

BAYESIAN CALIBRATION IN THE ESTIMATION OF THE AGE OF RHINOCEROS

J. L. DU PLESSIS AND A. J. VAN DER MERWE

*Department of Mathematical Statistics, Faculty of Science, University of the O.F.S.,
P.O. Box 339, Bloemfontein, 9300, Republic of South Africa*

(Received June 28, 1993; revised April 11, 1995)

Abstract. In this paper the Bayesian approach for nonlinear multivariate calibration will be illustrated. This goal will be achieved by applying the Gibbs sampler to the rhinoceros data given by Clarke (1992, *Biometrics*, 48(4), 1081–1094). It will be shown that the point estimates obtained from the profile likelihoods and those calculated from the marginal posterior densities using improper priors will in most cases be similar.

Key words and phrases: Anterior horn length, multivariate Bayesian calibration, nonlinear response model, credibility intervals, Gibbs sampler, posterior horn length, posterior distribution, rhinoceros data, Student-*t* family.

1. Introduction

The traditional calibration problem is concerned with estimating the unknown value of an explanatory variable x corresponding to an observed value of a response variable Y that is assumed to be functionally related to x . Since Krutchkoff's paper appeared in *Technometrics* (1967), a great deal has been written on the calibration problem, see for example, Hunter and Lamboy (1981), Scheffé (1973) and Williams (1969). Brown (1982) gives a broad review of the linear multivariate calibration problem and outlines both Bayesian and non-Bayesian solutions. Oman and Wax (1984) consider an application of linear multivariate calibration in estimating fetal age, given two measurements, femur length and biparietal diameter on human babies. They comment on the nonlinear response model as being very difficult to analyse.

This paper is concerned with obtaining an estimate for an unknown value x corresponding to an observed Y , where Y and x are related through a nonlinear function. A Bayesian approach will be used to make inferences about the unknown x . The example that will be used is one from Clarke (1992). In this problem a sample of rhinoceros of known ages x_i , $i = 1, \dots, n$ have measurements made on the anterior (Y_{ai}) and posterior horn (Y_{pi}). This data are to be used in future to predict a new fixed but unknown age x_0 (denoted by ξ) corresponding to the observed $\mathbf{y}_0 = [y_{a0} \ y_{p0}]'$.

The model proposed by Clarke (1992) is the classic multi-response model, (Bates and Watts (1988), Chapter 4) where the response functions are nonlinear in the unknown parameters. Clarke augmented the original data set with the additional measurements y_{a0} and y_{p0} and a guessed age for ξ_0 of an extra animal whose age is needed. By varying values of ξ_0 he generated varying values of the maximum likelihood, conditional on ξ_0 , and these define the profile likelihood. This profile likelihood can be used and interpreted in a number of ways, all basically intended to illustrate the reliability of the age estimate.

As mentioned by Clarke “The step of associating the profile likelihood with classical confidence limits for the true value ξ_0 is difficult and only approximate solutions are available. To that extent this paper should not be regarded as proposing a definitive answer”. Smith and Corbett (1987) pointed out that the profile likelihood often fails to tend to zero as $\xi_0 \rightarrow \pm\infty$ so it cannot strictly be renormalized and treated as a posterior density. It therefore seems obvious that emphasis should be placed on Bayesian estimation of the model parameters.

Technical difficulties arising in the calculation of the marginal posterior densities needed for Bayesian inference have long served as an impediment to the wider application of the Bayesian framework to real data. The reason for this is that the integration operation plays a fundamental role in Bayesian statistics. In the last few years there have been a number of advances in numerical integration and analytic approximation techniques for such calculations but implementation of these approaches typically requires sophisticated numerical or analytic approximation expertise and possibly specialist software.

Recently due to work of Gelfand and Smith (1991), Gelfand *et al.* (1990), Carlin *et al.* (1992) and Gelfand *et al.* (1992), the Gibbs sampler has been shown as a useful tool for applied Bayesian inference in a broad variety of statistical problems. The Gibbs sampler is implicit in the work of Hastings (1970) and made popular in the image processing context by Geman and Geman (1984). The Gibbs sampler is an adaptive Monte Carlo integration technique. The typical objective of the sampler is to collect a sufficiently large enough number of parameter realizations from conditional posterior densities in order to obtain accurate estimates of the marginal posterior densities. The principal requirement of the sampler is that all conditional densities must be available in the sense that random variates can be generated from them.

In what follows a Bayesian procedure for nonlinear multivariate calibration will be illustrated. This goal will be achieved by applying the Gibbs sampler to the rhinoceros data given in Clarke (1992). As will be seen the point estimates obtained from the profile likelihoods and those calculated from the marginal posterior densities using improper priors will in most cases be similar.

Using a similar notation to that of Clarke, the assumed model for the n -sample calibrating data can be written as

$$(1.1) \quad Y_{ji} = f_j(\boldsymbol{\theta}_j, x_i) + \varepsilon_{ji},$$

where Y_{ji} denote the value of the variate j , $j = a$ or p on animal i , $i = 1, \dots, n$ of age x_i . f_j are nonlinear functions which may or may not differ for various j , $\boldsymbol{\theta}_j$ are unknown parameters of dimension m , some of which may be common for

various j , and the random errors ε_{ji} are normally distributed with zero means, independent for different i but having covariance matrix Σ of dimension 2×2 , when i is common.

