

BAYESIAN PREDICTION ANALYSIS FOR GROWTH CURVE MODEL USING NONINFORMATIVE PRIORS*

GWOWEN SHIEH AND JACK C. LEE

*Department of Management Science and Institute of Statistics, National Chiao Tung University,
Hsinchu, Taiwan 30050, R.O.C.*

(Received November 16, 1998; revised July 17, 2000)

Abstract. We apply a Bayesian approach to the problem of prediction in an unbalanced growth curve model using noninformative priors. Due to the complexity of the model, no analytic forms of the predictive densities are available. We propose both approximations and a prediction-oriented Metropolis-Hastings sampling algorithm for two types of prediction, namely the prediction of future observations for a new subject and the prediction of future values for a partially observed subject. They are illustrated and compared through real data and simulation studies. Two of the approximations compare favorably with the approximation in Fearn (1975, *Biometrika*, **62**, 89–100) and are very comparable to the more accurate Rao-Blackwellization from Metropolis-Hastings sampling algorithm.

Key words and phrases: Approximations, Metropolis-Hastings, posterior, random coefficient regression, Rao-Blackwellization.

1. Introduction

The growth curve model (hereafter referred to as the GC model) or generalized MANOVA model proposed by Potthoff and Roy (1964) is

$$(1.1) \quad E(Y_{p \times N}) = X_{p \times m} \tau_{m \times r} A_{r \times N},$$

where τ is unknown, X and A are the within-subject and between-subject design matrices with ranks $m < p$ and $r < N$, respectively. Furthermore the columns of Y are independently normally distributed with a common unknown covariance matrix Σ . The GC model has strong connection with the classical MANOVA model and has been studied intensively in the literature by Rao (1965, 1966), Khatri (1966), Grizzle and Allen (1969), Geisser (1970), Lee and Geisser (1972) and von Rosen (1991), among others. Yet many studies typically have unbalanced designs or missing data that make these standard multivariate procedures inapplicable. Hence it is of great interest and demand to have generalizations of GC model which allows unbalanced data either by designs or by chance, such as Kleinbaum (1973), Fearn (1975), Strenio *et al.* (1983), Reinsel (1985), Jennrich and Schluchter (1986), Vonesh and Carter (1987) and Carter *et al.* (1992).

In this paper we conduct a Bayesian prediction analysis for unbalanced GC model with random coefficient structure. Whereas Geisser (1970) and Lee and Geisser (1972)

*This research was partially supported by National Science Council Grant 87-2119-M-009-001 and was presented at the 1998 Taipei International Statistical Symposium.

considered the Bayesian prediction analysis of completely balanced GC model under the assumptions that covariance is either an arbitrary unknown positive definite matrix or of the Rao's simple structure. We propose both approximations and a method based on the Metropolis-Hastings (MH) sampling algorithm (Chib and Greenberg (1995)). We like to emphasize that the Bayesian formulation of the model considered here is closely related to the Bayesian hierarchical linear model discussed in Fearn (1975) and Gelfand *et al.* ((1990), Section 6). By comparison, our model is more general and includes theirs as special cases. For the Bayesian framework, we use noninformative priors but they assumed proper prior for coefficient parameters with vague values for the hyperparameters.

Instead of relying on all full conditionals, we simplify the joint conditional and reduce the conditional generations to two sets in the execution of Markov chain Monte Carlo (MCMC) methods. The prediction of future observations and estimation of parameters can still be conducted as other applications of MCMC methods, for example, Gelfand *et al.* ((1990), Section 6), Gilks *et al.* (1993) and Yang and Chen (1995). The MCMC methods are the state of the art and have been shown to make possible the use of flexible Bayesian methods that would otherwise be computationally difficult. It seems pointless to study approximations for Bayesian models when MCMC methods are applicable. In the sense of "enjoy the omelet without messing the kitchen," it is the motive to have useful approximations as long as the trade-off between accuracy and time is acceptable and especially when time is limited.

In Section 2 we describe the GC model in a form of hierarchical Bayes linear model to cover unbalanced cases. In Section 3 we discuss predictive densities and propose both approximations and the MH sampling algorithm for prediction purpose. We then present numerical results in Section 4 through real data and simulations to illustrate the performance of the proposed approaches. Finally in Section 5 some concluding remarks are made.

2. The model

Consider a random coefficient model

$$(2.1) \quad y_i = X_i \beta_i + \varepsilon_i,$$

where y_i is a $p_i \times 1$ vector of observations, X_i is the $p_i \times m$ ($m < p_i$) within-subject design matrix, β_i is the $m \times 1$ vector of random coefficients, and ε_i is the vector of errors for subject $i = 1, \dots, N$. Furthermore, we assume β_i and ε_i , $i = 1, \dots, N$, are independent and

$$(2.2) \quad \beta_i \sim N_m(\tau a_i, \Gamma),$$

$$(2.3) \quad \varepsilon_i \sim N_{p_i}(0, \sigma^2 I_{p_i}),$$

where τ is the $m \times r$ unknown parameter matrix of baseline covariates, a_i is the $r \times 1$ ($r < N$) between-subject design vector, and Γ and σ^2 are the positive definite covariance matrix and variance component, respectively. For the completely balanced case, $X_i = X$, one can write model (2.1)–(2.3) in the form of (1.1) with $\Sigma = X\Gamma X' + \sigma^2 I$.

To complete a Bayesian formulation of model (2.1)–(2.3), one needs to specify priors for (τ, Γ, σ^2) . We consider a Jeffrey's type of noninformative prior

$$(2.4) \quad [\tau, \Gamma^{-1}, \sigma^{-2}] \sim |\Gamma^{-1}|^{-(m+1)/2} (\sigma^{-2})^{-1}.$$

As remarked earlier, both Fearn (1975) and Gelfand *et al.* ((1990), Section 6) assumed the following prior for Γ^{-1} ,

$$(2.5) \quad [\Gamma^{-1}] \sim \text{Wishart}_m(R^{-1}, \rho)$$

with degrees of freedom $\rho \geq m$ and $m \times m$ precision matrix R . Note that both ρ and R need to be specified. They completed the model by setting a diagonal matrix of large enough components for diffuse purpose. From our experiences, there are no dramatic differences in using different diffuse priors, however it is one extra step that needs some insight and experimentation. It is interesting to note that they assumed the identical noninformative prior for the inverse of variance component σ^{-2} as in (2.4). It is obvious that the prior $|\Gamma^{-1}|^{-(m+1)/2}$ reduces to $(\sigma^{-2})^{-1}$ when $m = 1$ but (2.5) doesn't unless for a special value of ρ and a special matrix R with very large diagonal elements.

