# On the Prediction of Growth Curves

by

Jack C. Lee and Seymour Geisser
National Chiao Tung University and University of Minnesota

# On the Prediction of Growth Curves

Jack C. Lee*
National Chiao Tung University

Seymour Geisser*
University of Minnesota

## Abstract

The subject of this paper is on generalized linear growth curve models. In the past, these types of growth curves have been studied by earlier authors including J. Wishart, G.E.P. Box and C.R. Rao. Further attention was directed towards this subject ever since the publication of the seminal paper on the subject by R.F. Potthoff and S.N. Roy in 1964 which modeled the growth curve as a generalized multivariate linear model. However, the emphasis up until 1970 was on the estimation and testing of parameters. In this paper we mainly review work on prediction within the context of these growth models over the last 25 years.

# 1 Introduction

The subject of this paper is generalized linear growth curve models. In the past, these types of growth curves have been studied by many authors including Wishart (1938), Box (1950) and Rao (1958). Further attention was directed towards this subject since the publication of the seminal paper of Potthoff and Roy (1964) that modeled the growth curve as a generalized multivariate linear model. However, all of the attention up until Geisser (1970) was directed towards estimation and testing parameters. In this paper we mainly review work on prediction within the context of these growth curve models over the last 25 years.

In Section 2 we shall briefly review the estimation of parameters of growth curve models. In subsequent sections we turn our attention to various problems associated with prediction.

# 2 Covariance Structures and Parameter Estimation

Let $Y_i$ be a $p_i \times 1$ vector of observations and

$$Y_i = X_i \beta_a + \epsilon_i \tag{2.1}$$

for $i = 1, \ldots, n_a^*$; $n_a^* = n_1 + \cdots + n_a$; $a = 1, \ldots, r$

where $X_i$ is a known $p_i \times m$ design matrix, $\beta_a$ is an $m \times 1$ vector of unknown regression parameters, and $\epsilon_i$ is a $p_i \times 1$ vector of errors that are independently distributed as $N(.; 0, \Omega_i)$. Let $\Sigma_i$ be the covariance matrix of $Y_i$, then the form of $\Sigma_i$ depends on the assumptions regrading $\Omega_i$ as well as $\beta_a$. For example, in the fixed effects model, $\beta_a$ is assumed fixed but unknown, then $\Sigma_i = \Omega_i$. On the other hand, in the random effects model, if $\beta_a$ is assumed to have $E(\beta_a) = \tau$, $\text{cov}(\beta_a) = \Gamma$, then $\Sigma_i = X_i \Gamma X_i' + \Omega_i$.

The linear growth function as specified by (2.1) allows for unbalanced situations, i.e., individual $Y_i$ can have different design matrix $X_i$. However, we will restrict our attention to the balanced case, i.e., $X_i = X$, for the rest of the paper.

In the balanced case for the general linear growth curve, the model can be specified as

$$\underset{p \times N}{Y} = \underset{p \times m}{X} \underset{m \times r}{\tau} \underset{r \times N}{A} + \underset{p \times N}{\epsilon}, \tag{2.2}$$

where $\tau$ is the unknown matrix of growth function parameters, $X$ and $A$ are known design matrices of ranks $m < p$ and $r < N$, respectively. Furthermore, the columns of $\epsilon$ are each

independent and $p$-variate normal with mean vector 0 and common covariance matrix $\Sigma$. Hence, $Y$ is normally distributed with mean function $X\tau A$ and covariance matrix $\Sigma \otimes I$, where $\otimes$ denotes the Kronecker product. Usually $p$ is the number of time points observed on each individual, $m$ and $r$, which specify the degree of polynomial in time and the number of distinct groups respectively, are assumed known. The design matrices $X$ and $A$ will therefore characterize the degree of the growth function and the distinct grouping out of $N$ independent vectorial observations. For example, if each individual is observed at $t_1, \ldots, t_p$ and there are two different linear growth functions, then $A$ consists of $N_1$ columns of $(1,0)$ and $N_2$ columns of $(0,1)$, $N_1 + N_2 = N$, and

$$
X = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_p \end{bmatrix}.
\tag{2.3}
$$

The model as specified by (2.2) with $\Sigma$ assumed to be an arbitrary unkown positive definite (p.d.) matrix and with $\Sigma = \sigma^2 C$, where $C = (\rho^{|a-b|})$, was proposed by Potthoff and Roy (1964). It was subsequently considered by many authors including Rao (1965, 1966, 1967, 1987), Khatri (1966, 1973), Geisser (1970, 1980, 1981) Lee and Geisser (1972, 1975), Fearn (1975), Zerbe (1979), Laird and Ware (1982), Lee and Tan (1984), Jenrich and Schluchter (1986), Lee (1988, 1991) and Keramidas and Lee (1990, 1995), among others. As indicated earlier, an arbitrary p.d. $\Sigma$ arises when $\text{cov}(Y_i) = \text{cov}(\epsilon_i) = \Sigma$ in the model specified by (2.1). As for the serial covariance structure in which $\Sigma = \sigma^2 C$, $C = (\rho^{|a-b|})$, it arises from the situation when the error terms $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})'$ in the model specified by (2.1) follows the relationship $\epsilon_{ij} = \rho \epsilon_{i,j-1} + \eta_{ij}$ where $\eta_{ij}$ are uncorrelated errors with common variance $\sigma_\eta^2$.

Although the focus of this paper is the prediction of future observations assumed to have been generated from a growth curve model, we will still consider the estimation of parameters as they are needed in some instances. For an arbitrary p.d. $\Sigma$, the MLEs of the parameters are

$$
\hat{\tau} = (X'S^{-1}X)^{-1}X'S^{-1}YA'(AA')^{-1},
\tag{2.4}
$$

where

$$
S = Y(I - A'(AA')^{-1}A)Y'
$$

3

and

$$\hat{\Sigma} = (Y - X\hat{\tau}A)(Y - X\hat{\tau}A)'/N. \tag{2.5}$$

An alternative estimate for $\Sigma$ is the posterior expectation of $\Sigma$, when the prior $p(\tau, \Sigma^{-1}) \propto |\Sigma|^{(p+1)/2}$ is assumed. Consequently, it can be shown that

$$
\begin{aligned}
E(\Sigma|Y) &= (N - p - 1)^{-1}\{(Y - X\hat{\tau}A)(Y - X\hat{\tau}A)' \\
&+ (N - m - r - 1)^{-1}X(X'S^{-1}X)X'[tr\, G^{-1}AA']\},
\end{aligned} \tag{2.6}
$$

where

$$G^{-1} = (AA')^{-1} + T_2'(Z'SZ)^{-1}T_2$$

$$T_2 = Z'YA'(AA')^{-1}, \tag{2.7}$$

and $Z_{p \times p - m}$ is any matrix of rank $p - m$ such that $X'Z = 0$.

Comparison of (2.5) and (2.6) shows that the difference $E(\Sigma|Y) - \hat{\Sigma}$ is always non-negative definite.

The MLE of $\tau$, as given in (2.4), was obtained by Khatri (1966) and was also derived by Rao (1967) as a covariance adjusted estimator. It can be shown that $\hat{\tau}$ is unbiased, i.e., $E(\hat{\tau}) = \tau$, although it is not a linear function of $Y$. Another unbiased estimator of $\tau$, which is a linear function of $Y$, is

$$T_1 = BYA'(AA')^{-1}, \tag{2.8}$$

where

$$B = (X'X)^{-1}X'.$$

Since $E(T_1) = E(\hat{\tau}) = \tau$, the comparison of their covariance matrices would be instructive as to which would be a more desirable estimator for $\tau$. Rao (1967) showed that $T_1$ is preferable if and only if the following condition holds

$$\Sigma = X\Gamma X' + Z\theta Z' + \sigma^2 I, \tag{2.9}$$

where $\Gamma$ and $\theta$ are arbitrary unknown symmetric matrices. Such a situation arises when the following mixed model is considered,

$$Y_i = X\alpha + X\beta + Z\gamma + \epsilon_i \tag{2.10}$$
$$\text{for } i = 1, \ldots, n_a^*; \; a = 1, \ldots, r$$

4

where $\alpha$ is a column of $\tau$, and $\beta, \gamma$ and $\epsilon_i$ are all uncorrelated random vectors such that $E(\beta) = 0$, $\text{cov}(\beta) = \Gamma$, $E(\gamma) = 0$, $\text{cov}(\gamma) = \theta$, $E(\epsilon_i) = 0$ and $\text{cov}(\epsilon_i) = \sigma^2 I$.