The observation of the prediction experiment is assumed to follow the same assumptions as those of the calibration experiment. In particular

$$(1.2) \quad y_{j0} = f_j(\boldsymbol{\theta}_j, \xi) + \varepsilon_{j0},$$

where ε_{j0} are independently distributed from ε_{ji} .

The marginal posterior distribution of ξ cannot be derived in closed form so the Gibbs sampler will be used. Missing values in the response variables are treated as unknown parameters and the Gibbs sampler again provides a natural framework for obtaining approximations of the unconditional posterior densities.

2. A Bayesian solution

For the rhinoceros data the model proposed by Clarke (1992), is

$$(2.1) \quad \begin{aligned} Y_{ai} &= \theta_1 + \theta_2 \theta_3^{x_i} + \varepsilon_{ai}, \\ Y_{pi} &= \theta_1 + \theta_4 \theta_5^{x_i} + \varepsilon_{pi}, \quad i = 1, \dots, n \end{aligned}$$

where $Y = \log_e$ (horn length) and $x = \text{age in year}$.

The least square analysis used by Clarke (1992) showed that there is no benefit to be gained by fitting separate asymptotes to the two curves and consequently the model was reduced to five parameters. In the analysis of experimental data we assume that the n experimental design variables x_i , $i = 1, \dots, n$ are fixed and known, so we can form the $n \times 2$ observation matrix \mathbf{Y} with the (j, i) th element Y_{ji} . Model (2.1) for the calibration problem can thus be written as

$$(2.2) \quad \mathbf{y}_r = \mathbf{1}\theta_1 + \mathbf{F}_\theta \boldsymbol{\theta} + \boldsymbol{\varepsilon}_r$$

where \mathbf{y}_r is a $(2n \times 1)$ column vector obtained from stacking the rows of the response matrix \mathbf{Y} into one column. \mathbf{F}_θ is a $(2n \times 2)$ matrix of functions, nonlinear in θ_3 and θ_5 and $\boldsymbol{\theta}(2 \times 1) = [\theta_2 \ \theta_4]'$. Furthermore $\mathbf{1}$ is a $(2n \times 1)$ column vector of ones and $\boldsymbol{\varepsilon}_r(2n \times 1) \sim N(\mathbf{0}, \mathbf{I}_n \otimes \Sigma)$.

Also the prediction model (2.1) can be written as

$$(2.3) \quad \mathbf{y}'_0 = \tilde{\mathbf{1}}'\theta_1 + \boldsymbol{\theta}'\mathbf{F}_\xi + \boldsymbol{\varepsilon}'_0$$

where $\tilde{\mathbf{1}}(2 \times 1)$ is a column vector of ones and $\mathbf{F}_\xi(2 \times 2)$ is a matrix nonlinear in θ_3, θ_5 and ξ . Furthermore $\boldsymbol{\varepsilon}_0(2 \times 1) \sim N(\mathbf{0}, \Sigma)$.

As mentioned in the last paragraph of Section 1 we would like to make inferences about ξ without deleting partially observed data. Assume that the missing values are in the last two rows of \mathbf{Y} , i.e.

$$\mathbf{Y} = \begin{bmatrix} \tilde{\mathbf{Y}} \\ \mathbf{y}'_V \\ \mathbf{y}'_L \end{bmatrix}, \quad \mathbf{F}_\theta = \begin{bmatrix} \tilde{\mathbf{F}}_\theta \\ \mathbf{F}_{\theta V} \\ \mathbf{F}_{\theta L} \end{bmatrix}$$

where $\tilde{\mathbf{Y}} = \begin{bmatrix} y_{a1} & y_{p1} \\ y_{a2} & y_{p2} \\ \vdots & \vdots \\ y_{a,n-2} & y_{p,n-2} \end{bmatrix}$ and $\tilde{\mathbf{F}}_\theta = \begin{bmatrix} \theta_3^{x_1} & 0 \\ 0 & \theta_5^{x_1} \\ \vdots & \vdots \\ \theta_3^{x_{n-2}} & 0 \\ 0 & \theta_5^{x_{n-2}} \end{bmatrix}$.

Also let the last two rows of Y be denoted by

$$\mathbf{y}'_V = [y_{a,n-1} \ y_{p,n-1}^V] = [y_{V(1)} \ y_{V(2)}], \quad \mathbf{y}'_L = [y_{an} \ y_{pn}^L] = [y_{L(1)} \ y_{L(2)}]$$

where $y_{p,n-1}^V, y_{pn}^L$ denote the missing values and

$$\mathbf{F}_{\theta V} = \begin{bmatrix} \theta_3^{x_{n-1}} & 0 \\ 0 & \theta_5^{x_{n-1}} \end{bmatrix}, \quad \mathbf{F}_{\theta L} = \begin{bmatrix} \theta_3^{x_n} & 0 \\ 0 & \theta_5^{x_n} \end{bmatrix},$$

$$\mathbf{F}_\xi = \begin{bmatrix} \theta_3^\xi & 0 \\ 0 & \theta_5^\xi \end{bmatrix} \quad \text{and} \quad \mathbf{y}_0 = \begin{bmatrix} y_{ao} \\ y_{po} \end{bmatrix}.$$

In the Bayesian treatment of the problem priors must be specified for the unknown parameters $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \Sigma$ and ξ .