A usual parameterization of Γ is $\Gamma = \sigma^2 \Gamma^*$ by factoring out σ^2 from Γ . Accordingly the prior $[\Gamma^{*-1}] \sim |\Gamma^{*-1}|^{-(m+1)/2}$ and (2.4) is of the same form

$$(2.6) \quad [\tau, \Gamma^{*-1}, \sigma^{-2}] \sim |\Gamma^{*-1}|^{-(m+1)/2} (\sigma^{-2})^{-1}.$$

From the model (2.1)–(2.3) and (2.6), the joint conditional density of $(\tau, \Gamma^{*-1}, \sigma^{-2})$ given $y = (y'_1, y'_2, \dots, y'_N)'$ is

$$(2.7) \quad p(\tau, \Gamma^{*-1}, \sigma^{-2} | y) \propto |\Gamma^{*-1}|^{-(m+1)/2} (\sigma^{-2})^{-1} \\ \cdot \prod_{i=1}^N \{ |\Sigma_i^{-1}|^{1/2} \exp^{-\frac{1}{2}} (\hat{\beta}_i - \tau a_i)' \Sigma_i^{-1} (\hat{\beta}_i - \tau a_i) \} \\ \cdot \prod_{i=1}^N \{ |\sigma^{-2} (Z_i' Z_i)|^{1/2} \exp^{-\frac{1}{2}} (D_i y_i)' \sigma^{-2} (Z_i' Z_i) (D_i y_i) \},$$

where $\Sigma_i = X_i \Gamma X_i' + \sigma^2 I_{p_i}$, $\hat{\beta}_i = (X_i' X_i)^{-1} X_i' y_i$, Z_i is $p_i \times (p_i - m)$ such that $Z_i' X_i = 0$ and $D_i = (Z_i' Z_i)^{-1} Z_i'$. This follows from the pseudo-augmented model $E(Y_i) = (X_i, Z_i) \begin{pmatrix} \tau \\ 0 \end{pmatrix} a_i$, see Lee and Geisser ((1972), Section 2). Due to the complexity of the joint posterior density, it is of practical importance to know under what conditions that the resulting posterior is proper. In a manner analogous to Yang and Chen (1995), it can be shown that the joint posterior (2.7) is proper if $N_1 > m + r$, $\sum_{i=1}^N p_i > r$, and there exists at least one full rank design matrix X_i such that $y_i \neq X_i (X_i' X_i)^{-1} X_i' y_i$, where N_1 is the number of full rank design matrix for X_i , $i = 1, \dots, N$.

As usual, the parameter estimation of Bayesian models can't proceed much further without certain approximations due to intractable marginal conditionals of parameters. In our case the parameter Γ^{*-1} is the most troublesome. After integrating out τ and σ^{-2} in (2.7), a useful conditional (posterior) of Γ^{*-1} can be derived

$$(2.8) \quad p(\Gamma^{*-1} | y) \propto |\Gamma^{*-1}|^{-(m+1)/2} |W|^{-1/2} |A^* W^{-1} A^*|^{-1/2} [\hat{\beta}' W^* \hat{\beta} + d]^{-(p-mr)/2},$$

where $W = \text{diag}(W_i)$, a block-diagonal matrix with $m \times m$ blocks of $W_i = \Gamma^* + (X_i' X_i)^{-1}$, $A^* = A' \otimes I_m$, \otimes denotes the Kronecker product, $\hat{B} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_N)$, $\hat{\beta} = \text{vec}(\hat{B})$, $\text{vec}()$ is a matrix operator which arranges the columns of a matrix into one long column, $W^* = W^{-1} - W^{-1} A^* (A^* W^{-1} A^*)^{-1} A^* W^{-1}$, $d = \sum_{i=1}^N y_i' (I_{p_i} - X_i (X_i' X_i)^{-1} X_i') y_i$, and $p = \sum_{i=1}^N p_i$.

For the case of parameter estimation, one can derive some estimate (for example, unbiased estimate, MLE or posterior mode) of Γ^{*-1} , and substitute into the following conditionals for the estimation of τ and σ^{-2} :

$$(2.9) \quad p(\sigma^{-2} | y, \Gamma^{*-1}) \equiv \text{Gamma} \left(\frac{p - mr}{2}, \frac{2}{h} \right),$$

where $h = \hat{\beta}' W^* \hat{\beta} + d$; and

$$(2.10) \quad p(\lambda | y, \Gamma^{*-1}) \equiv T_{mr}(\lambda^*, \Omega, p),$$

a multivariate T distribution, with $\lambda = \text{vec}(\tau)$, $\lambda^* = (A^* W^{-1} A^*)^{-1} A^* W^{-1} \hat{\beta}$, $\Omega = \frac{h}{v} (A^* W^{-1} A^*)^{-1}$, $v = p - mr$. Note that a $t \times 1$ vector T is said to possess a multivariate T -distribution with p degrees of freedom, i.e., $T \sim T_t(\mu, \Sigma, p)$ if

$$f(T) = (\pi v)^{-t/2} \Gamma^{-1} \left(\frac{v}{2} \right) \Gamma \left(\frac{p}{2} \right) |\Sigma|^{-1/2} [1 + (T - \mu)' (v \Sigma)^{-1} (T - \mu)]^{-p/2},$$

where $\Gamma(\cdot)$ is the GAMMA function and $v = p - t$. Accordingly, $E(T) = \mu$ and $\text{Cov}(T) = \frac{v}{v-2} \Sigma$. It can be shown that λ^* is the generalized least squares estimator of λ for the model defined in (2.1)–(2.3) with known $\{\sigma^2, \Gamma^*\}$, see equation (2.2) of Vonesh and Carter (1986). The variance-covariance matrix of λ^* is $\sigma^2 (A^* W^{-1} A^*)^{-1}$. We note that under the model assumptions $\lambda^* | (\sigma^2, \Gamma^*) \sim N_{mr}(\lambda, \sigma^2 (A^* W^{-1} A^*)^{-1})$.

Although (2.8) is nonstandard, as long as one can simulate observations from (2.8), the estimate of Γ^{*-1} and the Rao-Blackwellization estimates of σ^{-2} and $\lambda = \text{vec}(\tau)$ can be obtained through (2.9) and (2.10) (Casella and George (1992)). We shall not pursue this here but the estimation of σ^{-2} and τ will become by-products in the process of prediction of future observations in next section.

3. Approximations and the MH sampling algorithm for predictive densities

We shall now discuss the problem of prediction. Two types of prediction are considered, namely the prediction of future observations for a new subject and the prediction of future values for a partially observed subject. Because of the complexity of the predictive densities, we propose both approximations and a prediction-oriented MH sampling algorithm. The former represents a crude but easy-to-implement method while the latter may be the most sophisticated method so far.

