Biometrika (1997), **84**, 1, *pp.* 85–93
Printed in Great Britain

Hierarchical Bayesian semiparametric procedures for logistic regression

BY JOHN S. J. HSU

Department of Statistics and Applied Probability, University of California, Santa Barbara, California 93106, U.S.A.
e-mail: hsu@pstat.ucsb.edu

AND TOM LEONARD

Department of Mathematics and Statistics, University of Edinburgh, Edinburgh EH9 3JZ, U.K.
e-mail: leonard@maths.ed.ac.uk

SUMMARY

A simple procedure is proposed for exact computation to smooth Bayesian estimates for logistic regression functions, when these are not constrained to lie on a fitted regression surface. Exact finite sample inferences and predictions are available, together with an exact residual analysis. The prior distribution relates to O'Hagan's assumptions for a normal regression function. A global shrinkage parameter and local smoothness parameter can be evaluated from the current data by hierarchical Bayesian procedures. Consideration of the shrinkage parameter permits an overall check regarding a hypothesised regression model. No optimisation technique is needed, since Monte Carlo simulations from independent logistic distributions can be directly employed. The complexity of the computations does not substantively increase with the dimensionality of the design space.

Some key words: Bayesian smoothing; Gaussian process; Hierarchical Bayes; Logistic distribution; Logistic regression; Monte Carlo; Semiparametric.

1. LOGISTIC REGRESSION FUNCTIONS

Consider frequencies $y_1, \ldots, y_m$ which, given $\theta_1, \ldots, \theta_m$, possess independent binomial distributions with respective probabilities, $\theta_1, \ldots, \theta_m$, and sample sizes, $n_1, \ldots, n_m$. Suppose that the logits $\alpha_i = \text{logit}\,(\theta_i) = \log \theta_i - \log(1-\theta_i)$ satisfy

$$\alpha_i = \alpha(t_i) \quad (i = 1, \ldots, m), \tag{1·1}$$

where $t_1, \ldots, t_m$ are specified design vectors in a design space D, and the real-valued regression function $\alpha(t)$ is defined for all $t \in D$.

Also consider a linear hypothesis for $\alpha(t)$, and taking the form

$$H_0 : \alpha(t) \equiv \alpha_0(t) \equiv \phi'(t)\beta, \tag{1·2}$$

where β is a $p \times 1$ vector of unknown parameters, and $\phi(t) = (\phi_1(t), \ldots, \phi_p(t))'$ is a specified vector of basis functions. Estimates will be developed for $\alpha(t)$, and for

$$\theta(t) := \frac{\exp\{\alpha(t)\}}{1+\exp\{\alpha(t)\}}.$$

The estimates for $\alpha(t)$ and $\theta(t)$ have the following properties.

(a) They are not constrained to lie on a fitted surface, but compromise between perfectly fitting the data, and a surface fitted via a linear hypothesis. This is the concept of 'global shrinkage' towards a hypothesised model. See, for example, Leonard (1978).

(b) They are nevertheless infinitely differentiable, and sufficiently regular to be able to predict $\theta(t)$ at further points in the design space D. This is the concept of 'local smoothness' and will be created by the infinitely differentiable covariance kernel introduced below.

Generalising Leonard (1972), we develop a hierarchical Bayesian methodology for achieving objectives (a) and (b). Following the approaches of Silverman (1978) and Leonard (1978) to density estimation, and of Blight & Ott (1975) and O'Hagan (1978) to the smoothing of regression functions, it is assumed, in the prior assessment, that, given β and a covariance kernel C, $\alpha(t)$ follows a Gaussian process with mean value function $\alpha_0(t)$ satisfying (1·2) and covariance kernel $C(s, t)$, defined as $\sigma^2 K(s, t)$, for $s, t \in D$, where

$$K(s,t) = \exp\{-\gamma(s-t)'A(s-t)\}, \tag{1·3}$$

with A a positive definite matrix. For simplicity, β is assumed to be uniformly distributed over R^p. The global shrinkage parameter σ^2, and local smoothness parameter γ, can be evaluated from the current data set via hierarchical Bayesian procedures.