According to Mitchell (1967) and Ye and Berger (1991) flat priors can be used for θ_1, θ_2 and θ_4 but proper uniform distributions for θ_3 and θ_5 i.e. $0 \leq \theta_3 \leq 1$ and $0 \leq \theta_5 \leq 1$ might cause problems and can result in improper posterior distributions. Both these parameters should therefore lie in the open interval $(0, 1)$.

For this problem, θ_1 is positive and measures the maximum anterior and posterior horn length. The parameters θ_2 and θ_4 are negative, $-\theta_2$ and $-\theta_4$ being the maximum possible increases in anterior and posterior horn lengths respectively for a unit (year) increase in time, while $0 < \theta_3 < 1$ and $0 < \theta_5 < 1$ measure the influence of age on horn length. Note that "horn length" here refers to the natural logarithm of the measurable horn lengths.

Consulting zoologists at the University of the Orange Free State and the Kruger National Park as well as the curator of the Bloemfontein zoo, estimates of θ_3 and θ_4 ranged from 0,4 to 0,95 with median 0,8 which means that in practice θ_3 and θ_5 will not be equal to 1 and definitely never 0.

Although it was possible for Ye and Berger (1991) to derive the reference prior for the exponential regression model $Y = \theta_1 + \theta_2\theta_3^x + \varepsilon$ for the group ordering $\{\theta_3, \theta_1, \theta_2, \sigma_{11}\}$ reasonably easy and to show that it was only dependent on θ_3 and σ_{11} , the reference prior for the group ordering $\{\xi, \theta_3, \theta_5, \theta_1, \theta_2, \theta_4, \Sigma\}$, in the case of model (2.1) is very complicated and for all practical purposes impossible to obtain.

However by comparing the posterior density of θ_3 based on the reference prior for the model $Y = \theta_1 + \theta_2\theta_3^x + \varepsilon$ to the corresponding posterior using a prior uniformly distributed on the interval $[a, b]$ where $a = 0.05 > 0$ and $b = 0.095 < 1$ it was observed that the two posteriors are for all practical purposes identical. The same correspondence was observed for the parameter θ_5 .

In this paper we will therefore use as joint prior

$$(2.4) \quad p(\xi, \theta_3, \theta_5, \theta_1, \theta_2, \theta_4, \Sigma) \propto \Sigma^{-(p+1)/2} p(\xi)$$

where

$$-\infty \leq \theta_1 \leq \infty, \quad -\infty \leq \theta_2 \leq \infty, \quad -\infty \leq \theta_4 \leq \infty,$$

$$0.05 \leq \theta_3 \leq 0.95 \quad \text{and} \quad 0.05 \leq \theta_5 \leq 0.95.$$

Also the parameters $\theta_1, \theta_2, \theta_4$, occur as linear parameters in model (2.1) which is a further indication that flat priors can be used on them.

Define

$$(2.5) \quad \tilde{\boldsymbol{\theta}} = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]' \quad \text{and} \quad \boldsymbol{\theta} = [\theta_2, \theta_4]'$$

The joint density function of $\tilde{\mathbf{Y}}, \mathbf{y}_0, \mathbf{y}_L, \mathbf{y}_V, \tilde{\boldsymbol{\theta}}, \xi$ and $\boldsymbol{\Sigma}$ is therefore given by

$$(2.6) \quad f(\tilde{\mathbf{Y}}, \mathbf{y}_0, \mathbf{y}_L, \mathbf{y}_V, \tilde{\boldsymbol{\theta}}, \xi, \boldsymbol{\Sigma}) \\ \propto |\boldsymbol{\Sigma}|^{-(n+p+2)/2} \\ \cdot \exp \left\{ -\frac{1}{2} (\tilde{\mathbf{y}}_r - \mathbf{1}\theta_1 - \tilde{\mathbf{F}}_{\boldsymbol{\theta}}\boldsymbol{\theta})' (\mathbf{I}_{n-2} \otimes \boldsymbol{\Sigma})^{-1} (\tilde{\mathbf{y}}_r - \mathbf{1}\theta_1 - \tilde{\mathbf{F}}_{\boldsymbol{\theta}}\boldsymbol{\theta}) \right. \\ \quad - \frac{1}{2} (\mathbf{y}'_0 - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\xi})\boldsymbol{\Sigma}^{-1} (\mathbf{y}'_0 - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\xi})' \\ \quad - \frac{1}{2} (\mathbf{y}'_L - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\theta L})\boldsymbol{\Sigma}^{-1} (\mathbf{y}'_L - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\theta L})' \\ \quad \left. - \frac{1}{2} (\mathbf{y}'_V - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\theta V})\boldsymbol{\Sigma}^{-1} (\mathbf{y}'_V - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\theta V})' \right\} p(\xi).$$