First, suppose one is interested in the prediction of a future $p_V \times 1$ observation vector V to be drawn from the model (2.1)–(2.3). Based on the Bayesian formulation in (2.6) and follow essentially the same algebra leading to (2.8), one gets the predictive density of V given y and Γ^{*-1} ,

$$(3.1) \quad p(V | y, \Gamma^{*-1}) \equiv T_{p_V}(\mu_V, \Omega_V, p + p_V - mr),$$

where $\mu_V = X_V (a'_V \otimes I_m) \lambda^*$, $\Omega_V = \frac{h}{v} C$ with $C = X_V (a'_V \otimes I_m) (A^* W^{-1} A^*)^{-1} (a_V \otimes I_m) X'_V + (X_V \Gamma^* X'_V + I_{p_V})$ and h and v are defined in (2.9) and (2.10), respectively. Note that μ_V is based on the generalized least squares estimator λ^* and $(a'_V \otimes I_m) \lambda^*$ can be interpreted as the predictor of individual β_V . Furthermore, the associated variance-covariance is $\frac{h}{v-2} C$ and $\frac{h}{v-2}$ can be viewed as the estimate of σ^2 with known Γ^* .

Likewise, for the case of predicting the future values $V_2(p_{V_2} \times 1)$ of a partially observed subject $V_1(p_{V_1} \times 1, p_{V_1} > m)$, one gets the conditional predictive density of V_2 given y, V_1 and Γ^{*-1}

$$(3.2) \quad p(V_2 | y, V_1, \Gamma^{*-1}) \equiv T_{p_{V_2}}(\mu_{V_2}, \Omega_{V_2}, p + p_V - mr),$$

where

$$\begin{aligned} \mu_{V_2} &= X_{V_2} \{ \Lambda [(X'_{V_1} X_{V_1})^{-1} + \Lambda]^{-1} \hat{\beta}_{V_1} + (X'_{V_1} X_{V_1})^{-1} [(X'_{V_1} X_{V_1})^{-1} + \Lambda]^{-1} \\ &\quad \cdot (a'_V \otimes I_m) \lambda^* \} \\ &= X_{V_2} \{ \Gamma^* W_{V_1}^{-1} \hat{\beta}_{V_1} + (X'_{V_1} X_{V_1})^{-1} W_{V_1}^{-1} (a'_V \otimes I_m) \lambda^*_{yV_1} \}, \\ \hat{\beta}_{V_1} &= (X'_{V_1} X_{V_1})^{-1} X'_{V_1} V_1, \quad \Lambda = \Gamma^* + (a'_V \otimes I_m) (A^* W^{-1} A^*)^{-1} (a_V \otimes I_m), \\ W_{V_1} &= \Gamma^* + (X'_{V_1} X_{V_1})^{-1}, \\ \lambda^*_{yV_1} &= (A^*_{yV_1} W_{yV_1}^{-1} A^*_{yV_1})^{-1} A^*_{yV_1} W_{yV_1}^{-1} \hat{\beta}_{yV_1} \quad \text{with} \quad \hat{\beta}_{yV_1} = \begin{pmatrix} \hat{\beta} \\ \hat{\beta}_{V_1} \end{pmatrix}, \end{aligned}$$

$A^*_{yV_1}$ and W_{yV_1} are direct extensions of A^* and W , respectively, by including V_1 as the $(N + 1)$ -th observation, $\Omega_{V_2} = \frac{h_{yV_1}}{v_{yV_1}} C_{22.1}$, $h_{yV_1} = \hat{\beta}'_{yV_1} W^*_{yV_1} \hat{\beta}_{yV_1} + d_{yV_1}$, $W^*_{yV_1}$ and d_{yV_1} are extensions of W^* and d in (2.9), respectively, to accommodate the $(N + 1)$ -th observation V_1 , $v_{yV_1} = p + p_{V_1} - mr$, $C_{22.1} = C_{22} - C_{21} C_{11}^{-1} C_{12} = X_{V_2} [(X'_{V_1} X_{V_1}) + \Lambda^{-1}]^{-1} X'_{V_2} + I_{p_{V_2}}$.

It can be shown that $[(X'_{V_1} X_{V_1}) + \Lambda^{-1}]^{-1} = (X'_{V_1} X_{V_1})^{-1} - (X'_{V_1} X_{V_1})^{-1} W_{V_1}^{-1} (X'_{V_1} X_{V_1})^{-1} + (X'_{V_1} X_{V_1})^{-1} W_{V_1}^{-1} (a'_V \otimes I_m) (A^*_{yV_1} W_{yV_1}^{-1} A^*_{yV_1})^{-1} (a_V \otimes I_m) W_{V_1}^{-1} (X'_{V_1} X_{V_1})^{-1}$, which is quite useful for deriving different forms of μ_{V_2} shown above. Accordingly, μ_{V_2} is constructed from X_{V_2} and the individual predictor of β_V , which is a weighted average of $\hat{\beta}_{V_1}$ and λ^* or $\hat{\beta}_{V_1}$ and $\lambda^*_{yV_1}$. For the estimation of variance components, it is important to note that the associated variance-covariance is $\frac{h_{yV_1}}{v_{yV_1}-2} C_{22.1}$ and again $\frac{h_{yV_1}}{v_{yV_1}-2}$ can be viewed as the estimate of σ^2 with known Γ^* . In general, it agrees with that of Bondeson (1990) and Vonesh and Chinchilli ((1997), Section 6.2.4) except that they were based on the conditional distribution of $\{\sigma^2, \Gamma^*\}$ while here it is relied on the conditional distribution of Γ^* alone. Hence the only difference is that σ^2 is replaced by proper estimate in terms of Γ^* .