We develop an exact analysis of the above model, using Monte Carlo techniques, without using any iterative procedures for maximising likelihoods or posterior densities. An exact residual analysis is also developed for investigating the linear hypothesis (1·2), providing an alternative to approximate analyses recommended by Cox & Snell (1968), Cox (1970) and Cook (1977). Furthermore, the marginal posterior density of σ^2 facilitates an overall check of the validity of the hypothesised model.

O'Sullivan, Yandell & Raynor (1986) and Gu (1990) discuss elegant methods for smoothing $\alpha(t)$ based upon splines, involving optimisation techniques for Bayesian posterior modes, but do not attempt to deal with the concepts in (a) and (b) separately, since their choice of null space is defined by their choices of covariance kernel, or to develop finite sample inference procedures. They are however able to provide useful point estimates, where a single smoothing parameter is chosen by generalised cross-validation.

2. BAYESIAN ANALYSIS

All distributions in this section are conditional on σ^2 and γ. For any $t \in D$, the conditional distribution of $\alpha(t)$, given $\tilde{\alpha}_m = (\alpha_1, \ldots, \alpha_m)'$ and β, is the same as the corresponding conditional prior distribution, that is, normal with mean

$$\alpha^*(t) = \phi'(t)\beta + e^*(t), \tag{2·1}$$

and variance $\sigma^2\omega^*(t)$, where

$$\omega^*(t) = K(t,t) - \kappa'(t)K^{-1}\kappa(t), \tag{2·2}$$

$$e^*(t) = \kappa'(t)K^{-1}(\tilde{\alpha}_m - X\beta), \tag{2·3}$$

$$\kappa(t) = (K(t,t_1), \ldots, K(t,t_m))'. \tag{2·4}$$

Here K is the $m \times m$ matrix with (i, j)th element equal to $K(t_i, t_j)$, and X is the $m \times p$ matrix with (i, j)th element equal to $\phi_j(t_i)$. The corresponding conditional distribution of the 'parametric residual'

$$e(t) := \alpha(t) - \phi'(t)\beta \tag{2·5}$$

is normal with mean $e^*(t)$ and variance $\sigma^2\omega^*(t)$.

Furthermore, the conditional posterior distribution of β, given $\tilde{\alpha}_m$, is multivariate normal with mean vector

$$\beta^*(\tilde{\alpha}_m) := (X'K^{-1}X)^{-1}X'K^{-1}\tilde{\alpha}_m, \tag{2·6}$$

and covariance matrix $\sigma^2 D$, where

$$D = (X'K^{-1}X)^{-1}, \tag{2·7}$$

so that the posterior distibution of $\alpha(t)$, given $\tilde{\alpha}_m$, is normal, with mean

$$\tilde{\alpha}(t) := \{b'(t) + \kappa'(t)Q\}\tilde{\alpha}_m, \tag{2·8}$$

and variance $\sigma^2\hat{\omega}(t)$, where

$$\hat{\omega}(t) = \omega^*(t) + \{\phi'(t) - \kappa'(t)K^{-1}X\}D\{\phi(t) - X'K^{-1}\kappa(t)\}, \tag{2·9}$$

$$b'(t) = \phi'(t)DX'K^{-1}, \tag{2·10}$$

$$Q = K^{-1} - K^{-1}X(X'K^{-1}X)^{-1}X'K^{-1}. \tag{2·11}$$

Note that the posterior density of $\tilde{\alpha}_m$, unconditional on β, is

$$\pi(\tilde{\alpha}_m \mid y) \propto l(\tilde{\alpha}_m \mid y)\exp\{-\tilde{\alpha}_m' Q\tilde{\alpha}_m/(2\sigma^2)\}, \tag{2·12}$$

where

$$l(\tilde{\alpha}_m \mid y) \propto \exp\left\{\tilde{\alpha}_m' y - \sum_{i=1}^{m} n_i \log(1 + e^{\alpha_i})\right\}. \tag{2·13}$$