Geisser (1970) noted that

$$X(X'X)^{-1}X' + Z(Z'Z)^{-1}Z' = I$$

and hence claimed that, without loss of generality, the special structure for $\Sigma$ as given by (2.9) can be defined as

$$\Sigma = X\Gamma X' + Z\theta Z', \tag{2.11}$$

which was later termed Rao's simple structure (RSS) by Lee and Geisser (1972). They also showed that the likelihood ratio criterion for testing

$$H_0 : \Sigma = X\Gamma X' + Z\theta Z' \quad \text{vs} \quad H_1 : \Sigma \text{ is p.d.}$$

is

$$\lambda = \frac{|(X'S^{-1}X)^{-1}|}{|BSB'|} \tag{2.12}$$

and is distributed under $H_0$ as $U_{m,p-m,N-p+m-r}$, a product of independent beta variates. The two covariance structures considered so far are essentially the extremes for the growth curve model, one completely general and the other rather close to the independent situation. We will next consider a few other important structures.

The uniform covariance structure is defined as

$$\Sigma = \sigma^2[(1-\rho)I + \rho ee'], \tag{2.13}$$

where $e' = (1, \ldots, 1)$, $-1/(p-1) < \rho < 1$. The uniform covariance structure arises from the situation when the error term $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})'$ in the model specified by (2.1) has the property that $\text{cov}(\epsilon_{ij}, \epsilon_{ik}) = \sigma^2\rho$ for $j \neq k$, and $\text{var}(\epsilon_{ij}) = \sigma^2$. This indicates that the error components of $\epsilon_i$ are exchangeable. This covariance structure is very close to RSS. When $X = (e, X_2)$, that is, if there is a constant term in the growth function, then the MLEs of $\tau, \sigma^2$, and $\rho$ are, respectively, given by Lee (1988) as

$$\hat{\tau} = T_1 = BYA'(AA')^{-1},$$

$$\hat{\sigma}^2 = \text{tr } S^*/pN \tag{2.14}$$

$$\hat{\rho} = (e'S^*e - \text{tr } S^*)/(p-1)\text{tr } S^*,$$

where

$$S^* = S + ZDYA'(AA')^{-1}AY'D'Z', \qquad (2.15)$$
$$D = (Z'Z)^{-1}Z'.$$

Thus, as long as the constant term is in the growth function, which is typical in practice, the MLEs of the parameters are expressed in explicit form. Also, the MLE of $\tau$ is $T_1$, exactly as in the RSS case and consequently does not require covariance adjustment. However, according to our experience, this covariance structure may not be very practical, as least as far as the prediction of future observations are concerned. The likelihood ratio criterion for testing this particular covariance structure can be obtained easily but will be omitted here.

The covariance structure, which also caught the attention of Potthoff and Roy (1964), Lee and Geisser (1975) and Lee (1988, 1991), is the serial covariance structure which is defined as

$$\Sigma = \sigma^2 C, \qquad (2.16)$$

where

$$C = (\rho^{|a-b|}).$$

The MLEs of the parameters are given by Lee (1988) as

$$\hat{\tau} = (X'\hat{C}^{-1}X)^{-1}X'\hat{C}^{-1}YA'(AA')^{-1} \qquad (2.17)$$
$$\hat{\sigma}^2 = [tr\,(X'\hat{C}^{-1}X)^{-1}X'\hat{C}^{-1}S\hat{C}^{-1}X + tr\,(Z'\hat{C}Z)^{-1}Z'YY'Z]/pN$$

where

$$\hat{C} = (\hat{\rho}^{|a-b|}),$$

and $\hat{\rho}$ is obtained by maximizing the profile likelihood function

$$L_{\max}(\rho) = (\hat{\sigma}^2(\rho))^{-pN/2}(1 - \rho^2)^{-N(p-1)/s}, \qquad (2.18)$$

and $\hat{\sigma}^2(\rho)$ is the $\hat{\sigma}^2$ given by (2.17) with $\hat{\rho}$ replaced by $\rho$. We thus see that the MLEs of $\tau$ and $\sigma^2$ depend on the MLE of $\rho$, which can be obtained by a one-dimensional search.

No iterations are needed and the computations are relatively easy. A likelihood ratio test for testing

$$H_0 : \Sigma = \sigma^2 C \quad \text{vs} \quad H_1 : \Sigma \text{ is p.d.}$$

is $-2 \ln \lambda$ which is distributed under $H_0$ as $\chi^2_\nu$, $\nu = p(p+1)/2 - 2$ when $N \to \infty$, where

$$\lambda = N^{-pN/2}(\hat{\sigma}^2)^{-pN/2}(1 - \hat{\rho}^2)^{-(p-1)N/2}|(B'Z)|^{-N}|Z'YY'Z|^{N/2}|X'S^{-1}X|^{-N/2}. \quad (2.19)$$

This was obtained by Lee (1991).

The other covariance structures included in this paper are obtained from the different assumptions on $\Omega = \text{cov}(\epsilon_i)$ and $\beta_a$ in the model specified by (2.1). For example, when $\Omega = \sigma^2 C$, and $\beta_a$ is considered as a random vector with mean vector $E(\beta_a) = \tau$ and covariance matrix $\text{cov}(\beta_a) = \Gamma$, then

$$\Sigma = X\Gamma X' + \sigma^2 C \qquad (2.20)$$

The estimation of the parameters for this model is quite complex as indicated by Keramidas and Lee (1995). Two somewhat more parsimonious models involve either the intercept random or the slopes random. The two resulting covariance structures are

$$\Sigma = \gamma ee' + \sigma^2 C \qquad (2.21)$$

and

$$\Sigma = X_2 \Phi X_2 + \sigma^2 C, \qquad (2.22)$$

where

$$X = (e, X_2).$$

The above three covariance structures were considered by Chi and Reinsel (1989).

In (2.20) if we set $\Omega = \sigma^2 I$, then we have the following covariance structure

$$\Sigma = X\Gamma X' + \sigma^2 I, \qquad (2.23)$$

which was considered by Rao (1967) and later promoted by Fearn (1975) as a natural covariance matrix resulting from a two stage hierarchical model in which each individual has a separate growth curve and the parameters associated with each curve are exchangeable, i.e., they are realizations of another normal distribution. It should be pointed out,

however, that although the model started out with a separate curve for each individual, there is still a common growth curve at the end. The only difference is the final covariance structure of $\Sigma$ which is different than $\text{cov}(\epsilon_i)$, i.e., $\Sigma = X\Gamma X' + \sigma^2 I$ rather than $\Sigma = \text{cov}(\epsilon_i)$, as the fixed effects model assumes.

The MLEs of the parameters for the covariance structures (2.20) through (2.23) are not available in closed form. However, using BMDV(BMDP, 1988), one can obtain these estimates. Hence the MLEs of the parameters can be obtained in principle. Nevertheless, convergence could be a problem at times.

## 3  Prediction of Future Observations

In this paper we will consider three types of prediction for the generalized growth curve model as specified by (2.2) when the covariance matrix $\Sigma$ has different structures. Let $V$ be a set of $p \times K$ future observations drawn from the generalized growth curve model, i.e., the set of future observations is such that given the parameters $\tau$ and $\Sigma$,

$$E(V) \;=\; X\tau F, \tag{3.1}$$

where $F$ is a known $r \times K$ matrix, usually formed by some columns of $A$, and the columns of $V$ are independent and multivariate normal with a common covariance matrix $\Sigma$. Geisser (1970, 1980) and Lee (1982) considered prediction of $V$, given $Y$ as the sample, from a Bayesian viewpoint.

The second type of prediction for the generalized growth curve model is to predict $V^{(2)}$ given $V^{(1)}$ and $Y$, if $V$ is partitioned as

$$V \;=\; \begin{bmatrix} V^{(1)} \\ V^{(2)} \end{bmatrix}, \tag{3.2}$$

where $V^{(i)}$ is $p_i \times K (i = 1, 2)$ and $p_1 + p_2 = p$. If $p$ is interpreted as the number of times being observed, then this problem is concerned with the growth curves for future individuals for subsets of size $p_2$, having observed subsets of size $p_1$. This type of prediction is called conditional prediction of $V^{(2)}$ given $V^{(1)}$ and $Y$. When $p_2 < p$ and $K = 1$, it is also called conditional prediction of the unobserved portion of a partially observed vector. This type of prediction has been considered by many authors including Lee and Geisser (1972, 1975), Fearn (1975), Rao (1975, 1987), Geisser (1981), Rei (1984), Lee (1988, 1991), and Keramidas and Lee (1995), among others.

The third type of prediction is somewhat different. It is concerned with predicting future values for the observed individuals. Let $y$, of dimension $q \times n$, be a set of $n(\leq N)$ future observations whose previous $p$-dimensional observations are a subset of $Y$. We are interested in predicting $y$ given $Y$. This is time series type of prediction and thus is very important in practice as growth curve data are often time series in nature. However, $\Sigma$ needs to be structured in order to predict $y$ in a reasonable manner. For example, if $\Sigma$ is assumed arbitrary p.d., then predictions can not be made simply because information is lacking in regard to the relationship between $y$ and $Y$. This type of prediction is called extended prediction of $y$ given $Y$, because the prediction is made beyond the observed time range of the sample $Y$. Extended prediction of $y$ given $Y$ has been considered by Rao (1977, 1987, 1984), Lee (1988, 1991) and Keramidas and Lee (1990, 1995).