3.1 Approximations

As the estimation of parameters discussed in the previous section, there is no closed or easily recognizable predictive densities of $P(V | y)$ and $P(V_2 | y, V_1)$. But one can substitute Γ^{*-1} with some sensible estimate $\hat{\Gamma}^{*-1}$ given y or given y and V_1 into (3.1) and (3.2). Then one has the approximate predictive densities

$$(3.3) \quad p(V | y, \hat{\Gamma}^{*-1}) \equiv T_{p_V}(\hat{\mu}_V, \hat{\Omega}_V, p + p_V - mr),$$

and

$$(3.4) \quad p(V_2 | y, V_1, \hat{\Gamma}^{*-1}) \equiv T_{p_{V_2}}(\hat{\mu}_{V_2}, \hat{\Omega}_{V_2}, p + p_V - mr).$$

For the choices of $\hat{\Gamma}^{*-1}$, the posterior modes of (2.8) and its extensions $p(\Gamma^{*-1} | y, V_1)$ are possible candidates. However it will involve some iterative procedures inevitably. To keep the approximations as simple as possible, we turn to non-iterative solutions:

the unbiased (Vonesh and Carter (1987)) and empirical Bayes (EB) estimators (Reinsel (1985) and Rao (1987)). Hence the predictive density $P(V | y)$ is approximated by

$$(3.5) \quad T_{UB}(V | y) \equiv T_{p_V}(\hat{\mu}_V, \hat{\Omega}_V, p + p_V - mr),$$

with $\hat{\Gamma}^{*-1} = \hat{\Gamma}_{UB}^{*-1} = \hat{\sigma}_{UB}^2 \hat{\Gamma}_{UB}^{-1}$, $\hat{\Gamma}_{UB} = \frac{1}{N-r} [S_y - \sum_{i=1}^N (1 - a_i'(AA')^{-1}a_i) \hat{\sigma}_{UB}^2 (X_i'X_i)^{-1}]$, $S_y = \hat{B}(I_N - A'(AA')^{-1}A)\hat{B}'$, $\hat{\sigma}_{UB}^2 = d/k$, and $k = \sum_{i=1}^N (p_i - m)$, or

$$(3.6) \quad T_{EB}(V | y) \equiv T_{p_V}(\hat{\mu}_V, \hat{\Omega}_V, p + p_V - mr),$$

with $\hat{\Gamma}^{*-1} = \hat{\Gamma}_{EB}^{*-1} = \hat{\sigma}_{EB}^2 \hat{\Gamma}_{EB}^{-1}$, $\hat{\Gamma}_{EB} = \frac{1}{N-r-m-1} [S_y - \sum_{i=1}^N (1 - a_i'(AA')^{-1}a_i) \hat{\sigma}_{EB}^2 (X_i'X_i)^{-1}]$ and $\hat{\sigma}_{EB}^2 = d/(k+2)$.

Meanwhile, the conditional predictive density $P(V_2 | y, V_1)$ is approximated by

$$(3.7) \quad T_{UB}(V_2 | y, V_1) \equiv T_{p_{V_2}}(\hat{\mu}_{V_2}, \hat{\Omega}_{V_2}, p + p_V - mr),$$

with $\hat{\Gamma}^{*-1} = \hat{\Gamma}_{UB}^{*-1} = \hat{\sigma}_{UB}^2 \hat{\Gamma}_{UB}^{-1}$, $\hat{\Gamma}_{UB} = \frac{1}{N+1-r} [S_{yV_1} - \sum_{i=1}^{N+1} (1 - a_i'(HH')^{-1}a_i) \hat{\sigma}_{UB}^2 (X_i'X_i)^{-1}]$, $S_{yV_1} = (\hat{B}, \hat{\beta}_{V_1})(I_{N+1} - H'(HH')^{-1}H)(\hat{B}, \hat{\beta}_{V_1})'$, $H = (A, a_{V_1})$, $\hat{\sigma}_{UB}^2 = d_{yV_1}/k_2$, $k_2 = \sum_{i=1}^{N+1} (p_i - m)$, or

$$(3.8) \quad T_{EB}(V_2 | y, V_1) \equiv T_{p_{V_2}}(\hat{\mu}_{V_2}, \hat{\Omega}_{V_2}, p + p_V - mr),$$

with $\hat{\Gamma}^{*-1} = \hat{\Gamma}_{EB}^{*-1} = \hat{\sigma}_{EB}^2 \hat{\Gamma}_{EB}^{-1}$, $\hat{\Gamma}_{EB} = \frac{1}{N+1-r-m-1} [S_{yV_1} - \sum_{i=1}^{N+1} (1 - a_i'(HH')^{-1}a_i) \hat{\sigma}_{EB}^2 (X_i'X_i)^{-1}]$, $\hat{\sigma}_{EB}^2 = d_{yV_1}/(k_2 + 2)$.

For the $T_{UB}(V | y)$ and $T_{UB}(V_2 | y, V_1)$ approximations, $\hat{\Gamma}_{UB}$ is an unbiased estimator of Γ (Reinsel (1985)) and is termed the residual method of moments (RMM) in Vonesh and Chinchilli ((1997), Section 6.2). It should be noted that this estimator is similar to the restricted maximum likelihood estimation in that it corrects for small sample bias. In fact, for balanced and complete data ($X_i = X$), it can be shown that $\hat{\sigma}_{UB}^2$ and $\hat{\Gamma}_{UB}$ are the REML estimates. For unbalanced data, $\hat{\Gamma}_{UB}$ is unbiased and consistent. It has properties similar to the REML estimates in that it is based on a set of linearly independent error contrasts $(I_N - A'(AA')^{-1}A)\hat{B}'$ or $(I_{N+1} - H'(HH')^{-1}H)(\hat{B}, \hat{\beta}_{V_1})'$, see Vonesh and Carter (1987) and Vonesh and Chinchilli ((1997), Section 6.2) for further details.

Next, for the $T_{EB}(V | y)$ and $T_{EB}(V_2 | y, V_1)$ approximations, $\hat{\Gamma}_{EB}$ is constructed from $S_y/(N - r - m - 1)$ and $d/(k + 2)$ or $S_{yV_1}/(N + 1 - r - m - 1)$ and $d_{yV_1}/(k_2 + 2)$. As discussed in Reinsel (1985), it is not possible to give a closed form for the optimal value of the "shrinking constant" in the unbalanced case. This is solely motivated by the optimal property of EB estimates for balanced data.