The likelihood (2·13) is proportional to the density of a proper distribution. Since the second contribution to the product on the right-hand side of (2·12) is bounded above by unity, the density (2·12) will remain proper. All posterior moments, given σ^2 and γ, if they exist, and probabilities for $\alpha(t)$ and elements of β, may be computed by integrating the corresponding quantities, when $\tilde{\alpha}_m$ is known, with respect to the posterior density in (2·12). These integrations can be performed exactly, using the Monte Carlo techniques of § 3. Similar techniques can be applied to the parametric residual $e(t)$, in (2·5), since, given $\tilde{\alpha}_m$, this possesses a normal posterior density with mean

$$\hat{e}(t) := \kappa'(t)Q\tilde{\alpha}_m, \tag{2·14}$$

and variance $\sigma^2\hat{\omega}^*(t)$, where

$$\hat{\omega}^*(t) = K(t, t) - \kappa'(t)Q\kappa(t). \tag{2·15}$$

An integrated likelihood may be obtained for σ^2 and γ by integrating out β from the joint distribution of y, $\tilde{\alpha}_m$ and β, giving the joint likelihood

$$l(\sigma^2, \gamma, \tilde{\alpha}_m \mid y) \propto (\sigma^2)^{-\frac{1}{2}(m-p)}|K|^{-\frac{1}{2}}|X'K^{-1}X|^{-\frac{1}{2}} l(\tilde{\alpha}_m \mid y)\exp\{-\tilde{\alpha}_m' Q\tilde{\alpha}_m/(2\sigma^2)\}. \tag{2·16}$$

A further integration with respect to $\tilde{\alpha}_m$ yields the required integrated likelihood.

3. HIERARCHICAL BAYESIAN PROCEDURES

Further to the assumptions of § 1, it is supposed, in the prior assessment, that σ^2 and γ are independent. Furthermore, it is assumed that $\nu\lambda/\sigma^2$ possesses a chi-squared distribution with ν degrees of freedom, and that $\rho := e^{-a_0\gamma}$ possesses a beta distribution with parameters a_1 and a_2. Our prior distribution therefore requires the specification of five prior parameters, ν, λ, a_0, a_1 and a_2.

Under the above prior assumptions, the joint distribution of ρ and $\tilde{\alpha}_m$ is

$$\pi(\rho, \tilde{\alpha}_m \mid y) \propto \rho^{a_1-1}(1-\rho)^{a_2-1}|K|^{-\frac{1}{2}}|X'K^{-1}X|^{-\frac{1}{2}}\Omega(\tilde{\alpha}_m)l(\tilde{\alpha}_m \mid y), \tag{3·1}$$

where

$$\Omega(\tilde{\alpha}_m) = (\nu\lambda + \tilde{\alpha}_m' Q\tilde{\alpha}_m)^{-\frac{1}{2}(\nu+m-p)}. \tag{3·2}$$

Consequently, the marginal posterior density of ρ is

$$\pi(\rho \mid y) = c\rho^{a_1-1}(1-\rho)^{a_2-1}|K|^{-\frac{1}{2}}|X'K^{-1}X|^{-\frac{1}{2}}E^*\{\Omega(\tilde{\alpha}_m)\} \quad (0<\rho<1), \tag{3·3}$$

where c is a constant of proportionality, and E^* denotes expectation with respect to $\tilde{\alpha}_m = (\alpha_1, \ldots, \alpha_m)'$ when, for $i = 1, \ldots, m$, the $\theta_i = e^{\alpha_i}/(1+e^{\alpha_i})$ possess independent beta distributions with respective parameters $y_i + 1$ and $n_i - y_i + 1$. The last term in (3·3) can be therefore evaluated by Monte Carlo simulations from the corresponding logistic distributions for the α_i, for any particular value of ρ. Then the proportionality constant c can be evaluated via a one-dimensional numerical integration with respect to ρ.

Furthermore, the posterior expectation given ρ of any transformation $\eta = g(\tilde{\alpha}_m)$ of $\tilde{\alpha}_m$ is

$$E(\eta \mid \rho, y) = \frac{E^*\{g(\tilde{\alpha}_m)\Omega(\tilde{\alpha}_m)\}}{E^*\{\Omega(\tilde{\alpha}_m)\}}, \tag{3·4}$$

and the unconditional posterior expectation of η is

$$E(\eta \mid y) = \int_0^1 E(\eta \mid \rho, y)\pi(\rho \mid y)\,d\rho. \tag{3·5}$$

Similarly, the unconditional posterior expectation of $\alpha(t)$ may be obtained by replacing $\tilde{\alpha}_m$ in (2·8) by its posterior expectation given ρ, as calculated via (3·4), and then averaging the entire expression in (2·8) with respect to the density for ρ in (3·3).