Let $x$, of dimension $q \times m$, be a design matrix corresponding to $y$, $Y = (Y_1, \ldots, Y_N)$, $A = (A_1, \ldots, A_N)$, $y = (y_1, \ldots, y_n)$, and assume that for $i \leq n$,

$$E \begin{pmatrix} Y_i \\ y_i \end{pmatrix} = \begin{pmatrix} X \\ x \end{pmatrix} \tau A_i, \tag{3.3}$$

$$\Sigma^* = \operatorname{cov} \begin{pmatrix} Y_i \\ y_i \end{pmatrix} = \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix}. \tag{3.4}$$

We note in passing that $\Sigma_{11}^* = \Sigma$. From (3.4) it is clear that in order to make an inference about $y$, the form of $\Sigma_{21}^*$ has to be known, i.e., the covariance structure for $\Sigma$ has to be extendable to future values.

# 4 $\Sigma$ is Arbitrary p.d.

From a Bayesian viewpoint Geisser (1970) considered the estimation problem for $\tau$ and the predictive inference for $V$. Using a convenient prior, (Geisser and Cornfield, 1963; Geisser, 1965),

$$g(\tau, \Sigma^{-1}) \propto |\Sigma|^{(p+1)/2}, \tag{4.1}$$

Geisser (1970) showed that the predictive density of $V$ given $Y$ is

$$f(V|Y) \propto |Z'[YY' + (V - X\hat{\tau}F)(V - X\hat{\tau}F)']Z|^{-(N+K-m)/2} |G_1|^{m/2}$$
$$\times |I + G_1(V - X\hat{\tau}F)'S^{-1}X(X'S^{-1}X)^{-1}X'S^{-1}(V - X\hat{\tau}F)|^{-(N+K-r)/2}, \tag{4.2}$$

where

$$G_1^{-1} = [I - F'(HH')^{-1}F]^{-1} + (V_1 - Z'\hat{V})'(Z'SZ)^{-1}(V_2 - Z'\hat{V})$$

$$H = (A, F), \quad V_1 = BV, \quad V_2 = Z'\hat{V}, \quad \hat{V} = YA'(AA')^{-1}F.$$

It is clear that the predictive density of $V$ given $Y$ is a product of three general determinantal, or matrix $T$ densities and is extremely complex. Drawings from the three general determinantal distributions can be accomplished as suggested by Zellner et al (1988). However, since the dimension of $V$ is $p \times k$, the Monte Carlo method may be hard to use in practice for the predictive region of $V$. Instead, we will suggest the following. Geisser (1970) showed that $E(V|Y) = X\hat{\tau}F$. He also showed that $X\hat{\tau}F$ is not the mode of the predictive distribution of $V$ given $Y$. However, numerical procedures, via two-dimensional search, for calculating the predictive mode when $K = 1$ were given by Lee and Geisser (1972). Geisser (1970) also showed that for

$$Q = BYA'(AA')^{-1} + BSZ(Z'SZ)^{-1}Z'(V - YA'(AA')^{-1}F),$$

$$U_1 = |I + (X'S^{-1}X)(BV - Q)G_1(BV - Q)'| \tag{4.3}$$

is distributed as $U_{m,K,N-r}$ and is independent of

$$U_2 = |I + V'Z(Z'YY'Z)^{-1}Z'V| \tag{4.4}$$

which is distributed as $U_{p-m,K,N-m}$. Hence, $U_1 + U_2$ is distributed as the sum of two independent $U$ variates. For an excellent approximation to this distribution and its special case, a linear combination of two independent $F$ variates, see Lee and Hu (1995).

For conditional prediction of $V^{(2)}$ given $V^{(1)}$ and $Y$, we note that $f(V^{(2)}|V^{(1)}, Y) \propto f(V|Y)$ and hence the predictive distribution of $V^{(2)}$ given $V^{(1)}$ and $Y$ is at least as complicated as the predicative distribution of $V$ given $Y$. Therefore, some approximations are in order. Lee and Geisser (1972) showed that conditional on $\Sigma^{-1}$ and $Y$, $V$ is distributed as normal with mean

$$\mu_a = X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}YA'(AA')^{-1}F, \tag{4.5}$$

and covariance matrix

$$\Sigma_a = X(X'\Sigma^{-1}X)^{-1}X' \otimes M^{-1} + [XB\Sigma Z(Z'\Sigma Z)^{-1}Z'\Sigma B'X'$$
$$+ D'Z'\Sigma B'X' + XB\Sigma ZD + D'Z'\Sigma ZD] \otimes I_k, \tag{4.6}$$

10

where

$$M = I - F'(AA' + FF')^{-1}F.$$

We thus see that the predictive distribution of $V$ given $Y$ is approximately $N(.; \hat{\mu}_a, \hat{\Sigma}_a)$ where $\hat{\mu}_a$ and $\hat{\Sigma}_a$ are obtained by replacing $\Sigma$ by its estimate, either the MLE or the posterior expectation as given by (2.5) or (2.6), respectively. Hence, a predictive region for $V$ given $Y$ as well as for $V^{(2)}$ given $V^{(1)}$ and $Y$ can be obtained through standard normal theory, with appropriate rearrangement of $V$ and the corresponding covariance matrix $\hat{\Sigma}_a$.

As for extended prediction of $y$ given $Y$, the prediction can not be made because there is no information concerning the dependence structure between $y$ and the observations in the sample. Hence, other covariance structures need to be assumed to facilitate extended prediction.

# 5    $\Sigma$ has RSS

Using the convenient prior,

$$g(\Gamma^{-1}, \theta^{-1}, \tau) \quad \propto \quad |\Gamma|^{(m+1)/2} |\theta|^{(p-m+1)/2}, \tag{5.1}$$

Geisser (1970) showed that the predictive density of $V$ given the sample $Y$ is

$$f(V|Y) \quad \propto \quad |Z'[YY' + (V - XT_1F)(V - XT_1F)']Z|^{-(N+K)/2}$$
$$\times \quad |I + (I - F'(HH')^{-1}F)(V - XT_1F)'B'(BSB')^{-1}B(V - XT_1F)|^{-(N+K-r)/2}. \tag{5.2}$$

It is clear that $XT_1F$ is the mean and the mode of $f(V|Y)$. It can also be shown that a posteriori

$$U_1 = |I + (I - F'(HH')^{-1}F)(V - XT_1F)'B'(BSB')^{-1}B(V - XT_1F)|^{-1} \tag{5.3}$$

and

$$U_2 = |I + (V - XT_1F)'Z'(Z'YY'Z)^{-1}Z'(V - XT_1F)|^{-1} \tag{5.4}$$

are independently distributed as $U_{m,K,N-r}$ and $U_{p-m,K,N}$, respectively. Hence a predictive region for $V$ can be obtained through $U_1 + U_2$ which is distributed as the sum of two

11

independent $U$ variates. For the special case in which $K = F = r = 1$, a predictive region for a future vectorial observation can be obtained from

$$Q = (V - XT_1)'[N(N+1)^{-1}X(X'SX)^{-1}X' + Z(Z'YY'Z)^{-1}Z'](V - XT_1) \quad (5.5)$$

which is distributed as a linear combination of two independent $F$ variates. It is noted that $Q$ can be written free of $Z$ by following identity

$$Z(Z'YY'Z)^{-1}Z' = (YY')^{-1} - (YY')^{-1}X[X'(YY')^{-1}X]^{-1}X'(YY')^{-1}. \quad (5.6)$$

Accurate approximations to the above distributions have been obtained by Lee and Hu (1995).

An alternative representation for the predictive density of $V$ given $Y$, when $K = 1$, was obtained by Lee and Geisser (1972) as

$$f(V|Y) = \int_0^\infty f(V|Y,t)g(t|Y)dt \quad (5.7)$$

where

$$
\begin{aligned}
F(V|Y,t) &= St(.; XT_1F, (1+\gamma t)(2N+2-r-p)^{-1}J^{-1}, 2N+2-r), \\
J &= t\gamma MX(X'SX)^{-1}X' + Z(Z'YY'Z')^{-1}Z', \\
M &= 1 - F'(AA' + FF')^{-1}F, \\
\gamma &= (N+1-p+m)^{-1}(N+1-r-m), \\
G(t|Y) &= F(.; N+1-r-m, N+1-p+m),
\end{aligned}
\quad (5.8)
$$

and $St(.; \mu, \Sigma, v+p)$ is a multivariate Student $t$ distribution with density

$$f(T) = (\pi\nu)^{-p/2}\Gamma^{-1}(\nu/2)\Gamma(N/2)|\Sigma|^{-1/2}|1 + (T-\mu)'(\nu\Sigma)^{-1}(T-\mu)|^{-N/2} \quad (5.9)$$

where $T$ is $p \times 1$ and $\nu = N - p$.

