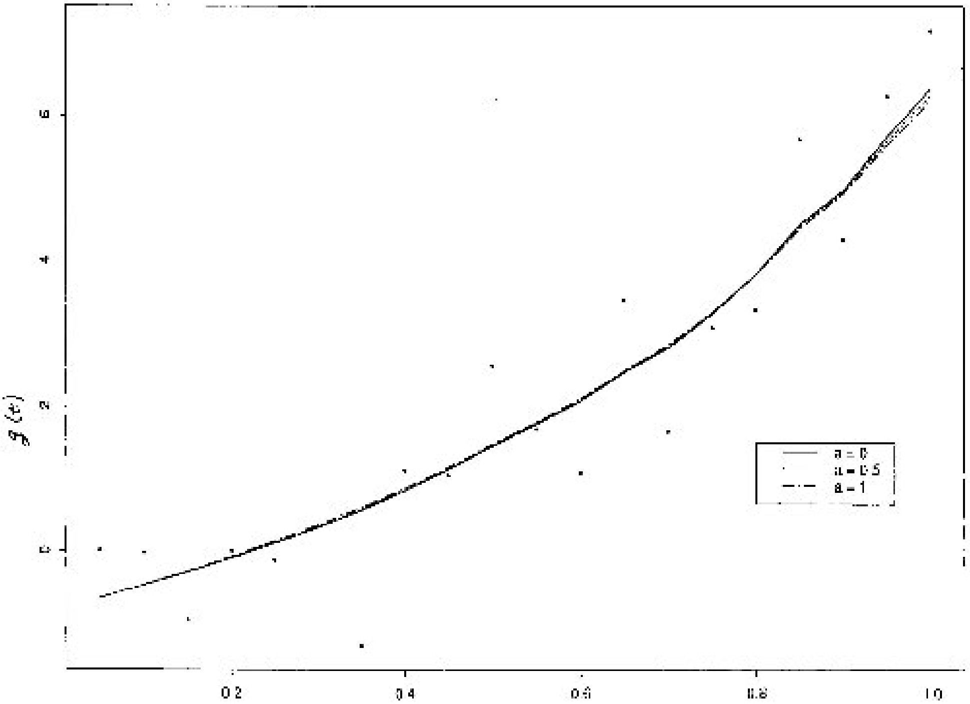

FIGURE 4: Sensitivity to the choice of *a*.

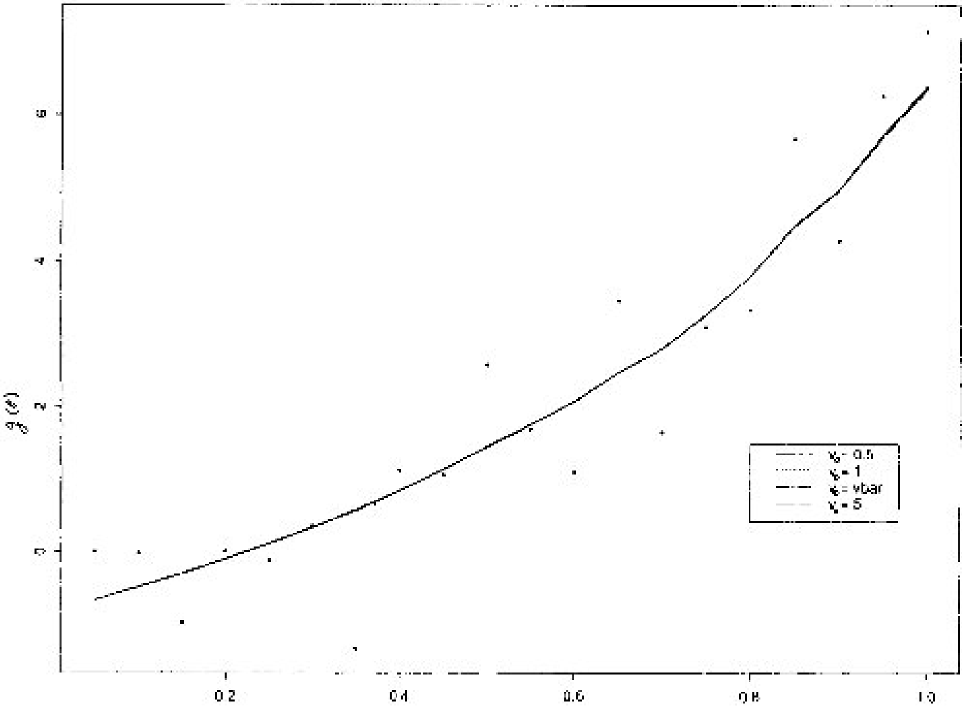

FIGURE 5: Sensitivity to the choice of v_0.