The marginal posterior distribution $f(\xi | \tilde{\mathbf{Y}}, \mathbf{y}_0, \mathbf{y}_{L(1)}, \mathbf{y}_{V(1)})$ cannot be obtained analytically and because of the many unknowns very difficult to calculate numerically. To implement the Gibbs sampler we need the full conditionals given by

$$(2.7) \quad \boldsymbol{\Sigma}^{-1} | \tilde{\mathbf{Y}}, \mathbf{y}_0, \mathbf{y}_L, \mathbf{y}_V, \tilde{\boldsymbol{\theta}}, \xi \sim W(\mathbf{R}^{-1}, n+1)$$

where

$$\mathbf{R} = (\tilde{\mathbf{Y}} - \tilde{\mathbf{M}})' (\tilde{\mathbf{Y}} - \tilde{\mathbf{M}}) + (\mathbf{y}'_0 - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\xi})' (\mathbf{y}'_0 - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\xi}) \\ + (\mathbf{y}'_L - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\theta L})' (\mathbf{y}'_L - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\theta L}) \\ + (\mathbf{y}'_V - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\theta V})' (\mathbf{y}'_V - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\theta V})$$

and

$$\tilde{\mathbf{M}} = \begin{bmatrix} \theta_1 + \theta_2\theta_3^{x_1} & \theta_1 + \theta_4\theta_5^{x_1} \\ \vdots & \vdots \\ \theta_1 + \theta_2\theta_3^{x_{n-2}} & \theta_1 + \theta_4\theta_5^{x_{n-2}} \end{bmatrix}.$$

By choosing a normal prior $\xi \sim N(\mu_0, \sigma_0^2)$, the full conditional for ξ is given by

$$(2.8) \quad f(\xi | \tilde{\mathbf{Y}}, \mathbf{y}_0, \mathbf{y}_L, \tilde{\boldsymbol{\theta}}, \boldsymbol{\Sigma}) \\ \propto \exp \left\{ -\frac{1}{2} [(\mathbf{y}'_0 - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\xi})\boldsymbol{\Sigma}^{-1} (\mathbf{y}'_0 - \tilde{\mathbf{1}}'\theta_1 - \boldsymbol{\theta}'\mathbf{F}_{\xi})' \right. \\ \quad \left. + (\xi - \mu_0)^2 / \sigma_0^2] \right\}.$$

The full conditionals for θ_1 through θ_5 are given by

$$(2.9) \quad \boldsymbol{\theta} | \theta_1, \theta_3, \theta_5, \tilde{\mathbf{Y}}, \mathbf{y}_0, \mathbf{y}_L, \mathbf{y}_V, \xi, \boldsymbol{\Sigma} \sim N(\tilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}})$$

where

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}} = & \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} [\tilde{\mathbf{F}}' (\mathbf{I}_{n-2} \otimes \boldsymbol{\Sigma})^{-1} (\tilde{\mathbf{y}}_r - \mathbf{1}' \boldsymbol{\theta}_1) + \mathbf{F}_{\xi} \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_0 - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1)' \\ & + \mathbf{F}_{\boldsymbol{\theta}L} \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_L - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1)' + \mathbf{F}_{\boldsymbol{\theta}V} \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_V - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1)'] \end{aligned}$$

and

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}} = [\tilde{\mathbf{F}}' (\mathbf{I}_{n-2} \otimes \boldsymbol{\Sigma})' \tilde{\mathbf{F}} + \mathbf{F}_{\xi} \boldsymbol{\Sigma}^{-1} \mathbf{F}'_{\xi} + \mathbf{F}_{\boldsymbol{\theta}L} \boldsymbol{\Sigma}^{-1} \mathbf{F}'_{\boldsymbol{\theta}L} + \mathbf{F}_{\boldsymbol{\theta}V} \boldsymbol{\Sigma}^{-1} \mathbf{F}'_{\boldsymbol{\theta}V}]^{-}$$

$$(2.10) \quad \boldsymbol{\theta}_1 \mid \boldsymbol{\theta}, \boldsymbol{\theta}_3, \boldsymbol{\theta}_5, \tilde{\mathbf{Y}}, \mathbf{y}_0, \mathbf{y}_L, \mathbf{y}_V, \xi, \boldsymbol{\Sigma} \sim N(\tilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}_1}, \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\theta}_1}^2)$$

where

$$\begin{aligned} \tilde{\boldsymbol{\mu}}_{\boldsymbol{\theta}_1} = & \tilde{\boldsymbol{\sigma}}_{\boldsymbol{\theta}_1}^2 [\mathbf{1}' (\mathbf{I}_{n-2} \otimes \boldsymbol{\Sigma})^{-1} (\tilde{\mathbf{y}}_r - \tilde{\mathbf{F}} \boldsymbol{\theta}) + \mathbf{1}' \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_0 - \boldsymbol{\theta}' \mathbf{F}_{\xi})' \\ & + \tilde{\mathbf{1}}' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_L - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}L}) + \tilde{\mathbf{1}}' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_V - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}V})] \end{aligned}$$

and

$$\tilde{\boldsymbol{\sigma}}_{\boldsymbol{\theta}_1}^2 = [\mathbf{1}' (\mathbf{I}_{n-2} \otimes \boldsymbol{\Sigma})^{-1} \mathbf{1} + 3 \tilde{\mathbf{1}}' \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{1}}]^{-1}.$$

$$(2.11) \quad f(\boldsymbol{\theta}_3 \mid \boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_5, \tilde{\mathbf{Y}}, \mathbf{y}_0, \mathbf{y}_L, \mathbf{y}_V, \xi, \boldsymbol{\Sigma})$$