Thus, the predictive density of $V$ given $Y$ is expressed as an average of a multivariate Student $t$ density over an $F$ density. As noted in Lyung and Box (1980), since $F$ is nearly symmetric and well concentrated, a reasonable approximation for the predictive density of $V$ given $Y$ is

$$
\begin{aligned}
F(V|Y) &\doteq F(V|Y,\hat{t}) \\
&= St(.; XT_1F, (1+\gamma\hat{t})(2N+2-r-p)^{-1}J^{-1}(\hat{t}), 2N+2-r), \quad (5.10)
\end{aligned}
$$

where $\hat{t}$ is either the mean or the mode of the $F$ dsitribution and they are $(N+1-p+m)/(N-1-p+m)$ or $(N-r-m-1)(N+1-p+m)/(N+1-r-m)(N+3-p+m)$, respectively, and $J(\hat{t})$ is the value of $J$ evaluated at $t = \hat{t}$. From (5.10) we have

$$F(V^{(2)}|V^{(1)},Y) \doteq St(.;\mu_{r2\cdot1}(\hat{t}),b(\hat{t})(2N+2-r-p_2)^{-1}J_{22}^{-1}(\hat{t}),2N+2-r) \quad (5.11)$$

where

$$\mu_{r2\cdot1} = X^{(2)}T_1F - J_{22}^{-1}J_{21}(V^{(1)} - X^{(1)}T_1F),$$
$$b = 1 + \gamma t + (V^{(1)} - X^{(1)}T_1F)'J_{11\cdot2}(V^{(1)} - X^{(1)}T_1F), \quad (5.12)$$
$$X = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}, \ X^{(i)} \text{ is } p_i \times m, \ p_1 + p_2 = p,$$
$$J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}, \ J_{ij} \text{ is of dimension } p_i \times p_j, \ J_{11\cdot2} = J_{11} - J_{12}J_{22}^{-1}J_{21}.$$

Thus, an approximate point estimate of $V^{(2)}$ given $V^{(1)}$ and $Y$ is $\mu_{r2\cdot1}(\hat{t})$ and an approximate predictive region for $V^{(2)}$ can be obtained through

$$(V^{(2)} - \mu_{r2\cdot1}(\hat{t}))'[b^{-1}(\hat{t})J_{22}(\hat{t})](V^{(2)} - \mu_{r2\cdot1}(\hat{t})) \quad (5.13)$$

which is distributed as $(2N+2-r-p_2)^{-1}p_2 F(p_2, 2N+2-r-p_2)$. A better approximation for the predictive density of $V$ given $Y$ is

$$f(V|Y) \doteq \frac{1}{L}\sum_{i=1}^{L} f(V|Y,t^{(i)}) \quad (5.14)$$

where $f(V|Y,t^{(i)})$ is a multivariate Student $t$ density as given in (5.10) and $t^{(i)}$ is the $i$th draw from $F(N+1-r-m, N+1-p+m)$. It is noted that (5.14) will clearly be better than the approximation given by (5.10) as $L = 1$ with $t^{(i)} = \hat{t}$ will be a special case, see e.g., Liu (1995). However, an average of $L$ multivariate Student $t$ densities is no longer a multivariate Student $t$ density and hence a predictive region based on the approximation (5.14) will be harder to obtain in practice.

For the conditional distribution of $V^{(2)}$ given $V^{(1)}$ and $Y$, we also have

$$f(V^{(2)}|V^{(1)},Y) = \int_0^\infty f(V^{(2)}|V^{(1)},r,t)g(t|V^{(1)},Y)dt, \quad (5.15)$$

where

$$F(V^{(2)}|V^{(1)},Y,t) = St(.;\mu_{s2\cdot1},b(2N+2-r-p_2)^{-1}J_{22}^{-1},2N+2-r),$$
$$\mu_{s2\cdot1} = X^{(2)}T_1F - J_{22}^{-1}J_{21}(V^{(1)} - X^{(1)}T_1F), \quad (5.16)$$

13

$$b = 1 + \gamma t + (V^{(1)} - X^{(1)}T_1 F)' J_{11\cdot 2} (V^{(1)} - X^{(1)}T_1 F),$$

$$g(t|V^{(1)}, Y) = [f(V^{(1)}|Y)]^{-1} \pi^{p_2/2} t^{(N+1-r)/2-1}$$
$$\times \ \Gamma[\frac{1}{2}(2N+1-r-p_2)] b^{-(2N+2-r-p_2)/2} C_0 |J_{22}|^{-\frac{1}{2}}, \qquad (5.17)$$

$$C_0 = \pi^{-p/2} M^{m/2} |BSB'|^{-\frac{1}{2}} |DYY'D'|^{-\frac{1}{2}}$$
$$\times \ \Gamma^{-1}[\frac{1}{2}(N+1-r-m)] \Gamma^{-1}[\frac{1}{2}(N+1-p+m)]$$
$$\times \ \text{mod} \left| \begin{pmatrix} B \\ D \end{pmatrix} \right| \qquad (5.18)$$

$$f(V^{(1)}|Y) = \int_0^\infty f(V^{(1)}|Y,t) g(t|Y) dt, \qquad (5.19)$$

$$F(V^{(1)}|Y,t) = St(.; X^{(1)}T_1 F, (1+\gamma t)(2N+2-r-p)^{-1} J_{11\cdot 2}^{-1},$$
$$2N+2-r-p_2), \qquad (5.20)$$

$$G(t|Y) = F(.; N+1-r-m, N+1-p+m).$$

As in (5.10), a reasonable approximation for the conditional predictive distribution of $V^{(2)}$ given $V^{(1)}$ and $Y$ is

$$F(V^{(2)}|V^{(1)}, Y) \doteq F(V^{(2)}|V^{(1)}, Y, t_0)$$
$$= St(.; \mu_{s2\cdot 1}(t_0), b(t_0)(2N+2-r-p_2)^{-1} J_{22}^{-1}(t_0),$$
$$2N+2-r) \qquad (5.21)$$

where $t_0$ maximizes $g(t|V^{(1)}, Y)$. Similar to (5.14), a better approximation for the conditional predictive density of $V^{(2)}$ given $V^{(1)}$ and $Y$ is

$$f(V^{(2)}|V^{(1)}, Y) \doteq \frac{1}{L} \sum_{i=1}^{L} f(V^{(2)}|V^{(1)}, Y, t^{(i)}) \qquad (5.22)$$

where $t^{(i)}$ is the $i$th draw from $g(t|V^{(1)}, Y)$. A comparison among (5.11), (5.15), (5.21) and (5.22), via a real data set, will be given in Section 12.

It is also noted that numerical prodecures are given for $K = 1$ by Lee and Geisser (1972) for obtaining the predictive mode of $V^{(2)}$ given $V^{(1)}$ and $Y$ and an exact solution for the particularly interesting special case $p_2 = 1$. As in the arbitrary p.d. case, extended prediction for $y$ given $Y$ is not available for the RSS case. This is due to the fact that the dependence between $y$ and its previous observations is not defined.

# 6  $\Sigma$ has Uniform Covariance Structure

For the rest of this paper we will restrict our attention to the special case in which $K = 1$, i.e., there is one future vectorial observation to be predicted. When $\Sigma$ has the uniform covariance structure, it has the form as given in (2.13). The predictive distribution of $V$ given $Y$ can be approximated as

$$F(V|Y) \doteq F(V|Y,\hat{\tau},\hat{\rho},\hat{\sigma}^2). \tag{6.1}$$

Thus, the predictive distribution of $V$ given $Y$ is approximately

$$F(V|Y) \doteq N(.;XT_1F,\hat{\sigma}^2[(1-\hat{\rho})I + \hat{\rho}ee']) $$

and the conditional distribution of $V^{(2)}$ given $V^{(1)}$ and $Y$ is

$$F(V^{(2)}|V^{(1)},Y) \doteq N(.;\mu_{u2\cdot1},\Sigma_{u22\cdot1}), \tag{6.2}$$

where

$$\mu_{u2\cdot1} = X^{(2)}T_1F + \frac{\hat{\rho}}{1+(p_1-1)\hat{\rho}}e_{p_2}e'_{p_1}(V^{(1)} - X^{(1)}T_1F), \tag{6.3}$$

$$\Sigma_{u22\cdot1} = \hat{\sigma}^2[(1-\hat{\rho})I_{p_2} + \frac{\hat{\rho}(1-\hat{\rho})}{1+(p_1-1)\hat{\rho}}e_{p_2}e'_{p_2}].$$

Hence, the approximate means, $XT_1F$ and $\mu_{u2\cdot1}$, can be used as the point predictors for $V$ given $Y$, and $V^{(2)}$ given $V^{(1)}$ and $Y$, respectively, and the corresponding predictive regions can be obtained through standard normal theory.