However, on occasion both $\hat{\Gamma}_{UB}$ and $\hat{\Gamma}_{EB}$ can result in non-positive semi-definite estimate of Γ . To ensure having a positive semi-definite estimate of Γ , we make the adjustment suggested in Vonesh and Carter (1987) whenever necessary. Furthermore, Vonesh and Carter (1987) focused only on the problem of estimation and proposed two large sample tests for linear hypothesis testing of parameter λ . Nevertheless, instead of extending their results for prediction, Vonesh and Chinchilli ((1997), Section 6.2.4) proposed a t approximation for the prediction of single future value for a partially observed subject. This is motivated by knowledge that, when data are balanced and complete, it

has an exact t distribution. Here we extend it to accommodate simultaneous predictions of several future values for both types of prediction:

$$(3.9) \quad T_{VC}(V | y) \equiv T_{p_V}(\hat{\mu}_{V2.VC}, \hat{\Omega}_{V2.VC}, N - r + p_V),$$

where $\hat{\mu}_{V2.VC}$ is the same as $\hat{\mu}_V$ in (3.5), $\hat{\Omega}_{V2.VC} = \frac{N-r-2}{N-r} \hat{\sigma}_{UB}^2 \{X_{V2}[(X'_{V1}X_{V1}) + \Lambda]^{-1}X'_{V2} + I_{p_{V2}}\}$ with $\Lambda = \Lambda(\hat{\Gamma}^*)$, $\hat{\Gamma}^* = \hat{\Gamma}_{UB}/\hat{\sigma}_{UB}^2$, $\hat{\Gamma}_{UB}$ and $\hat{\sigma}_{UB}^2$ are defined in (3.5); and

$$(3.10) \quad T_{VC}(V_2 | y, V_1) \equiv T_{p_{V2}}(\hat{\mu}_{V2.VC}, \hat{\Omega}_{V2.VC}, N + 1 - r + p_{V2}),$$

where $\hat{\mu}_{V2.VC}$ is the same as $\hat{\mu}_{V2}$ in (3.7), $\hat{\Omega}_{V2.VC} = \frac{N+1-r-2}{N+1-r} \hat{\sigma}_{UB}^2 \{X_{V2}[(X'_{V1}X_{V1}) + \Lambda]^{-1}X'_{V2} + I_{p_{V2}}\}$ with $\Lambda = \Lambda(\hat{\Gamma}^*)$, $\hat{\Gamma}^* = \hat{\Gamma}_{UB}/\hat{\sigma}_{UB}^2$, $\hat{\Gamma}_{UB}$ and $\hat{\sigma}_{UB}^2$ are defined in (3.7). Essentially, T_{UB} and T_{VC} are very much alike. Both give identical predicted values, while the only differences are in the degrees of freedom and estimates of σ^2 .

In the vein of Bayesian prediction, normal approximation was proposed in Fearn (1975) for all $a_i = 1$. His proposed estimate of Γ is $(S_y + R)/(N + \rho - m - 2)$ or $(S_{yV1} + R)/(N + 1 + \rho - m - 2)$ depending on the type of prediction. Hence they are not corrected for the factor $\frac{N-1}{N} \sum_{i=1}^N \hat{\sigma}^2(X'_iX_i)^{-1}$ or $\frac{N}{N+1} \sum_{i=1}^{N+1} \hat{\sigma}^2(X'_iX_i)^{-1}$ as other aforementioned approximations. However, Fearn's approximation serves as a convenient benchmark to assessing the performance of the proposed T approximations and the MH sampling algorithm discussed next.

3.2 The MH sampling algorithm

Gelfand *et al.* (1990) showed the applications of Gibbs sampling in a wide range of normal data models, including the hierarchical growth curve mentioned earlier. The Bayesian marginal posterior and predictive densities can be extracted from the full conditional distributions after successively simulating observations from $[\beta | y, \tau, \Gamma^{-1}, \sigma^2]$, $[\tau | y, \beta, \Gamma^{-1}, \sigma^2]$, $[\Gamma^{-1} | y, \tau, \beta, \sigma^2]$, and $[\sigma^2 | y, \beta, \tau, \Gamma^{-1}]$. In particular, the predictive density of future observations V given τ and σ^2 is $N_{p_V}(X_V\tau a_V, \sigma^2 I_{p_V})$. Then the Rao-Blackwellized density and prediction of V can be obtained from the finite mixture (average) density of $N_{p_V}(X_V\tau_i a_V, \sigma_i^2 I_{p_V})$ from all stationary (τ_i, σ_i^2) .

In the same spirit of the Rao-Blackwellization, we turn things around and propose simplified sampling scheme by successively simulating only two sets of conditionals, namely $[V | y, \Gamma^{*-1}]$ and $[\Gamma^{*-1} | y, V]$, and $[V_2 | y, V_1, \Gamma^{*-1}]$ and $[\Gamma^{*-1} | y, V_1, V_2]$, for the prediction of V and V_2 , respectively. After the execution of the MH sampling scheme, the predictive density and related prediction of V and V_2 and the densities of parameters $(\Gamma^{*-1}, \sigma^2, \tau)$ and their estimation can be obtained through the Rao-Blackwellization of conditionals defined in (3.1), (3.2), (2.8), (2.9) and (2.10), respectively.

The sampling procedure proceeds as follows:

0. Choose an initial value of Γ^{*-1} .
1. Generate V given y and Γ^{*-1} from (3.1) or generate V_2 given y , V_1 and Γ^{*-1} from (3.2).
2. Generate Γ^{*-1} given y and V in the form of (2.8) using the MH algorithm with independence chain (Chib and Greenberg (1995), Section 5) and a candidate generating density $\text{Wishart}_m(R^{-1}, N + 1 - r)$, where $R = \frac{1}{N+1-r} [S_{yV1} - \sum_{i=1}^{N+1} (1 - a'_i(HH')^{-1}a_i)s^2(X'_iX_i)^{-1}]$ and $s^2 = d_{yV1}/k_2$.

In order to ensure the samples are drawn from the entire domains of the densities of V given y and V_2 given y and V_1 , Gelman and Rubin (1992) suggested using "overdispersed" initial values in multiple chains and assessing the convergence by using the

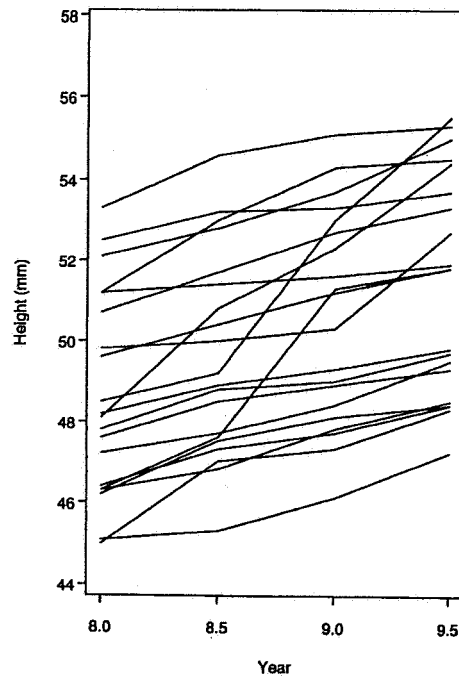


Fig. 1. Ramus heights of 20 boys.

potential scale reduction measure $\sqrt{\hat{R}}$. Following these guidelines we implement the sampling procedure using seven chains with seven different starting values of Γ^{*-1} and V_2 . The starting values for V_2 are the MLE and six others obtained from $\pm 2, \pm 4, \pm 6$ times of standard deviation. The starting values for Γ^{*-1} are generated from a Wishart distribution with mean $\hat{\Gamma}^{*-1}/(N+1-r)$ and degrees of freedom $N+1-r$, where $\hat{\Gamma}^{*-1}$ is the MLE of Γ^{*-1} . After a sufficiently long burn-in iterations in each chain we then use the remaining samples as simulated from the densities of V given y and V_2 given y and V_1 .