The posterior variance of $\alpha(t)$ given ρ and $\tilde{\alpha}_m$ is

$$\operatorname{var}\{\alpha(t) \mid \rho, \tilde{\alpha}_m, y\} = \hat{\omega}(t)E(\sigma^2 \mid \rho, \tilde{\alpha}_m, y), \tag{3·6}$$

where $\hat{\omega}(t)$ is defined in (2·9). However, the posterior distribution given ρ and $\tilde{\alpha}_m$ of $(\nu\lambda + \tilde{\alpha}_m' Q\tilde{\alpha}_m)/\sigma^2$ is chi-squared with $\nu + m - p$ degrees of freedom. Consequently,

$$\operatorname{var}\{\alpha(t) \mid \rho, \tilde{\alpha}_m, y\} = \hat{\omega}(t)\left(\frac{\nu\lambda + \tilde{\alpha}_m' Q\tilde{\alpha}_m}{\nu + m - p - 2}\right). \tag{3·7}$$

Therefore, the posterior variance of $\alpha(t)$ given only ρ is

$$\operatorname{var}\{\alpha(t) \mid \rho, y\} = \hat{\omega}(t)\left\{\frac{\nu\lambda + E(\tilde{\alpha}_m' Q\tilde{\alpha}_m \mid \rho, y)}{\nu + m - p - 2}\right\} + \operatorname{var}[\{b'(t) + \kappa'(t)Q\}\tilde{\alpha}_m \mid \rho, y]. \tag{3·8}$$

The expectation and variance in (3·8) can be computed as for (3·4). The unconditional posterior variance of $\alpha(t)$ may be found by averaging (3·8) with respect to the posterior density (3·3), and then adding the posterior variance of the posterior expectation of $\alpha(t)$ given ρ, where the variance again refers to the posterior density (3·3).

Since the posterior distribution of $\alpha(t)$ given σ^2, ρ and $\tilde{\alpha}_m$ is normal, with mean $\hat{\alpha}(t)$ and variance $\sigma^2\hat{\omega}(t)$, the joint posterior density of $\alpha(t)$ and $\tilde{\alpha}_m$ given ρ is

$$\pi\{\alpha(t), \tilde{\alpha}_m \mid \rho, y\} \propto |K|^{-\frac{1}{2}}|X'K^{-1}X|^{-\frac{1}{2}} l(\tilde{\alpha}_m \mid y) \times [\nu\lambda + \{\alpha(t) - \hat{\alpha}(t)\}^2/\hat{\omega}(t) + \tilde{\alpha}_m' Q \tilde{\alpha}_m]^{-\frac{1}{2}(\nu+m-p+1)}. \quad (3{\cdot}9)$$

It is consequently also possible to compute the posterior density of $\alpha(t)$, via simulations for $\tilde{\alpha}_m$ combined with numerical integrations for ρ. Similar procedures may be used to compute the unconditional posterior mean and variance of the parametric residual (2·14).

4. Numerical example

We consider a subset to the data previously analysed by Hasselblad, Stead & Crenson (1980) concerning an experiment relating the mortality of mice to nitrogen dioxide exposure (Larsen, Gardner & Coffin, 1979). The second column of Table 1 gives degree of NO_2 exposure, the third the time of exposure in hours, the fourth the number of dead mice, and the fifth the number of mice tested. Our analysis however uses log transforms as the explanatory variables. Each explanatory variable was standardised by subtracting the average, and dividing by the standard deviation.