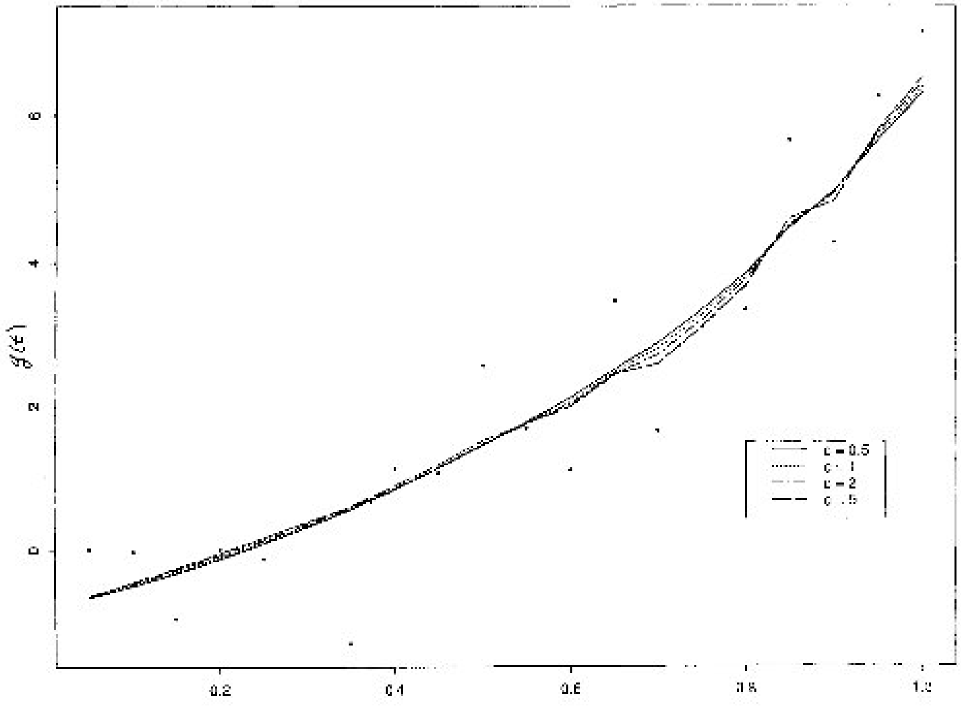

FIGURE 6: Sensitivity to the choice of c.

REFERENCES

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Delampady, M. (1987). Estimating a monotone regression function. Technical Report 54, Department of Statistics, University of British Columbia, Vancouver.

Goel, P.K., and DeGroot, M.H. (1981). Information about hyperparameter in hierarchical model. *J. Amer. Statist. Assoc.*, 76, 140–147.

Kitagawa, G., and Gersch, W. (1984). A smoothness priors–state space modeling of time series with trend and seasonality. *J. Amer. Statist. Assoc.*, 79, 378–389.

Lindley, D.V., and Smith, A.F.M. (1972). Bayes estimates for the linear model. *J. Roy. Statist. Soc. Ser. B*, 34, 1–41.

Linthurst, R.A., Landers, D.H., Eilers, J.M., Brakke, D.F., Overton, W.S., Meier, E.P., and Crowe, R.E. (1986). *Characteristics of Lakes in the Eastern United States, Volume I. Population Descriptions and Physico-chemical Relationships*. EPA/600/4-86/007a, U.S. Environmental Protection Agency, Washington.

Ma, P.H., and Zidek, J.V. (1988). Data for statistics research and instruction. Technical Report. Department of Statistics, University of British Columbia, Vancouver.

Nebebe, F., and Stroud, T.W.F. (1986). Bayes and empirical Bayes shrinkage estimation of regression coefficients. *Canad. J. Statist.*, 14, 267–280.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *J. Roy. Statist. Soc. Ser. B*, 40, 1–42.

Rao, C.R., and Wu, Y. (1989). A strongly consistent procedure for model selection in a regression problem. *Biometrika*, 76(2), 369–374.

Searle, S.R. (1982). *Matrix Algebra Useful for Statistics*. Wiley, New York.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B*, 40, 364–372.

Wahba, G. (1983). Bayesian "confidence intervals" for cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B*, 45, 113–150.

Weerahandi, S., and Zidek, J.V. (1988). Bayesian nonparametric smoothers. *Canad. J. Statist.*, 16, 61–74.

Received 20 December 1989
Revised 18 January 1991
Accepted 16 May 1991

Département de mathématiques et de statistique
Université de Montréal
C.P. 6128 Succ. A
Montréal, Québec H3C 3J7

Indian Statistical Institute
8th Mile Road, Mysore Road
RV College Post
Bangalore 560059
India