4. Numerical Results

In this section we apply the results in Section 3 to analyze both real and simulated data. For the purpose of illustration, only the conditional prediction is discussed, i.e., the prediction of future values V_2 given y and V_1 .

4.1 Ramus-bone data

This data set was first analyzed by Elston and Grizzle (1962) and consists of measurements of the ramus height (in mm) on 20 boys for ages 8, 8.5, 9 and 9.5 (See Fig. 1). Lee and Geisser (1975), Fearn (1975), Geisser (1981) and Lee (1988) also analyzed the data set. Since the observations are equally spaced in time, the within-subject design matrix can have the alternative form $X' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 \end{pmatrix}$. The between-subject design matrix A is a 1×20 vector consisting of all 1s. We continue the cross-validation analysis in Lee and Geisser (1975) and Fearn (1975) for predicting the last value of a partially observed vector by withholding one vector y_j as V and using the rest for predicting the

Table 1. The MAD and MSD between the predicted and the actual values of final measurements: ramus-bone data.

	T_{EB}	T_{UB}	Fearn	T_{VC}	ILS	MH	MHRB
MAD	0.5798	0.5885	0.5731	0.5885	0.5650	0.5835	0.5810
MSD	0.5643	0.5622	0.5601	0.5622	0.6405	0.5645	0.5618

last component V_2 of $V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$ and repeating this for all vectors y_j , $j = 1, \dots, 20$. In Fearn (1975) the hyperparameters in (2.5) were set as $R = \begin{pmatrix} 2 & 0 \\ 0 & 0.2 \end{pmatrix}$ and $\rho = m = 2$. His approximation is compared with the T approximations $T_{UB}(df = p + p_V - mr = 78)$, $T_{EB}(df = p + p_V - mr = 78)$ and $T_{VC}(df = N + 1 - r + p_{V2} = 20)$ defined in (3.7), (3.8) and (3.10), respectively. Furthermore, the mean of V_{2i} (MH) and conditional mean $\mu_{V2} | \Gamma_i^{*-1}$ of (3.2) (MHRB) represent the predicted values of V_2 from the presumably more accurate MH sampling algorithm, where V_{2i} and Γ_i^{*-1} are the last 500 samples of seven chains each with length 1500, with a total of 3500 samples. Besides, the predictor based on individual least squares estimator is also included since its prediction property is exactly known. It is denoted by $\hat{\mu}_{V2.ILS}$ (ILS) where

$$\hat{\mu}_{V2.ILS} = X_{V2}\hat{\beta}_{V1} \sim N_{p_{V2}}(X_{V2}\beta_V, \sigma^2\{X_{V2}(X'_{V1}X_{V1})^{-1}X'_{V2} + I_{p_{V2}}\})$$

with estimated variance-covariance $\hat{\sigma}_{UB}^2\{X_{V2}(X'_{V1}X_{V1})^{-1}X'_{V2} + I_{p_{V2}}\}$.

The following two discrepancy measures are calculated:

$$\text{MAD} = \sum_{j=1}^N |y_{pj} - \hat{y}_{pj}|/N \quad \text{and} \quad \text{MSD} = \sum_{j=1}^N (y_{pj} - \hat{y}_{pj})^2/N,$$

where y_{pj} and \hat{y}_{pj} are the p -th actual and predicted values of subject j . The results are listed in Table 1.

It is noted that the values of MAD and MSD of Fearn's approximation in Table 1 don't completely agree with the numbers (MAD = 0.5725 and MSD = 0.5590) in Table 3 of Fearn (1975). We believe that the discrepancy is due to individual rounding effects. The results are somewhat surprising because even predictors from the MH sampling algorithm, MH and MHRB, do not produce the smallest MAD and MSD. The individual least squares predictor has the minimum of MAD (0.5650), whereas Fearn's approximation has the minimum MSD (0.5601). Possible explanation of this may be found in Lee (1988). With the AR(1) covariance assumption, the values of MAD and MSD are 0.5638 and 0.5178, respectively. This indicates that there is a better fit of the data set than the random coefficient structure assumed here in terms of prediction. To have a profound study of these methods, we now turn to simulations.

4.2 Simulations

We will first extend the analysis of ramus-bone data by utilizing its data structure and the assumed parameters of τ , Γ , σ^2 to simulate one data set for more convincing comparisons, see Fig. 2. The MLEs are $\tau = \begin{pmatrix} 50.0750 \\ 0.4665 \end{pmatrix}$, $\Gamma = \begin{pmatrix} 6.2164 & 0.0899 \\ 0.0899 & 0.0752 \end{pmatrix}$ and $\sigma^2 = 0.1935$. For this simulated data, the prediction results are presented in Table 2.

From Table 2 it can be observed that the Rao-Blackwellized predictor MHRB has smallest MAD = 0.4414 and its MSD = 0.3016 is slightly larger than the MSD (0.2995)

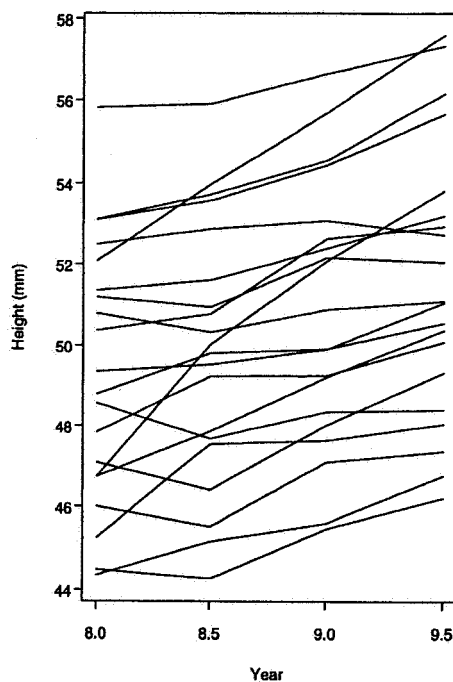


Fig. 2. Simulated ramus heights of 20 boys.

Table 2. The MAD and MSD between the predicted and the actual values of final measurements: simulated data based on ramus-bone data structure.