The main effects model

$$\alpha_i = \beta_0 + \beta_1 t_{i1} + \beta_2 t_{i2} \quad (i = 1, \ldots, 17) \quad (4{\cdot}1)$$

was fitted by maximum likelihood. The fit is slightly unsatisfactory, yielding $\chi^2 = 30{\cdot}9$ on 14 degrees of freedom. The corresponding fit to the probability of death is plotted in curves (A) of Fig. 1, as a function of time of exposure. Figures 1(a), (b) and (c) correspond to the three different levels of NO_2, 1·5, 3·5 and 7·0, respectively. They do not provide a convincing fit to the corresponding observed proportions, which are represented by the crosses.

A multiplicative interaction model

$$\alpha_i = \beta_0 + \beta_1 t_{i1} + \beta_2 t_{i2} + \beta_3 t_{i1} t_{i2} \quad (i = 1, \ldots, 17) \quad (4{\cdot}2)$$

fits much better, yielding $\chi^2 = 15{\cdot}11$ on 13 degrees of freedom. The corresponding fitted probability of death is plotted against time as curves (C) of Fig. 1. The fit is now excellent.

Table 1. *Mortality of mice exposed to* NO_2

Group	t_1	t_2	y	n	Group	t_1	t_2	y	n
1	1·5	96·0	44	120	10	3·5	7·0	152	280
2	1·5	168·0	37	80	11	3·5	14·0	55	80
3	1·5	336·0	43	80	12	3·5	24·0	98	140
4	1·5	504·0	35	60	13	3·5	48·0	121	160
5	3·5	0·5	29	100	14	7·0	0·5	52	120
6	3·5	1·0	53	200	15	7·0	1·0	62	120
7	3·5	2·0	13	40	16	7·0	1·5	61	120
8	3·5	3·0	75	200	17	7·0	2·0	86	120
9	3·5	5·0	23	40					

t_1, degree of NO_2 exposure in ppm; t_2, time of exposure in hours; y, number of dead mice; n, number of mice tested.