$$\begin{aligned} \propto \exp \left\{ -\frac{1}{2} \left\{ (\tilde{\mathbf{y}}_r - \mathbf{1} \boldsymbol{\theta}_1 - \tilde{\mathbf{F}}_{\boldsymbol{\theta}} \boldsymbol{\theta})' (\mathbf{I}_{n-2} \otimes \boldsymbol{\Sigma})^{-1} (\tilde{\mathbf{y}}_r - \mathbf{1} \boldsymbol{\theta}_1 - \tilde{\mathbf{F}}_{\boldsymbol{\theta}} \boldsymbol{\theta}) \right. \right. \\ + (\mathbf{y}'_0 - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\xi}) \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_0 - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\xi})' \\ + (\mathbf{y}'_L - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}L}) \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_L - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}L})' \\ \left. \left. + (\mathbf{y}'_V - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}V}) \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_V - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}V})' \right\} \right\}, \end{aligned}$$

$$(2.12) \quad f(\boldsymbol{\theta}_5 \mid \boldsymbol{\theta}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_3, \tilde{\mathbf{Y}}, \mathbf{y}_0, \mathbf{y}_L, \mathbf{y}_V, \xi, \boldsymbol{\Sigma})$$

$$\begin{aligned} \propto \exp \left\{ -\frac{1}{2} \left\{ (\tilde{\mathbf{y}}_r - \mathbf{1} \boldsymbol{\theta}_1 - \tilde{\mathbf{F}}_{\boldsymbol{\theta}} \boldsymbol{\theta})' (\mathbf{I}_{n-2} \otimes \boldsymbol{\Sigma})^{-1} (\tilde{\mathbf{y}}_r - \mathbf{1} \boldsymbol{\theta}_1 - \tilde{\mathbf{F}}_{\boldsymbol{\theta}} \boldsymbol{\theta}) \right. \right. \\ + (\mathbf{y}'_0 - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\xi}) \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_0 - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\xi})' \\ + (\mathbf{y}'_L - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}L}) \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_L - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}L})' \\ \left. \left. + (\mathbf{y}'_V - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}V}) \boldsymbol{\Sigma}^{-1} (\mathbf{y}'_V - \tilde{\mathbf{1}}' \boldsymbol{\theta}_1 - \boldsymbol{\theta}' \mathbf{F}_{\boldsymbol{\theta}V})' \right\} \right\}. \end{aligned}$$

The conditional density for the missing value $y_{L(2)}$ is

$$(2.13) \quad y_{L(2)} \mid \tilde{\mathbf{Y}}, \mathbf{y}_0, y_{L(1)}, \mathbf{y}_V, \boldsymbol{\theta}, \xi, \boldsymbol{\Sigma} \sim N(E(y_{L(2)}), \text{Var}(y_{L(2)}))$$

where

$$E(y_{L(2)}) = \mu_{L(2)} + \sigma_{21} \sigma_{11}^{-1} (y_{L(1)} - \mu_{L(1)})$$

and

$$\text{Var}(y_{L(2)}) = \sigma_{22} - \sigma_{21} \sigma_{11}^{-1} \sigma_{12},$$

(see for example Anderson (1984)).

Also for the missing value $y_{V(2)}$ it follows that

$$(2.14) \quad y_{V(2)} \mid \tilde{\mathbf{Y}}, \mathbf{y}_0, y_{V(1)}, \mathbf{y}_L, \tilde{\boldsymbol{\theta}}, \xi, \boldsymbol{\Sigma} \sim N(E(y_{V(2)}), \text{Var}(y_{V(2)}))$$

where

$$E(y_{V(2)}) = \mu_{V(2)} + \sigma_{21}\sigma_{11}^{-1}(y_{V(1)} - \mu_{V(1)}),$$

$$\text{Var}(y_{V(2)}) = \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12},$$

$$\boldsymbol{\mu}_V = \begin{bmatrix} \mu_{aV} \\ \mu_{pV} \end{bmatrix} = \begin{bmatrix} \theta_1 + \theta_2\theta_3^{x_n-1} \\ \theta_1 + \theta_4\theta_5^{x_n-1} \end{bmatrix},$$

$$\boldsymbol{\mu}_L = \begin{bmatrix} \mu_{aL} \\ \mu_{pL} \end{bmatrix} = \begin{bmatrix} \theta_1 + \theta_2\theta_3^{x_n} \\ \theta_1 + \theta_4\theta_5^{x_n} \end{bmatrix}$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}.$$

The conditional distributions of ξ , θ_3 and θ_5 , are however not in closed form, hence we only have the kernels of these densities. Random numbers can still be generated by using the weighted Bootstrap method which is discussed in Smith and Gelfand (1992) and Stephens and Smith (1992). Simulation from the Wishart distribution is easily done by using the algorithm of Odell and Feiveson (1966).

The Gibbs-sampler can now easily be implemented. The iterative process starts by using arbitrary starting values $\boldsymbol{\Sigma}^{(0)}$, $\tilde{\boldsymbol{\theta}}^{(0)}$, $\xi^{(0)}$, $\mathbf{y}_{L(2)}^{(0)}$, $\mathbf{y}_{V(2)}^{(0)}$ to calculate the first iteration using (2.7) to (2.14). After k iterations in which the conditional densities were updated at each iteration the Gibbs-sampler has generated the values $\boldsymbol{\Sigma}^{(k)}$, $\tilde{\boldsymbol{\theta}}^{(k)}$, $\xi^{(k)}$, $\mathbf{y}_{L(2)}^{(k)}$, $\mathbf{y}_{V(2)}^{(k)}$. The process is repeated m times.