Instead of the approximate mean as the point predictor for $V^{(2)}$ given $V^{(1)}$ and $Y$, an ad hoc predictor can be obtained by the weighted average of two independent predictors, one based on $V^{(1)}$ alone and the other based on $Y$, with the weights being the estimates of the relative precisions. This type of predictor was first proposed by Lee and Geisser (1975). The ad hoc predictor, when the uniform covariance structure is assumed, is

$$P_{ua} = (S_{m1}^{-1} + S_{m2}^{-1})^{-1}(S_{m1}^{-1}m_1 + S_{m1}^{-1}m_2), \tag{6.4}$$

where

$$m_1 = X^{(2)}T_1F,$$

$$m_2 = X^{(2)}\tau_vF + \Sigma_{u21}\Sigma_{u11}^{-1}(V^{(1)} - X^{(1)}\tau_vF),$$

$$\tau_v = (X^{(1)'}\Sigma_{u11}^{-1}X^{(1)})^{-1}X^{(1)'}\Sigma_{u11}^{-1}V^{(1)}F'(FF')^{-} \tag{6.5}$$

$$\Sigma_u = \hat{\sigma}^2[(1-\hat{\rho})I + \hat{\rho}ee'] = \begin{pmatrix} \Sigma_{u11} & \Sigma_{u12} \\ \Sigma_{u21} & \Sigma_{u22} \end{pmatrix},$$

and

$$S_{m1} = \Sigma_{u22} + X^{(2)}B\Sigma_u B'X^{(2)'}F'(AA')^{-1}F,$$

$$S_{m2} = \Sigma_{u22} + (b_1 B_2 + \Sigma_{u21})'\Sigma_{u11}^{-1}(b_1 B_2 + \Sigma_{u21})$$

$$\quad - (b_1 B_2 + \Sigma_{u21})\Sigma_{u11}^{-1}\Sigma_{u12} + \Sigma_{u21}\Sigma_{u11}^{-1}(b_1 B_2 + \Sigma_{u21})',$$

$$b_1 = F'(FF')^- F,$$

$$B_2 = (X^{(2)} - \Sigma_{u21}\Sigma_{u11}^{-1}X^{(1)})(X^{(1)'}\Sigma_{u11}^{-1}X^{(1)})^{-1}X^{(1)'},$$

where $\Sigma_{uij}$ is $p_i \times p_j$, $p_1 + p_2 = p$, $(FF')^-$ is the generalized inverse of $FF'$ and can be set as $FF'$ in the special case $F = (1,0)'$ or $F = (0,1)'$. Note that $F'(AA')^{-1}F$ is a scalar and $S_{mi}$ is the covariance matrix of the forecast error when $m_i$ is the predictor for $V^{(2)}$. An alternative formula for $S_{mi}$ was given by Lee and Geisser (1975). A justification for the ad hoc predictor is the fact that it is the optimum combination of two independent forecasts.

Next, consider extended prediction of $y$, given $Y$. This is a time series type of prediction. To make this type of prediction the covariance structure generally has to be extendable to future values of the individuals observed.

Let $x$, $q \times m$, be a design matrix corresponding to $y$, $Y = (Y_1, \ldots, Y_N)$, $A = (A_1, \ldots, A_N)$, $y = (y_1, \ldots, y_n)$, and assume that for $i \leq n$,

$$E\begin{pmatrix} Y_i \\ y_i \end{pmatrix} = \begin{pmatrix} X \\ x \end{pmatrix}\tau A_i, \tag{6.6}$$

and

$$\Sigma^* = \mathrm{cov}\begin{pmatrix} Y_i \\ y_i \end{pmatrix} = \begin{pmatrix} \Sigma_{11}^* & \Sigma_{12}^* \\ \Sigma_{21}^* & \Sigma_{22}^* \end{pmatrix} = \sigma^2\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \sigma^2\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}. \tag{6.7}$$

Similar to (6.1), the predictive distribution of $y_i$ given $Y$ can be approximated as

$$F(y_i|Y) = F(y_i|Y, \hat{\tau}, \hat{\Gamma}, \hat{\sigma^2}). \tag{6.8}$$

Thus, the predictive distribution of $y_i$ given $Y$ is approximately

$$F(y_i|Y) \doteq N(.; \mu_{u2\cdot1}^*, \Sigma_{u22\cdot1}^*), \tag{6.9}$$

where

$$\mu_{u2\cdot1}^* = xT_1 A_i + \frac{\hat{\rho}}{1 + (p-1)\hat{\rho}}e_q e_p'(Y_i - XT_1 A_i),$$

$$\Sigma_{u22\cdot1}^* = \hat{\sigma}^2[(1-\hat{\rho})I_q + \frac{\hat{\rho}(1-\hat{\rho})}{1+(p-1)\hat{\rho}}e_q e_q'], \tag{6.10}$$

16

and $T_1$, $\hat{\rho}$, and $\sigma^2$ are given in (2.14).

# 7 Serial Covariance Structure for $\Sigma$

When $\Sigma$ has the serial covariance structure it takes the form $\Sigma = \sigma^2 C$ as given in (2.6). The predictive distribution of $V$ given $Y$ can be approximated as

$$F(V|Y) \doteq F(V|Y, \hat{\tau}, \hat{\rho}, \hat{\sigma}^2) \tag{7.1}$$

where $\hat{\tau}$, $\hat{\rho}$ and $\hat{\sigma}^2$ are given in (2.17). Thus, the predictive distribution of $V$ given $Y$ is approximately

$$F(V|Y) \doteq N(.; X\hat{\tau}F, \sigma^2 C(\hat{\rho})) \tag{7.2}$$

and the conditional distribution of $V^{(2)}$ given $V^{(1)}$ and $Y$ is

$$F(V^{(2)}|V^{(1)}, Y) \doteq N(.; \mu_{s2\cdot1}, \Sigma_{s22\cdot1}) \tag{7.3}$$

where

$$
\begin{aligned}
\mu_{s2\cdot1} &= X^{(2)}\hat{\tau}F + (0, \hat{\eta})(V^{(1)} - X^{(1)}\hat{\tau}F), \\
\hat{\eta} &= (\hat{\rho}, \ldots, \hat{\rho}^{p_2})' \\
\Sigma_s &= \sigma^2 C(\hat{\rho}) = \begin{pmatrix} \Sigma_{s11} & \Sigma_{s12} \\ \Sigma_{s21} & \Sigma_{s22} \end{pmatrix}, \\
\Sigma_{s22\cdot1} &= \Sigma_{s22} - \Sigma_{s21}\Sigma_{s11}^{-1}\Sigma_{s12}.
\end{aligned}
\tag{7.4}
$$

Hence, the approximate means, $X\hat{\tau}F$ and $\mu_{s2\cdot1}$, can be used as the point predictors for $V$ given $Y$, and $V^{(2)}$ given $V^{(1)}$ and $Y$, respectively, and the corresponding predictive regions can be easily obtained through standard normal theory.

The ad hoc predictor, when the serial covariance structure holds, is

$$P_{sa} = (S_{m_1^*}^{-1} + S_{m_2^*}^{-1})^{-1}(S_{m_1^*}^{-1}m_1^* + S_{m_2^*}^{-1}m_2^*), \tag{7.5}$$

where $m_1^*$, $m_2^*$ and $S_{m_2^*}$ are defined as in $m_1$, $m_2$ and $S_{m_2}$, respectively, except that $T_1$ is replaced by $\hat{\tau}$ and $\Sigma_{uij}$ is replaced by $\Sigma_{sij}$, and

$$S_{m_1^*} = \Sigma_{s22} + X^{(2)}(X'\Sigma_s^{-1}X)^{-1}X^{(2)'}F'(AA')^{-1}F. \tag{7.6}$$

If the covariance structure is appropriate for the growth curve data at hand, the ad hoc predictor should be a very strong competitor as a point predictor for the unobserved

protion, $V^{(2)}$, of the partially observed vector $V$. This is due to the fact that it utilizes all the available data in a very reasonable fashion. This observation has been borne out in studies such as Lee (1988) for several real data sets. One disadvantage of the ad hoc predictor is that the predictive region is not available in a natural manner. However, since the covariance matrix of the forecast error, when $P_{sa}$ is the predictor for $V^{(2)}$, is approximately

$$\text{cov}(P_{sa}) \ \doteq \ (S_{m1}^{-1} + S_{m2}^{-1})^{-1}, \tag{7.7}$$

we have approximately,

$$F(P_{sa}|V^{(2)}, Y) \ \doteq \ N(.; V^{(2)}, (S_{m1}^{-1} + S_{m2}^{-1})^{-1}). \tag{7.8}$$

Hence, an approximate predictive region for $V^{(2)}$ given $V^{(1)}$ and $Y$ can be obtained through standard normal theory.