	T_{EB}	T_{UB}	Fearn	T_{VC}	ILS	MH	MHRB
MAD	0.4476	0.4418	0.4536	0.4418	0.5196	0.4441	0.4414
MSD	0.3052	0.2995	0.3092	0.2995	0.4067	0.3039	0.3016

Table 3. The coverage probabilities of prediction intervals.

	T_{EB}	T_{UB}	Fearn	T_{VC}	ILS	MHRB	
Nominal level	90%	0.8860 (0.0054)	0.8977 (0.0051)	0.8889 (0.0053)	0.9051 (0.0050)	0.9069 (0.0049)	0.8963 (0.0052)
	95%	0.9389 (0.0040)	0.9457 (0.0038)	0.9394 (0.0040)	0.9565 (0.0034)	0.9543 (0.0035)	0.9489 (0.0037)
	99%	0.9851 (0.0020)	0.9877 (0.0019)	0.9851 (0.0020)	0.9911 (0.0016)	0.9917 (0.0015)	0.9880 (0.0018)

of T_{UB} . The “averaged” predictor of MH samples, with MAD = 0.4441 and MSD = 0.3039, is slightly inferior. As expected, ILS is the worst with the largest MAD = 0.5196 and MSD = 0.4067. Among the four approximations, the T_{UB} and T_{VC} approximations are the best (MAD = 0.4438 and MSD = 0.3023) and are very comparable to MH and MHRB. However Fearn’s approximation has the largest MAD = 0.4536 and MSD =

Table 4. The MAD, MSD and coverage probabilities for predicting the last two components $V_2(2 \times 1) = \begin{pmatrix} V_{2.1} \\ V_{2.2} \end{pmatrix}$ with $N = 10$ and $X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 & 5 \end{pmatrix}$.

		T_{EB}	T_{UB}	Fearn	T_{VC}	ILS
MAD	$V_{2.1}$	0.6100	0.6080	0.6093	0.6080	0.6451
	$V_{2.2}$	0.8007	0.7962	0.7982	0.7962	0.8598
MSD	$V_{2.1}$	0.5862	0.5849	0.5841	0.5849	0.6499
	$V_{2.2}$	1.0010	0.9961	0.9940	0.9961	0.1450
Coverage Probabilities for $V_{2.1}$						
Nominal level	90%	0.8714 (0.0047)	0.8892 (0.0044)	0.8746 (0.0047)	0.9088 (0.0041)	0.8860 (0.0045)
	95%	0.9284 (0.0036)	0.9408 (0.0033)	0.9306 (0.0036)	0.9600 (0.0028)	0.9432 (0.0033)
	99%	0.9788 (0.0020)	0.9832 (0.0018)	0.9790 (0.0020)	0.9952 (0.0010)	0.9822 (0.0019)
Coverage Probabilities for V_2						
Nominal level	90%	0.8608 (0.0049)	0.8878 (0.0045)	0.8636 (0.0049)	0.9194 (0.0038)	0.8788 (0.0046)
	95%	0.9216 (0.0038)	0.9380 (0.0034)	0.9196 (0.0038)	0.9672 (0.0025)	0.9338 (0.0035)
	99%	0.9776 (0.0021)	0.9852 (0.0017)	0.9750 (0.0022)	0.9974 (0.0007)	0.9810 (0.0019)

0.3092. The T_{EB} approximation is slightly better than Fearn's with MAD = 0.4476 and MSD = 0.3052.

Next we compare approximate and the Rao-Blackwellized prediction intervals for the 20th subject by calculating the coverage probabilities of the 3500 simulated samples of V_2 from our MH sampling procedure. The results of three nominal coverage probabilities are listed in Table 3. The numbers in parentheses are the estimated standard errors. Among the six competitors, MHRB and T_{UB} are the two with closest coverage probabilities to the nominal level. The two frequentist approaches T_{VC} and ILS are also very good in achieving the nominal level, whereas T_{EB} and Fearn's approximation are slightly inferior.

Based on the analysis so far, the Rao-Blackwellization of the MH sampling algorithm performs consistently well in prediction accuracy and coverage probability. However, with the trade-off between accuracy and time involved, the approximations are quite reasonable and worth further investigation in order to make a solid conclusion.

The final comparisons solely focus on the performance among approximations. Utilizing the structure of ramus-bone data (with different numbers of subjects $N = 10$ and $N = 20$), we withhold only the last vector as V in each simulated data set and consider the prediction of the last two values for 5,000 iterations. Now the MAD, MSD and coverage probability are the averages of 5,000 absolute deviations, squared deviations and 0-1 values.

The comparisons of approximate prediction accuracy and coverage probabilities for $V_2(2 \times 1) = \begin{pmatrix} V_{2.1} \\ V_{2.2} \end{pmatrix}$ given y and V_1 with $N = 10$ are shown in Table 4 where $X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 & 5 \end{pmatrix}$ and $X'_{V_2} = \begin{pmatrix} 1 & 1 \\ 3 & 5 \end{pmatrix}$. The degrees of freedom for T_{EB} , T_{UB} and T_{VC} are 46, 46 and 9, respectively. From Table 4, it can be seen easily that the performance of the approximations is very much comparable in terms of MAD and MSD. Note that

Table 5. The MAD, MSD and coverage probabilities for predicting the last two components $V_2(2 \times 1) = \begin{pmatrix} V_{2.1} \\ V_{2.2} \end{pmatrix}$ with $N = 10$ and $X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 8 & 8.5 & 9 & 9.5 & 10 \end{pmatrix}$.

		T_{EB}	T_{UB}	Fearn	T_{VC}	ILS
MAD	$V_{2.1}$	0.4753	0.4635	0.4857	0.4635	0.6451
	$V_{2.2}$	0.5495	0.5284	0.5705	0.5284	0.8598
MSD	$V_{2.1}$	0.3592	0.3415	0.3742	0.3415	0.6499
	$V_{2.2}$	0.4839	0.4425	0.5177	0.4425	1.1450
Coverage Probabilities for $V_{2.1}$						
Nominal level	90%	0.8646 (0.0048)	0.8820 (0.0046)	0.8874 (0.0045)	0.9050 (0.0041)	0.8860 (0.0045)
	95%	0.9234 (0.0038)	0.9348 (0.0035)	0.9358 (0.0035)	0.9568 (0.0029)	0.9432 (0.0033)
	99%	0.9788 (0.0020)	0.9846 (0.0017)	0.9824 (0.0019)	0.9962 (0.0009)	0.9822 (0.0019)
Coverage Probabilities for V_2						
Nominal level	90%	0.8414 (0.0052)	0.8696 (0.0048)	0.8750 (0.0047)	0.9100 (0.0040)	0.8788 (0.0046)
	95%	0.9074 (0.0041)	0.9268 (0.0037)	0.9298 (0.0036)	0.9640 (0.0026)	0.9338 (0.0035)
	99%	0.9754 (0.0022)	0.9832 (0.0018)	0.9814 (0.0019)	0.9978 (0.0007)	0.9810 (0.0019)

the expected MSD of ILS for $V_{2.1}$ and $V_{2.2}$ are 0.6450 and 1.12875, respectively. They are very close to the simulated values 0.6499 and 1.1450, respectively.