The convergence of the algorithm is discussed in Gelfand and Smith (1991). The problem in this paper was done with $k = 30$ and $m = 1000$.

In a further study which is not included in this paper, we also compared the marginal posterior distributions for $k = 40$ and 50 as well as for an alternative Markov Chain Monte Carlo (MCMC) procedure in which after a long run (i.e. k sufficiently large) successive albeit correlated values of the random variable of interest was used to describe the unknown marginal distribution. As ‘‘burn-in-period’’ we used $k = 500$ iterations. The densities obtained, indicated that convergence had been achieved in all the cases which means that $k = 30$ and $m = 1000$ are appropriate for our problem.

It is less expensive to use successive values in the Gibbs sampling procedure. Gelman and Rubin (1992), however, warned ‘‘that particularly during the first tentative examination of a new problem, it can be argued that monitoring the evolutionary behaviour of several runs of the chain starting from a wide range of interval values is necessary’’. It is for this reason that we have decided to stick to the ‘‘traditional’’ procedure.

3. Example

The anterior and posterior horn lengths were recorded on ten rhinoceros of known age and only the anterior horn length for another two animals. The measured data as well as their natural logarithms, are recorded in Table 1. These data will then be used to predict the age of animals by measuring their horn lengths. It can be seen from the data that the length of the anterior horn does not increase very much after about 14 years of age. This is due to the natural wear on that horn during browsing. The posterior horn on the other hand, is protected from wear and increases steadily in length.

Table 1. Rhinoceros horn lengths.

Age	Anterior Horn - mm	\log_e (Ant. horn) Y_a	Posterior Horn - mm	\log_e (Post. Horn) Y_p
2	195	5.27	70	4.25
4	248	5.51	95	4.55
33	536	6.28	missing	missing
5	380	5.94	160	5.08
4	265	5.58	97	4.57
30	534	6.28	missing	missing
4	262	5.57	98	4.58
3.5	225	5.42	80	4.38
10	530	6.27	200	5.30
16	532	6.28	303	5.71
27	425	6.05	405	6.00
12	600	6.40	310	5.74

In our first example prediction for age of rhinoceros is based on a young animal with anterior horn length = 424 mm and posterior horn length = 172 mm, i.e. $\mathbf{y}_0 = [6.05 \ 5.15]'$ = $[\log_e(\text{anterior horn}) \ \log_e(\text{posterior horn})]'$. We will first consider a uniform prior on ξ and then proper priors on ξ . A uniform prior on ξ can be used provided the number of response variables are equal or greater than two (Brown (1982)). The estimated marginal posterior densities of the unknown parameters can easily be obtained by using Gibbs sampling. For lack of space the estimated posterior densities will not be displayed but are illustrated in du Plessis and van der Merwe (1994).

From the posterior density of ξ (age of distribution of young animal) it is observed that the mean = 8.32 years and the mode = 7.8 years. The 90% and 95% credibility intervals are 5.47–11.66 and 4.96–12.89 years respectively. Inspection of Table 2 (Clarke (1992), p. 1087) shows that for the young animal the estimated age is 7.9 years while the 90% and 95% confidence intervals are 5.4–11.3 and 4.9–12.3 years respectively. Comparing the results of the Bayes procedure with that

Table 2. Posterior estimates of the unknown parameters.

	Estimate
Maximum Hornlength	528 mm
Increase in Anterior Hornlength per year	6.4 mm
Increase in Posterior Hornlength per year	12.18 mm
Influence of Age on Anterior Hornlength	0.78
Influence of Age on Posterior Hornlength	0.92

of the profile likelihood, it can be concluded that the estimated ages for the two procedures are more or less equal but the Bayes intervals seem to be somewhat wider than those obtained from the profile likelihoods. Bayesian estimates of the other parameter values can also be obtained. The posterior modes of the parameters θ_1 through to θ_5 are 6.27, -1.86 , 0.78, -2.50 and 0.92 respectively which are almost identical to those given by Clarke, i.e. 6.29, -1.89 , 0.77, -2.50 and 0.91 respectively. As in the case of ξ the credibility intervals (using the full Bayesian approach) are somewhat wider than those obtained from the profile likelihoods. The posterior standard deviations of θ_1 through to θ_5 are 0.0857, 0.4040, 0.0570, 0.2077 and 0.017 while those calculated by Clarke are 0.067, 0.378, 0.051, 0.164 and 0.011 respectively.

As mentioned θ_1 measures the maximum horn length; $-\theta_2$ and $-\theta_4$ being the maximum possible increases in horn lengths while θ_1 and θ_3 measure the influence of age on horn length. Note that "horn length" here refers to the natural logarithm of the measurable horn lengths. Therefore in Table 2 the actual estimates are given.

The Gibbs sampler also automatically provides "predictive" densities of the missing horn lengths for the rhinoceros 30 and 33 years of age. For purposes of condensation these densities are not included here but the estimated horn lengths were 422 and 431 millimeter respectively.