We next consider extended prediction of $y_i$ given $Y$. As in the previous section, assume for $i \leq n$,

$$E \begin{pmatrix} Y_i \\ y_i \end{pmatrix} =\cdot \begin{pmatrix} X \\ x \end{pmatrix} \tau A_i \tag{7.9}$$

and

$$\Sigma^* \ = \ \text{cov} \begin{pmatrix} Y_i \\ y_i \end{pmatrix} = \sigma^2 \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \tag{7.10}$$

where

$$\begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \ = \ (\rho^{|a-b|}), a, b, = 1, \dots, p+q, \tag{7.11}$$

$$\tag{7.12}$$

$C_{11}$ is $p \times p$, $C_{12}$ is $p \times q$, $C_{22}$ is $q \times q$, and $C_{21} = C_{12}'$. Similar to the predictive distribution of $V$ given $Y$, the predictive distribution of $y_i$ given $Y$ can be approximated as

$$F(y_i|Y) \ \doteq \ F(y|Y, \hat{\tau}, \hat{\rho}, \hat{\sigma}^2). \tag{7.13}$$

Thus, the predictive distribution of $y_i$ given $Y$ is approximately

$$F(y_i|Y) \ \doteq \ N(.; \mu_{s2\cdot1}^*, \Sigma_{s22\cdot1}^*), \tag{7.14}$$

where

$$\mu^*_{s2\cdot1} = X\hat{\tau}A_i + (0,\hat{\eta}^*)(Y_i - X\hat{\tau}A_i), \quad (7.15)$$

$$\hat{\eta}^* = (\hat{\rho},\dots,\hat{\rho}^q)',$$

$$\Sigma^*_{s22\cdot1} = \Sigma^*_{s22} - \Sigma^*_{s21}\Sigma^{-1*}_{s11}\Sigma^*_{s12}, \quad (7.16)$$

$$\Sigma^*_s = \sigma^2 C(\hat{\rho}) = \begin{pmatrix} \hat{C}_{11} & \hat{C}_{12} \\ \hat{C}_{21} & \hat{C}_{22} \end{pmatrix}.$$

The approximate mean $\mu^*_{s2\cdot1}$ can be used as the point predictor for $y_i$ and the predictive region for $y_i$ can be obtained through standard normal theory.

# 8    Other Covariance Structures

In this section we consider the situation in which the covariance structures are obtained from the consideration of random (and mixed) effects for the regression parameters. They include the covariance structures given in (2.20) through (2.23). In addition to the random effects for the regression parameters, the growth curve models associated with these covariance structures rely heavily on numerical solutions for the estimation of parameters. In other words, the MLEs are difficult to obtain numerically and the convergence to the MLEs may not always be assured. Newer procedures such as Markov chain Monte Carlo maximum likelihood methods may alleviate the situation, Geyer and Thompson (1992).

Once the numerical estimates, $\hat{\tau}$ and $\hat{\Sigma}$, of the parameters, $\tau$ and $\Sigma$, are obtained, the predictive distributions of $V$ given $Y$, and $V^{(2)}$ given $V^{(1)}$ and $Y$ can be approximated as

$$F(V|Y) \doteq N(.; X\hat{\tau}F, \hat{\Sigma}), \quad (8.1)$$

and

$$F(V^{(2)}|V^{(1)}, Y) \doteq N(.; \hat{\mu}_{2\cdot1}, \hat{\Sigma}_{22\cdot1}), \quad (8.2)$$

where

$$\hat{\mu}_{2\cdot1} = X^{(2)}\hat{\tau}F + \hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}(V^{(1)} - X^{(1)}\hat{\tau}F), \quad (8.3)$$

$$\hat{\Sigma}_{22\cdot1} = \hat{\Sigma}_{22} - \hat{\Sigma}_{21}\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}. \quad (8.4)$$

As for extended prediction of $y$ given $Y$, the predictive distribution of $y$ given $Y$ can be approximated by

$$f(y|Y) \doteq F(y|Y, \hat{\tau}, \hat{\Sigma}). \quad (8.5)$$

Thus, the predictive distribution of $y_i$ given $Y$ is approximately

$$f(y_i|Y) \doteq N(.; \hat{\mu}^*_{2\cdot1}, \hat{\Sigma}^*_{22\cdot1}).$$

(8.6)

where

$$\hat{\mu}^*_{2\cdot1} = X\hat{\tau}A_i + \hat{\Sigma}^*_{21}\hat{\Sigma}^{-1*}_{11}(V^{(1)} - X^{(1)}\hat{\tau}F),$$

(8.7)

$$\hat{\Sigma}^*_{22\cdot1} = \hat{\Sigma}^*_{22} - \hat{\Sigma}^*_{21}\hat{\Sigma}^{-1*}_{11}\hat{\Sigma}^*_{12},$$

(8.8)

and $\Sigma^*_{ij}$ are defined in (3.4).

# 9  A Predictive Sample Reuse Approach for Conditional Prediction

This is a data analytic aparametric method which simulates the predictive process within the sample, given a complete lack of any distributional assumption. It is termed predictive sample reuse (PSR) by Geisser (1974, 1975) because each vector in the sample of size $N$ will be utilized $(N-1)$ times in the prediction process.

In this secion we are concerned with predicting one future vectorial observation, i.e., $K = 1$. Also, for ease of presentation let $V = Y_{N+1} = \begin{pmatrix} Y^{(1)}_{N+1} \\ Y^{(2)}_{N+1} \end{pmatrix}$ where $Y^{(i)}_{N+1}$ is $p_i \times 1$ and $p_1 + p_2 = p$. Suppose from the first $N$ vectors $Y_1, \ldots, Y_N$, we generate a data-based predictor of $Y^{(2)}_{N+1}$, denoted as $\hat{Y}^{(2)}_{(N)}$. Further, suppose another predictor of $Y^{(2)}_{N+1}$, denoted as $\hat{Y}^{(2)}_{N+1}$, depends only on the observed $Y^{(1)}_{N+1}$. The two independent predictors are then combined to produce a new predictor

$$\dot{Y}^{(2)}_{N+1} = f(\hat{Y}^{(2)}_{(N)}, \hat{Y}^{(2)}_{N+1}; \theta)$$

(9.1)

for $\theta \in \Theta$, $\Theta$ being the admissible domain of $\theta$ and $f$ an assumed predictive function. A particularly interesting case is

$$\dot{Y}^{(2)}_{N+1} = \theta\hat{Y}^{(2)}_{(N)} + (I - \theta)\hat{Y}^{(2)}_{N+1}$$

(9.2)

where $\theta$ is $p_2 \times p_2$ such that both $\theta$ and $I - \Theta$ are non-negative definite. Let

$$\dot{Y}^{(2)}_{\alpha} = \theta\hat{Y}^{(2)}_{(N-1,\alpha)} + (I - \theta)\hat{Y}^{(2)}_{\alpha}$$

(9.3)

where $\alpha = 1, \ldots, N$ and $\hat{Y}^{(2)}_{(N-1,\alpha)}$ is the predictor of $Y^{(2)}_{\alpha}$ based on $Y_1, \ldots, Y_{\alpha-1}, Y_{\alpha+1}, \ldots, Y_N$ and of the same functional form as $\hat{Y}^{(2)}_{(N)}$ and $\hat{Y}^{(2)}_{\alpha}$ is the predictor of $Y^{(2)}_{\alpha}$ and of the same

functional form as $\hat{Y}_{N+1}^{(2)}$. Further, define the discrepancy measure

$$D(\theta) \;=\; \sum_{\alpha=1}^{N} d(\dot{Y}_{\alpha}^{(2)}, Y_{\alpha}^{(2)}) \tag{9.4}$$

which is then minimized w.r.t. $\theta \in \Theta$. If $\hat{\theta}$ is the unique solution then the final predictor is

$$\tilde{Y}_{N+1}^{(2)} \;=\; \hat{\theta}\hat{Y}_{(N)}^{(2)} + (I - \hat{\theta})\hat{Y}_{N+1}^{(2)}. \tag{9.5}$$

A natural discrepancy measure, as utilized by Geisser (1975, 1980), is

$$D(\theta) \;=\; \sum_{\alpha=1}^{N} (\dot{Y}_{\alpha}^{(2)} - Y_{\alpha}^{(2)})'(\dot{Y}_{\alpha}^{(2)} - Y_{\alpha}^{(2)}). \tag{9.6}$$

With $D(\theta)$ given in (9.6) and $p_1 > m = 2$ and $p_2 = 1$, a solution for combining predictors based on simple least square predictors, appears in Geisser (1975) where

$$\hat{Y}_{(N)}^{(2)} \;=\; X^{(2)}T_1 \text{ and } \hat{Y}_{N+1}^{(2)} = X^{(2)}(X^{(1)'}X^{(1)})^{-1}X^{(1)'}Y_{N+1}^{(1)}.$$

A general solution may be obtained for other forms of $\hat{Y}_{(N)}^{(2)}$ and $\hat{Y}_{N+1}^{(2)}$, when $m < p_1$ and $p_2$ is arbitrary as

$$\hat{\theta} \;=\; [\sum_{\alpha=1}^{N}(\hat{Y}_{(N-1,\alpha)}^{(2)} - \hat{Y}_{\alpha}^{(2)})(\hat{Y}_{(N-1,\alpha)}^{(2)} - \hat{Y}_{\alpha}^{(2)})']$$
$$\times \; [\sum_{\alpha=1}^{N}(\hat{Y}_{(N-1,\alpha)}^{(2)} - \hat{Y}_{\alpha}^{(2)})(\hat{Y}_{(N-1,\alpha)}^{(2)} - \hat{Y}_{\alpha}^{(2)})']^{-1}, \tag{9.7}$$

provided it exists and satisfies the constraints.