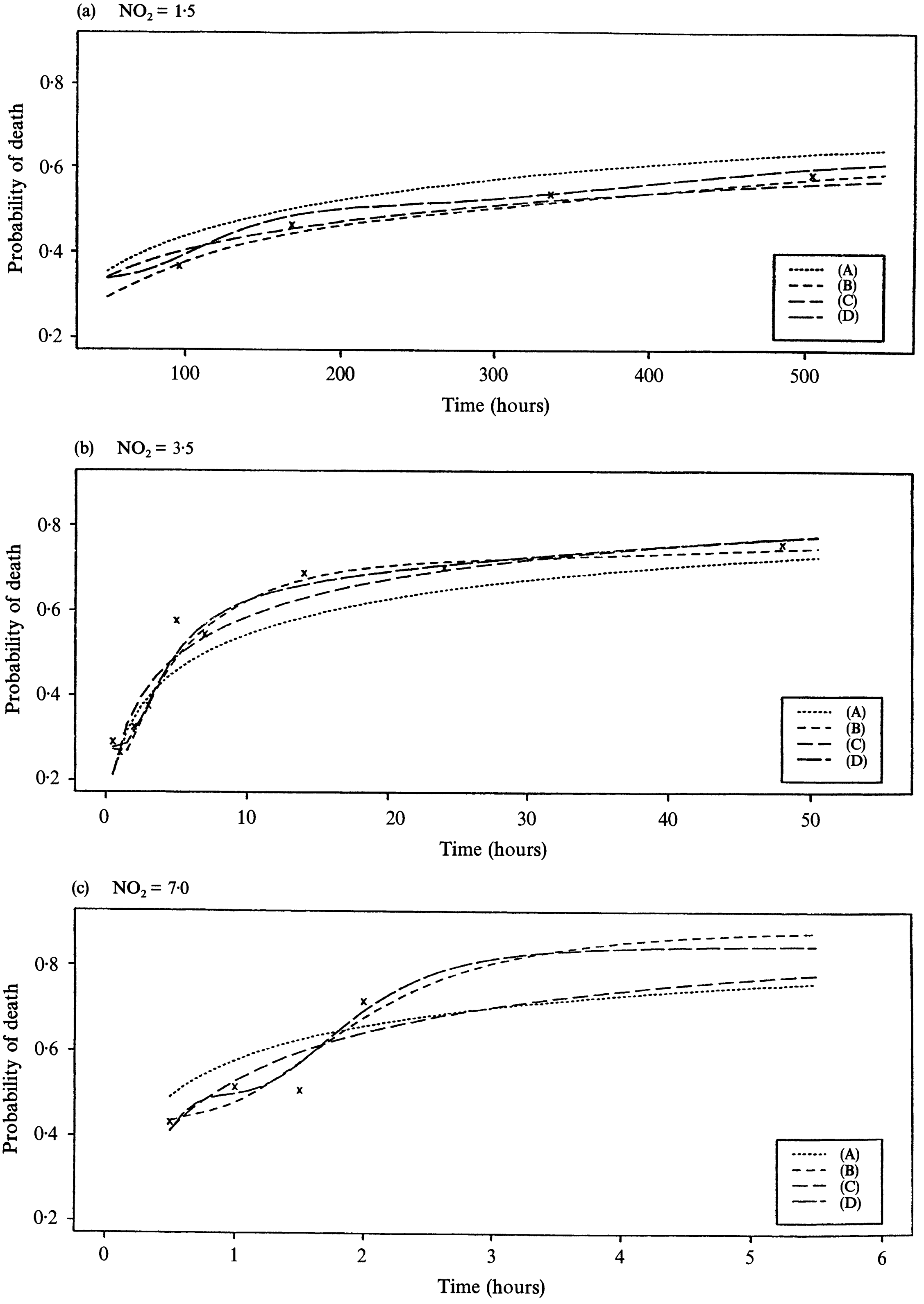

Fig. 1. Predictive probability of death against time for nitrogen dioxide exposure, $NO_2 = 1{\cdot}5$, $3{\cdot}5$ and $7{\cdot}0$. (A) maximum likelihood, without interaction; (B) hierarchical Bayesian, without interaction; (C) maximum likelihood, with interaction; (D) hierarchical Bayesian, with interaction. Crosses, observed proportions.

For our hierarchical Bayesian analyses, the choices $v = \lambda = a_0 = a_1 = a_2 = 1$ were made for the prior parameters, with A in (1·3) the identity matrix. For brevity of presentation, attention is confined to these particular choices.

We first took the main effects model (4·1) to represent the hypothesised model (1·2). The posterior mean value function of the probability of death, unconditional on σ^2 and γ, is given by curves (B) of Fig. 1. This also provides predictive probabilities of death for further mice, at the corresponding levels of NO_2 exposure, and time of exposure.

Note that, in each of Fig. 1(a), (b) and (c), the hierarchical Bayes curve (B) fits the data very well, while still behaving in a smooth manner. This suggests that, even though our first hierarchical Bayesian analysis does not assume a multiplicative interaction term, a more general interaction structure can be effectively modelled by our semiparametric procedure.

As a second hierarchical Bayesian analysis we used the multiplicative interaction model (4·2) as the null model in (1·2). Curves (D) of Fig. 1(a), (b) and (c) provide the exact posterior mean value function, unconditional on σ^2 and γ, of the probability of death. Note that curve (D) of Fig. 1(c) assumes a particularly interesting form in attempting to fit the data closely.

In our first hierarchical Bayesian analysis, without interaction, our exact posterior mean vector $(0{\cdot}0080,\ 0{\cdot}9301,\ 0{\cdot}9852)'$ for $\beta = (\beta_0, \beta_1, \beta_2)'$ somewhat smoothed the maximum likelihood vector $\hat{\beta} = (0{\cdot}0108,\ 0{\cdot}8891,\ 0{\cdot}9902)'$, to compensate for the uncertainty in the choice of model. However, G_1, the exact posterior covariance matrix of β, and G_2, the estimated covariance matrix of $\hat{\beta}$, were respectively

$$G_1 = \begin{pmatrix} 0{\cdot}04615 & 0{\cdot}00515 & 0{\cdot}00002 \\ 0{\cdot}00515 & 0{\cdot}11738 & 0{\cdot}09227 \\ 0{\cdot}00002 & 0{\cdot}09227 & 0{\cdot}10733 \end{pmatrix}, \quad G_2 = \begin{pmatrix} 0{\cdot}00211 & -0{\cdot}00010 & -0{\cdot}00014 \\ -0{\cdot}00010 & 0{\cdot}00628 & 0{\cdot}00532 \\ -0{\cdot}00014 & 0{\cdot}00532 & 0{\cdot}00663 \end{pmatrix}.$$

The matrix G_1 possesses substantially higher diagonal terms than G_2. Hence, when the null model is not well specified, our analysis compensates for this by telling us that we know less about β. This feature is unavailable from a maximum likelihood analysis. The posterior means of β_1 and β_2 nevertheless fall 2·715 and 3·007 posterior standard errors above zero, confirming that both explanatory variables should be included in the analysis.