For illustrative purposes we provide a Bayesian analysis based on two different prior specifications ($\xi \sim N(0, 100)$ and $\xi \sim N(10, 36)$) for the age of the young animal. The priors for the other parameters are again taken to be constant. A comparison between these posterior densities and the posterior density of ξ using the uniform prior show that the normal priors provide some information about ξ . The corresponding posteriors are more peaked with the result that the 95% credibility intervals 4.82–11.48 years and 5.28–12.56 years are somewhat shorter. The modes of the two posteriors are 7.56 years and 7.92 years respectively. As could be expected the first prior ($\xi \sim N(0, 100)$) shifted the posterior somewhat to the left while the second prior moved it to the right.

In our second example various prior specifications are compared for the age distribution of the old animal with anterior horn length = 528.48 mm and posterior horn length = 365.04 mm. The priors that will be used are $\xi \propto \text{constant}$, $\xi \sim N(0, 1000)$, $\xi \sim N(0, 100)$, $\xi \sim N(16, 100)$ and $\xi \sim N(16, 36)$.

The posterior density for $\xi \propto \text{constant}$ has got a very long tail to the right. We could not normalize the distribution so that no credibility intervals could be calculated from it. The mode of the curve is however 19.0 years which is more or

less the same as the estimated 19.2 years given by Clarke. For the second prior ($\xi \sim N(0, 1000)$) the posterior mode is also 18.8 while the 95% credibility interval is 11.40–40.40 years. The corresponding 95% interval calculated from the profile likelihood is 12.8–31.7. The Bayes interval is again somewhat wider. The posterior modes for the remaining cases $\xi \sim N(0, 100)$, $\xi \sim N(16, 100)$ and $\xi \sim N(16, 36)$ are 16.4, 18.4 and 17.9 years respectively while the 95% credibility intervals are 10.62–24.11, 11.60–28.41 and 12.37–25.20 years respectively. For these three cases the 95% credibility intervals are somewhat shorter than the corresponding intervals given by Clarke. A comparison of the modes of the posterior densities and the lengths of the credibility intervals show that the priors provide substantial information about ξ for the old animal. For further details see du Plessis and van der Merwe (1994).

To accommodate the possibility of outlying rhinos the assumption of Gaussian errors will be relaxed in the direction of the Student- t family. Consider the series of independent errors $\boldsymbol{\varepsilon}_i \mid \boldsymbol{\Sigma}, \lambda_i \sim N(\mathbf{0}, \lambda_i \boldsymbol{\Sigma})$, ($i = 1, \dots, n$). By placing a prior on λ_i enables a wide variety of model error densities $f(\boldsymbol{\varepsilon}_i \mid \boldsymbol{\Sigma})$ to emerge as scale mixtures of normal distributions (Andrews and Mallows (1974), Carlin and Polson (1991) and Wakefield *et al.* (1994)).

$$f(\boldsymbol{\varepsilon}_i \mid \boldsymbol{\Sigma}) = \int p(\boldsymbol{\varepsilon}_i \mid \boldsymbol{\Sigma}, \lambda_i) p(\lambda_i) d\lambda_i \quad (i = 1, \dots, n).$$

For this example it is assumed that $\nu \lambda_i^{-1} \sim \chi_\nu^2$ so that $\boldsymbol{\varepsilon}_i \mid \boldsymbol{\Sigma} \sim t_\nu(\mathbf{0}, \boldsymbol{\Sigma})$, a multivariate Student- t -distribution with mean $\mathbf{0}$, covariance matrix $\boldsymbol{\Sigma}$ and degrees of freedom ν . As in Wakefield *et al.* (1994) we will take ν equal to 2.

The conditional posterior density of λ_i will now be an Inverse Gamma density and the remaining hierarchical structure is defined exactly as in Section 2 except that λ_i , ($i = 1, \dots, n$) will occur in the expressions of the conditional densities defined in equations (2.7)–(2.14).

As mentioned one of the principal aims of this illustration is the detection of outlying rhinos by using the Student- t -model. According to Wakefield *et al.* (1994) the scale parameter λ_i^{-1} or $\log_{10}(\lambda_i^{-1})$ is a good global indicator of outliers. The prior expectation of $\log_{10}(\lambda_i^{-1})$ is 0, so that a $\log_{10}(\lambda_i^{-1})$ -value substantially below zero indicates that the i -th rhinoceros is likely to be an outlier. Interval estimates of $\log_{10}(\lambda_i^{-1})$ are easily constructed. These interval estimates can be thought of as describing the marginal posterior distributions of the $\log_{10}(\lambda_i^{-1})$. In Table 3 $E\{\log_{10}(\lambda_i^{-1})\}$ against the rhinoceros number i with the 95% posterior density intervals are given. Since none of the $\log_{10}(\lambda_i^{-1})$ parameter values are substantially below zero it can be concluded that there are no outlying rhinoceros.

For the Student- t model it was observed that the mean of the posterior density of ξ (age of young animal) is 9.37 years and the mode = 8.22 years while the 90% and 95% credibility intervals are 5.60–16.38 years and 4.87–20.60 years respectively.

A comparison of this density with the posterior density of ξ assuming Gaussian errors and a uniform prior on ξ shows that the central values are more or less the same but the credibility intervals are somewhat wider for the Student- t model.

Table 3. $E\{\log_{10}(\lambda_i^{-1})\}$ against case number i with the 95% credibility interval.