# 10  Transformations on Y

Motivated by real data considerations for forecasting technological subsitutions, Keramidas and Lee (1995) considered the following power transformations:

$$\underset{p \times N}{Y}{}^{(\lambda)} \;=\; \underset{p \times m}{X} \; \underset{m \times r}{\tau} \; \underset{r \times N}{A} + \underset{p \times N}{\epsilon}, \tag{10.1}$$

where

$$Y^{(\lambda)} \;=\; (Y_1^{(\lambda_1)}, \dots, Y_N^{(\lambda_N)})$$

and

$$Y_i^{(\lambda_i)} \;=\; (Y_{1i}^{(\lambda_i)}, \dots, Y_{pi}^{(\lambda_i)})',$$

21

with

$$
\begin{aligned}
Y_{ji}^{(\lambda_i)} &= \frac{(Y_{ji}+\gamma)^{\lambda_i}-1}{\lambda_i} \quad \text{when } \lambda_i \neq 0, \\
&= \ln(Y_{ji}+\gamma) \quad \text{when } \lambda_i = 0,
\end{aligned}
\tag{10.2}
$$

and $\gamma$ is assumed to be a known constant such that $Y_{ij}+\gamma > 0$ for all $j$ and $i$, $\lambda_i$ is an unknown parameter. Furthermore, the columns of $\epsilon$ are each independent and $p$-variate normal with mean vector 0 and common covariance $\Sigma = \sigma^2 C$, as defined in (2.16).

The MLEs of the parameters $\tau$, $\sigma^2$, $\rho$ and $\lambda$ are

$$
\hat{\tau} = (X'\hat{C}^{-1}X)^{-1}X'\hat{C}^{-1}Y^{(\lambda)}A'(AA')^{-1}
\tag{10.3}
$$

$$
\hat{\sigma}^2 = \frac{1}{pN}[tr\,(X'\hat{C}^{-1}X)^{-1}X'\hat{C}^{-1}S\hat{C}^{-1}X + tr\,(Z'\hat{C}Z)^{-1}Z'(Y^{(\lambda)})(Y^{(\lambda)})'Z],
\tag{10.4}
$$

where

$$
S = Y^{(\lambda)}(I - A'(AA')^{-1}A)Y^{(\lambda)'},
\tag{10.5}
$$

$$
\hat{C} = (\hat{\rho}^{|a-b|}),
\tag{10.6}
$$

and $\hat{\rho}$ and $\hat{\lambda}$ are obtained by maximizing the profile likelihood function

$$
L_{\max}(\rho,\lambda) = (\hat{\sigma}^2(\rho,\lambda))^{-pN/2}(1-\rho^2)^{-N(p-1)/2}J,
\tag{10.7}
$$

where $J$, the Jacobin of the transformation from $Y^{(\lambda)}$ to $Y$, is defined as

$$
J = \prod_{j=1}^{N}\prod_{i=1}^{p}Y_{ij}^{\lambda_i-1}
$$

and $\sigma^2(\rho,\lambda)$ is the $\hat{\sigma}^2$ given by (10.4) with $\hat{\rho}$ and $\hat{\lambda}$ replaced by $\rho$ and $\lambda$, respectively.

We thus see that the MLEs of $\tau$ and $\sigma^2$ are expressed in explicit forms and require no iteration. The MLEs of $\rho$ and $\lambda$ can be obtained by a numerical search. Once $\hat{\rho}$ and $\hat{\lambda}$ are obtained, the joint MLEs of $\tau$, $\sigma^2$, $\rho$ and $\lambda$ are established. Hence the most important step is to carry out the maximization of the profile likelihood function $L_{\max}(\rho,\lambda)$ as given in (10.7), which is an extension of (2.18). It is noted that in practice the maximization is easier than it looks, because the number of power transformations is less than $N$ in practice. It was proposed by Keramidas and Lee (1990) that the number of power transformations be identical to the number of groups, that is, that there be only $r$ different $\lambda$'s, even though there are $N$ independent vectors. This means that the same power transformation will be applied to each of the observations in the same group. Also, since

$p$ is usually quite small, the MLEs of $\lambda$ is very hard to obtain if $N$ different $\lambda$'s are allowed. Thus, in (10.7) we are dealing with a $(r+1)$-dimensional search only.

For extended prediction of $y$ given $Y$, we assume that for $i \leq n$,

$$E\left[\begin{array}{c} Y_i^{(\lambda)} \\ y_i^{(\lambda)} \end{array}\right] = \left(\begin{array}{c} X \\ x \end{array}\right)\tau A_i, \tag{10.8}$$

$$\operatorname{cov}\left[\begin{array}{c} Y_i^{(\lambda)} \\ y_i^{(\lambda)} \end{array}\right] = \sigma^2\left(\begin{array}{cc} C_{11} & C_{12} \\ C_{21} & C_{22} \end{array}\right) \tag{10.9}$$

where $C_{ij}$ is specified in (7.11). From the conditional expectation of $y_i$ given $Y$ we obtained the following predictor for $y_i$,

$$\begin{aligned} \hat{y}_i &= \{e + \hat{\lambda}_i[x\hat{\tau}A_i + (0,\hat{\eta}^*)(Y_i^{(\hat{\lambda}_i)} - x\hat{\tau}A_i)]\}^{1/\hat{\lambda}_i} \quad \text{when } \lambda_i \neq 0 \\ &= \exp[x\hat{\tau}A_i + (0,\hat{\eta}^*)(Y_i^{(\hat{\lambda}_i)} - x\hat{\tau}A_i)] \qquad \text{when } \lambda_i = 0, \end{aligned} \tag{10.10}$$

where $e' = (1,\ldots,1)$, $\hat{\eta}^* = (\hat{\rho},\ldots,\hat{\rho}^q)'$, and $0$ is a $q \times (p-1)$ matrix with all $0$ elements. In (10.10) we use the convention that $b^a = (b_1^a,\ldots,b_p^a)'$.

The model described in this section has been successfully applied to forecasting penetrations of telephone switching systems by Keramidas and Lee (1988), (1990).

# 11   Model Selection and Classification

The prediction problems considered so far can be useful in model selection for the growth curve data. Let $M_1,\ldots,M_g$ be the $g$ possible growth curve models that could have generated the growth curve data at hand. Then the selected model, say $M_{\alpha^*}$, is the one that yields the best predictive accuracy when the sample reuse procedure is used. This approach was utilized by Lee and Geisser (1975), Fearn (1975), Geisser (1981), Rao (1987), and Lee (1988, 1991), among others for the comparison of models based on the performance of the conditional prediction of $V^{(2)}$ given $V^{(1)}$ and $Y$. For more detailed discussion of model selection based on PSR, see Geisser and Eddy (1979).

The discrepancy measure by the PSR method is

$$D_\alpha = \frac{1}{p_2 N}\sum_{j=1}^N (Y_j^{(2)} - \hat{Y}_{(j)}^{(2)}(\alpha))'(Y_j^{(2)} - \hat{Y}_{(j)}^{(2)}(\alpha)), \tag{11.1}$$

where

$$\hat{Y}_{(j)}^{(2)}(\alpha) = X^{(2)}\hat{\tau}_{(j)}(\alpha)A_j + \hat{\Sigma}_{21(j)}(\alpha)\hat{\Sigma}_{11(j)}^{-1}(\alpha)(Y_{(j)}^{(1)} - X^{(1)}\hat{\tau}_{(j)}(\alpha)A_j), \tag{11.2}$$

$$A = (A_1,\ldots,A_N), \ X = \left(\begin{array}{c} X^{(1)} \\ X^{(2)} \end{array}\right), \ Y_i = \left(\begin{array}{c} Y_i^{(1)} \\ Y_i^{(2)} \end{array}\right), \ \Sigma = \left[\begin{array}{cc} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{array}\right],$$

$X^{(a)}$ is $p_a \times m$, $Y_i^{(a)}$ is $p_a \times 1$, $\Sigma_{ab}$ is $p_a \times p_b$, and $\hat{\tau}_{(j)}(\alpha)$ and $\hat{\Sigma}_{ab}(j)(\alpha)$ are the estimates of $\tau$ and $\Sigma_{ab}$ under the model $M_\alpha$ with sample $Y_{(j)} = (Y_1, \ldots, Y_{j-1}, Y_{j+1}, \ldots, Y_N)$. The model $M_{\alpha^*}$ corresponding to the minimum discrepancy is chosen as the most appropriate for the data. It is noted that in the prediction process, $Y_j^{(2)}$ is viewed as $V^{(2)}$ and $Y_j^{(1)}$ as $V^{(1)}$ and $Y_{(j)}$ as $Y$ in the conditional prediction of $V^{(2)}$ given $V^{(1)}$ and $Y$.