Our second hierarchical Bayesian analysis, with interaction, yielded posterior mean vector $(0{\cdot}1975, 0{\cdot}8671, 1{\cdot}1013, 0{\cdot}1778)'$ for $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)'$, compared with the maximum likelihood vector $\hat{\beta} = (0{\cdot}1577, 0{\cdot}8902, 1{\cdot}1043, 0{\cdot}1759)'$. In particular the posterior mean 0·1778 for the multiplicative interaction coefficient compares with a posterior standard deviation of 0·1775, so that the multiplicative interaction term should not obviously be included. The exact posterior covariance matrix G_1 of β, not reported here, still inflates the estimated covariance matrix G_2, but the inflation is not quite as substantial as observed for our first hierarchical Bayesian analysis.

Curves (A) and (B) in Fig. 2 describe the posterior densities for the local smoothness parameter γ, under two choices of null model described above. Note that curve (B) is concentrated on lower values of γ. Hence more local smoothing is suggested when an interaction term is included in the null model. This attempts to compensate for the effects of the global shrinkage towards a more complicated null model.

5. Residual analysis

The middle dotted curve in Fig. 3 denotes the exact posterior mean value function of the parametric residual function (2·5) plotted against time, at the third level ($NO_2 = 7$) of

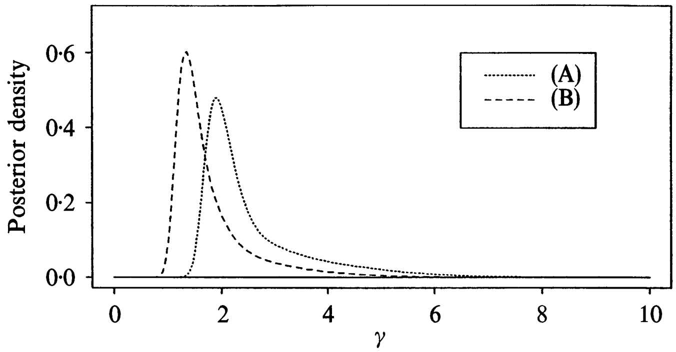

Fig. 2. Posterior densities of smooth parameter γ. (A) hierarchical Bayesian, without interaction; (B) hierarchical Bayesian, with interaction.

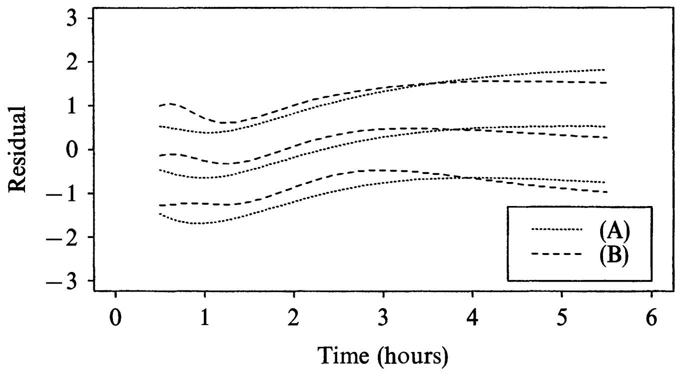

Fig. 3. Residual analysis ($NO_2 = 7{\cdot}0$). (A) hierarchical Bayesian, without interaction; (B) hierarchical Bayesian, with interaction.