Case Number i	$E\{\log_{10}(\lambda_i^{-1})\}$	95% Credibility Interval
1	1.1002	-0.2245 - 4.2812
2	0.6097	-0.9116 - 2.1193
3	1.1650	-0.0180 - 4.3252
4	0.3986	-0.8465 - 2.1654
5	0.4329	-0.7886 - 2.6482
6	0.5523	-0.6418 - 2.6961
7	0.9471	-0.2963 - 3.8344
8	0.5577	-0.7014 - 2.5776
9	1.0632	-0.2261 - 4.0227
10	0.9949	-0.0594 - 3.8232
11	0.5663	-0.6863 - 3.0241
12	0.5715	-0.9517 - 3.0387
13	0.6415	-0.7914 - 2.8559

Table 4. Posterior estimates of ξ , the age (years) of the two rhinoceros.

	Mode	95% Credibility Interval
Young Animal - Uniform Prior	7.8	4.96 - 12.89
Young Animal - N(0, 100) Prior	7.56	4.82 - 11.48
Young Animal - N(10,36) Prior	7.92	5.28 - 12.56
Young Animal - Uniform Prior Student- t Model	8.22	4.87 - 20.60
Old Animal - Uniform Prior	19.0	
Old Animal - N(0, 1000) Prior	18.8	11.40 - 40.40
Old Animal - N(16, 36) Prior	17.9	12.37 - 25.20
Old Animal - N(0, 100) Prior	16.4	10.62 - 24.11
Old Animal - N(16, 100) Prior	18.4	11.60 - 28.41

In Table 4 posterior estimates of the age distributions of the two rhinoceros are given. For further details about the posterior densities see du Plessis and van der Merwe (1994).

Acknowledgements

The authors wish to express their gratitude to the referees and editor for their helpful comments in revising the paper.

REFERENCES

- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (2nd ed.), Wiley, New York.

- Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normality, *J. Roy. Statist. Soc. Ser. B*, **36**, 99–102.
- Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*, Wiley, New York.
- Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*, Addison Wesley, Massachusetts.
- Brown, P. J. (1982). Multivariate calibration, *J. Roy. Statist. Soc. Ser. B*, **46**(3), 287–321.
- Carlin, B. P. and Polson, N. G. (1991). Inference for non conjugate Bayesian models using the Gibbs sample, *Canad. J. Statist.*, **19**(4), 399–405.
- Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992). Hierarchical Bayes analysis of change point problems, *Applied Statistics*, **41**(2), 389–405.
- Clarke, G. P. Y. (1992). Inverse estimates from a multi-response model, *Biometrics*, **48**(4), 1081–1094.
- du Plessis, J. L. and van der Merwe, A. J. (1994). An example of non linear Bayesian calibration—estimating the age of rhinoceros, Tech. Report No. 212, Department of Mathematical Statistics, University of the Orange Free State, South Africa.
- Gelfand, A. E., Hills, S. E., Racine-Poon, A. and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling, *J. Amer. Statist. Assoc.*, **85**, 972–985.
- Gelfand, A. E. and Smith, A. F. M. (1991). Gibbs sampling for marginal posterior expectations, *Comm. Statist. Theory Methods*, **20**(5, 6) 1747–1766.
- Gelfand, A. E., Smith, A. F. M. and Lee, T. M. (1992). Bayesian analysis of constrained parameters and truncated data problems using Gibbs sampling, *J. Amer. Statist. Assoc.*, **87**, 523–532.
- Gelman, A. and Rubin, D. R. (1992). A single series from the Gibbs sampler provides a false sense of security, *Bayesian Statistics 4* (ed. J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith), 627–635, Oxford University Press, Oxford.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Hastings, W. K. (1970). Monte Carlo sample methods using Markov chain and their applications, *Biometrika*, **57**, 97–109.
- Hunter, W. G. and Lamboy, W. F. (1981). A Bayesian analysis of the linear calibration problem, *Technometrics*, **23**, 323–350.
- Krutchkoff, R. C. (1967). Classical and inverse regression methods of calibration, *Technometrics*, **9**, 425–439.
- Mitchell, A. (1967). Discussion of paper by I. J. Good, *J. R. Statist. Soc.*, **B29**, 423–424.
- Odell, P. L. and Feiveson, A. H. (1966). A numerical procedure to generate a sample covariance matrix, *J. Amer. Statist. Assoc.*, **61**, 198–203.
- Oman, S. D. and Wax, Y. (1984). Estimating fetal age by ultrasound measurements: An example of multivariate calibration, *Biometrics*, **40**, 947–960.
- Scheffé, H. (1973). A statistical theory of calibration, *Ann. Statist.*, **1**, 1–37.
- Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling-resampling perspective, *Amer. Statist.*, **46**(2), 84–88.
- Smith, R. L. and Corbett, M. (1987). Measuring marathon courses: An application of statistical calibration, *Applied Statistics*, **36**, 283–295.
- Stephens, D. A. and Smith, A. F. M. (1992). Sampling-resampling techniques for the computation of posterior densities in normal means problems, *Test*, **1**(1), 1–18.
- Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. and Gelfand, A. E. (1994). Bayesian analysis of linear and non-linear population models by using the Gibbs sampler, *Applied Statistics*, **43**(1), 201–221.
- Williams, E. J. (1969). A note on regression methods in calibration, *Technometrics*, **11**, 189–192.
- Ye, K. and Berger, J. O. (1991). Noninformative priors for inferences in exponential regression models, *Biometrika*, **78**(3), 645–656.