For the comparison of coverage probabilities, T_{UB} and T_{VC} are considerably better than T_{EB} and Fearn's approximation for all three nominal levels. Nevertheless, T_{UB} tends to be lower than the nominal level while T_{VC} tends to be higher.

As mentioned in Subsection 3.1, there is some potential bias in Fearn's estimate of Γ . The sensitivity of such estimate in terms of different design matrices is not known. Therefore we set the design matrix as $X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 8 & 8.5 & 9 & 9.5 & 10 \end{pmatrix}$ to reflect the actual age scale and rerun the simulation with the identical seed. The results are reported in Table 5. Since ILS is the only one that is invariant to such transformation, it produces exactly the same outcomes as in Table 4. But all the other predictors turn to have smaller MAD and MSD. Comparatively, Fearn's method has much larger MAD and MSD for predicting both components of V_2 than the other three T approximations. On the contrary, Fearn's coverage probabilities become slightly better. However both T_{EB} and T_{UB} have lower simultaneous coverage probability for V_2 at the 90% nominal level and nearly the same as in Table 4 for other cases. Overall T_{VC} is very consistent in coverage probabilities. In general, the situation in simulation with $N = 20$ is quite similar only with smaller discrepancy in every respect due to larger sample size. Hence they are not reported here.

Based on these findings, the differences among the approximations are small, but it shows a clear pattern that the two approaches with RMM estimate of Γ^* , namely T_{UB} and T_{VC} , perform better than the other two approximations. They will be useful in case the more accurate and more involved Rao-Blackwellization from the MH sampling algorithm is not available.

5. Concluding Remarks

In this article we consider a Bayesian approach to the problem of prediction in an unbalanced growth curve model using noninformative priors. Three approximations and the Metropolis-Hastings sampling algorithm are proposed. Two of the multivariate T approximations using RMM estimate of Γ^* are shown to be better than the normal approximation proposed in Fearn (1975), and are slightly outperformed by the Rao-Blackwellization from the Metropolis-Hastings sampling algorithm. The proposed prediction-oriented Metropolis-Hastings sampling algorithm is efficient, accurate, and more simplified than the other Markov chain Monte Carlo methods in similar studies. Although posterior distributions, parameter estimation, predictive densities and related prediction can be conducted through the Rao-Blackwellization of the proposed Metropolis-Hastings sampling algorithm, they are computationally intensive. By contrast, all the computations in the approximations are non-iterative and avoid numerical integration.

REFERENCES

- Bondeson, J. (1990). Prediction in random coefficient regression models, *Biometrical J.*, **32**, 387–405.
- Carter, R. L., Resnick, M. B., Ariet, M., Shieh, G. and Vonesh, E. F. (1992). A random coefficient growth curve analysis of mental development in low-birth-weight infants, *Statistics in Medicine*, **11**, 243–256.
- Casella, G. and George, E. (1992). Explaining the Gibbs sampler, *Amer. Statist.*, **46**, 167–174.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm, *Amer. Statist.*, **49**, 327–335.
- Elston, R. C. and Grizzle, J. E. (1962). Estimation of time response curves and their confidence band, *Biometrics*, **18**, 148–159.
- Fearn, T. (1975). A Bayesian approach to growth curves, *Biometrika*, **62**, 89–100.
- Geisser, S. (1970). Bayesian analysis of growth curves, *Sankhyā Ser. A*, **32**, 53–64.
- Geisser, S. (1981). Sample reuse procedures for prediction of the unobserved portion of a partially observed vector, *Biometrika*, **68**, 243–250.
- 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.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences, *Statist. Sci.*, **7**, 457–472.
- Gilks, W. R., Wang, C. C., Yvonnet, B. and Cuursaget, P. (1993). Random-effects models for longitudinal data using Gibbs sampling, *Biometrics*, **49**, 441–453.
- Grizzle, J. E. and Allen, D. M. (1969). Analysis of growth and response curves, *Biometrics*, **25**, 357–381.
- Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices, *Biometrics*, **42**, 805–820.
- Khatri, C. G. (1966). A note on MANOVA model applied to problems in growth curve, *Ann. Inst. Statist. Math.*, **18**, 75–86.
- Kleinbaum, D. G. (1973). A generalization of the growth curve model which allows missing data, *J. Multivariate Anal.*, **3**, 117–124.
- Lee, J. C. (1988). Prediction and estimation of growth curves with special covariance structures, *J. Amer. Statist. Assoc.*, **83**, 432–440.
- Lee, J. C. and Geisser, S. (1972). Growth curve prediction, *Sankhyā Ser. A*, **34**, 393–412.
- Lee, J. C. and Geisser, S. (1975). Applications of growth curve prediction, *Sankhyā Ser. A*, **37**, 239–256.
- Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika*, **51**, 313–326.
- Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves, *Biometrika*, **52**, 447–458.

- Rao, C. R. (1966). Covariance adjustment and related problems in multivariate analysis, *Multivariate Analysis*, Vol. 1, (ed. P. R. Krishnaiah), 87–103, Academic Press, New York.
- Rao, C. R. (1987). Prediction of future observations in growth curve models, *Statist. Sci.*, **2**, 434–471.
- Reinsel, G. C. (1985). Mean squared error properties of empirical Bayes estimators in a multivariate random effects general linear model, *J. Amer. Statist. Assoc.*, **80**, 642–650.
- Strenio, J. F., Weisberg, H. I. and Bryk, A. S. (1983). Empirical Bayes estimation of individual growth-curve parameters and their relationship to covariates, *Biometrics*, **39**, 71–86.
- von Rosen, D. (1991). The growth curve model: A review, *Comm. Statist., Theory Methods*, **20**, 2791–2822.
- Vonesh, E. F. and Carter, R. L. (1987). Efficient inference for a random coefficient growth curve model with unbalanced data, *Biometrics*, **43**, 617–628.
- Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*, Marcel Dekker, New York.
- Yang, R. and Chen, M. H. (1995). Bayesian analysis for random coefficient regression models using noninformative priors, *J. Multivariate Anal.*, **55**, 283–311.