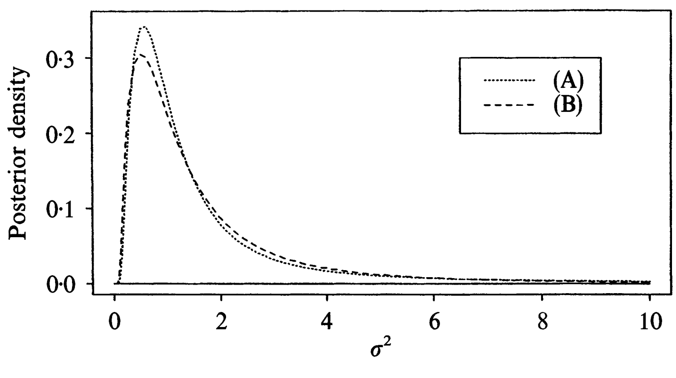

Fig. 4. Posterior densities of global shrinkage parameter σ^2. (A) hierarchical Bayesian, without interaction; (B) hierarchical Bayesian, with interaction.

nitrogen dioxide when the interaction term is absent in our null model. The two outer dotted curves are two posterior standard deviations from the posterior mean. The dashed curves in Fig. 3 represent similar functions when the interaction term is present in the null model. The results indicate that, while a main effects model cannot be refuted, the multiplicative interaction model is slightly superior. For example, the posterior mean value function of the parametric residual function is on the whole closer to zero when an interaction term is included in the null model. Furthermore, the posterior standard errors of the parametric residual function are not substantially increased by the inclusion of an interaction term.

Corresponding residual curves at the other two levels of NO_2 are not reported here. They indicate similar conclusions regarding the slight preferability of the interaction model. This conclusion is confirmed by the overall model checks provided by the curves in Fig. 4. These provide the posterior densities of the global shrinkage parameter σ^2 when the null model is (A) the main effects model, and (B) the multiplicative interaction model. Note that curve (B) is slightly more concentrated towards zero. Overall, our hierarchical Bayesian analysis is much less prepared than maximum likelihood methodology to distinguish between the two null models investigated here.

Acknowledgement

The authors would like to thank Bernie Silverman, Tony O'Hagan, Finbarr O'Sullivan, Brian Yandell and Grace Wahba for helpful discussions, and a referee for indicating the advantages of a hierarchical Bayesian approach.

References

Blight, B. J. N. & Ott, L. (1975). A Bayesian approach to model inadequacy for polynomial regression. *Biometrika* **62**, 79–88.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics* **19**, 15–8.

Cox, D. R. (1970). *The Analysis of Binary Data*. London: Chapman and Hall.

Cox, D. R. & Snell, E. J. (1968). A general definition of residuals (with Discussion). *J. R. Statist. Soc.* B **30**, 248–75.

Gu, C. (1990). Adaptive spline smoothing in non-Gaussian regression models. *J. Am. Statist. Assoc.* **85**, 801–7.

Hasselblad, V., Stead, A. G. & Crenson, J. P. (1980). Multiple probit analysis with a non-zero background. *Biometrics* **36**, 659–63.

Larsen, R. I., Gardner, D. E. & Coffin, D. L. (1979). An air quality data analysis system for interrelating effects, standards, and needed source reductions. Part 5: No. 2 mortality of mice. *J. Air Pollut. Control Assoc.* **39**, 113–7.

Leonard, T. (1972). Bayesian methods for binomial data. *Biometrika* **55**, 581–9.

Leonard, T. (1978). Density estimation, stochastic processes, and prior information. *J. R. Statist. Soc.* B **40**, 113–46.

Leonard, T. & Hsu, J. S. J. (1992). Bayesian inference for a covariance matrix. *Ann. Statist.* **20**, 1669–96.

Leonard, T., Hsu, J. S. J. & Tsui, K. (1989). Bayesian marginal inference. *J. Am. Statist. Assoc.* **84**, 1051–7.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *J. R. Statist. Soc.* B **40**, 1–25.

O'Sullivan, F., Yandell, B. & Raynor, W. (1986). Automatic smoothing of regression functions in generalized linear models. *J. Am. Statist. Assoc.* **81**, 96–103.

Silverman, B. W. (1978). Density ratios, empirical likelihood and cot death. *Appl. Statist.* **27**, 26–33.

[*Received August* 1994. *Revised June* 1995]