This procedure can be applied to any number of competing models, nested or not. One possible drawback is that the choice of $p_1$ and $p_2$ may not be unique. A different choice of $p_1$ and $p_2$ could produce a different selection result. However, for most growth curve problems, the responses are time series in nature, and hence the most practical choice of $p_2$ is 1, i.e., $V^{(2)}$ represents the last component of the vector $V$ (for $K = 1$). This choice of $p_2$ is particularly appealing if the ultimate goal in the modeling effort is the prediction of future values for each of the $N$ vectors.

Instead of conditional prediction of $V^{(2)}$ given $V^{(1)}$ and $Y$, prediction can be made on the entire vector $V$. For this purpose, the discrepancy measure by the PSR procedure is

$$D_\alpha = \frac{1}{pN} \sum_{j=1}^{N} (Y_j - X\hat{\tau}_{(j)}(\alpha)A_j)'(Y_j - X\hat{\tau}_{(j)}(\alpha)A_j). \tag{11.3}$$

Again, the selected model $M_{\alpha^*}$ corresponds to the minimum discrepancy measure.

An advantage of this discrepancy measure is that the prediction is made on the whole vector. Hence there is no indeterminacy in the selection of $p_2$ as encountered in (11.1). However, the prediction of $V$ given $Y$ may not be as sensitive to the appropriateness of the covariance structure as conditional prediction of $V^{(2)}$ given $V^{(1)}$ and $Y$, or extended prediction of $y$ given $Y$, which is the subject of the next discussion.

The discrepancy measure for extended prediction of $y$, when $q = 1$ and $n = N$, is

$$D_\alpha = (Y^{(2)} - \hat{Y^{(2)}})(Y^{(2)} - \hat{Y^{(2)}})'/N \tag{11.4}$$

where $\hat{Y^{(2)}} = X^{(2)}\hat{\tau}(\alpha)A + \hat{\Sigma}_{21}(\alpha)\hat{\Sigma}_{11}^{-1}(\alpha)(Y^{(1)} - X^{(1)}\hat{\tau}(\alpha)A)$. $\Sigma_{ab}$ is defined as before with $p_1 = p - 1$, $p_2 = 1$, $\hat{\tau}(\alpha)$ and $\hat{\Sigma}_{ab}(\alpha)$ are estimates of $\tau$ and $\Sigma_{ab}$ with the sample $Y^{(1)}$ under model $M_\alpha$, $Y = (Y^{(1)'}, Y^{(2)'})'$, $Y^{(1)}$ is $(p-1) \times N$, $Y^{(2)}$ is $1 \times N$, $X = (X^{(1)'}, X^{(2)'})'$, $X^{(1)}$ is $(p-1) \times m$, and $X^{(2)}$ is $1 \times m$. The model $M_{\alpha^*}$ corresponding to the minimum discrepancy measure is chosen as the most appropriate for the data. It is noted that $Y^{(2)}$ is viewed as $y$ and $Y^{(1)}$ as $Y$ in extended prediction of $y$ given $Y$. When $p > m + 2$, the following pseudo-cross-validation procedure advocated by Keramidas and Lee (1990), which is prequential in nature (see Dawid, 1984), can be used. The discrepancy measure

for this procedure is

$$D_\alpha \;=\; \frac{1}{(p-m-1)N} \sum_{j=m+2}^{p} (Y_j - \hat{Y}_j)(Y_j - \hat{Y}_j)'. \tag{11.5}$$

where $Y = (Y_1', \ldots, Y_p')'$, $\hat{Y}_j$ is the prediction of $Y_j$ and is obtained in exactly the same manner as $Y^{(2)}$ in (11.4) with the sample $(Y_1', \ldots, Y_{j-1}')'$. Thus, the prediction is still extended in nature except it is done in a sequential manner. In the prediction of $Y_j$, $p_1 = j-1$, $p_2 = 1$, and $X^{(2)}$ is the first $j-1$ rows of $X$, if all the available data are used as the sample in the prediction process. In case only a sbuset of available data is used as the sample, the choice of the design matrix $X^{(2)}$, $p_1$ and $p_2$ should be self-evident. As in the previous case, the model $M_{\alpha^*}$ corresponding to the minimum discrepancy measure is chosen as the most appropriate for the data.

When the data involve samples from two or more populations, model selection can be based on the ability to classify the data at hand. The model selected is the one that produces the smallest probability of misclassification using the PSR procedure.

Here we consider the situation where $g$ growth curves, as defined by (2.2), have been observed and a future observation matrix $V$, of dimension $p \times K$, is known to be drawn from one of $g$ populations, $\pi_1, \ldots, \pi_g$, with prior probability $q_1, \ldots, q_g$, respectively. It is also assumed that

$$V \;\sim\; N(X\tau_i F_i, \Sigma_i \otimes I),$$

where $F_i$ is a known design matrix formed by some columns of $A_i$, if $V$ is from $\pi_i$. For the selection of a model using the PSR procedure, we will set $K = 1$.

Let $Y = (Y_1, \ldots, Y_g) = (y_1, y_2, \ldots, y_N)$, where $N = \sum_{j=1}^{g} N_j$ and $\theta$ is the collection of parameters. Then $y_j$ is classified as $\pi_\alpha$ if

$$q_\alpha f_\alpha(y_j | Y_{(j)}, \hat{\theta}_{\alpha_{(j)}}, \pi_\alpha) \;>\; q_i f_i(y_j | Y_{(j)}, \hat{\theta}_{\alpha(j)}, \pi_i) \text{ for all } i \neq \alpha. \tag{11.6}$$

In case $g = 2$ and $\theta$ is known, (11.6) is equivalent to classifying $y_j$ into $\pi_1$ if

$$(y_j - X\tau_2 F_2)' \Sigma_2^{-1} (y_j - X\tau_2 F_2) + \log(|\Sigma_2|)$$
$$> (y_j - X\tau_1 F_1)' \Sigma_1^{-1} (y_j - X\tau_1 F_1) + \log(|\Sigma_1|) + 2\log(q_2/q_1). \tag{11.7}$$

If $\Sigma_1 = \Sigma_2 = \Sigma$, then (11.7) is further reduced to

$$[y_j - \frac{1}{2}(\tau_1 F_1 + \tau_2 F_2)]' X' \Sigma^{-1} X (\tau_1 F_1 - \tau_2 F_2) > \log(q_2/q_1) \tag{11.8}$$

25

which is an extension of equation (5) of Anderson (1984, p.205).

This procedure is quite practical, because the ability to classify the data correctly is very important. However, the procedure cannot be applied if there is no clear partitioning of the data into several distinct groups. Also, the sample size for each subgroup could be greatly reduced if the number of groups is large. For more detail of growth curve classification the reader is refered to Lee (1982).

## 12 Applications to Real Data

Several data sets have been used in the literature for the three types of prediction considered in this paper. The illustrative examples can be found in Lee and Geisser (1975), Fearn (1975), Rao (1977, 1984, 1987), Lee (1988,1991), Keramidas and Lee (1988, 1990, 1992). The data are all from biological applications with the exception of those in Keramidas and Lee (1988, 1990) in which the model was useful in technological substitutions.

Here we will restrict our attention to conditional prediction of $V^{(2)}$ given $V^{(1)}$ and $Y$ for the dental data, as reported by Potthoff and Roy (1964). Dental measurements were made on 11 girls and 16 boys at ages 8, 10, 12 and 14 years. Each measurement is the distance (in mm) from the center of the pituitary to the pteryomaxillary fissure. As noted in Lee and Geisser (1975), individual 20 ( the 9th boy ) is suspected to be an aberrant observation and will be excluded in this illustration. Since the measurements are taken once every two years, the design matrix $X$ is

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -3 & -1 & 1 & 3 \end{pmatrix}'$$  (12.1)

The design matrix $A$ is a $2 \times 26$ matrix composed of 11 (1,0) columns, followed by 15 (0,1) columns when both girls and boys are assumed to have a common covariance matrix. If the covariance matrices are distinct for girls and boys, the design matrix $A$ is a $1 \times 11$ vector for girls and a $1 \times 15$ vector, both consisting of all 1s.

Let $Y = (Y_1, \ldots, Y_{11}, Y_{12}, \ldots, Y_{26})$ where $Y_1$ through $Y_{11}$ represent the dental measurements of the girls and $Y_{12}$ through $Y_{26}$ are those of the boys. In conditional prediction of $V^{(2)}$ given $V^{(1)}$ and $Y$ we will consider the special case in which $V = (V^{(1)'}, V^{(2)'})'$, $V^{(1)}$ is $(p-1) \times 1$ and $V^{(2)}$ is $1 \times 1$. The discrepancy measure given by (11.1) will be applied with $p_2 = 1$.

We will first apply the results (5.11), (5.15), (5.21) and (5.22) to the case in which

$V^{(1)} = (26, 25, 29)'$ [corresponding to the first boy] and $V^{(1)} = (24.5, 25, 28)'$ [corresponding to the last girl]. The exact and approximate predictive densities of $V^{(2)}$ given $V^{(1)}$
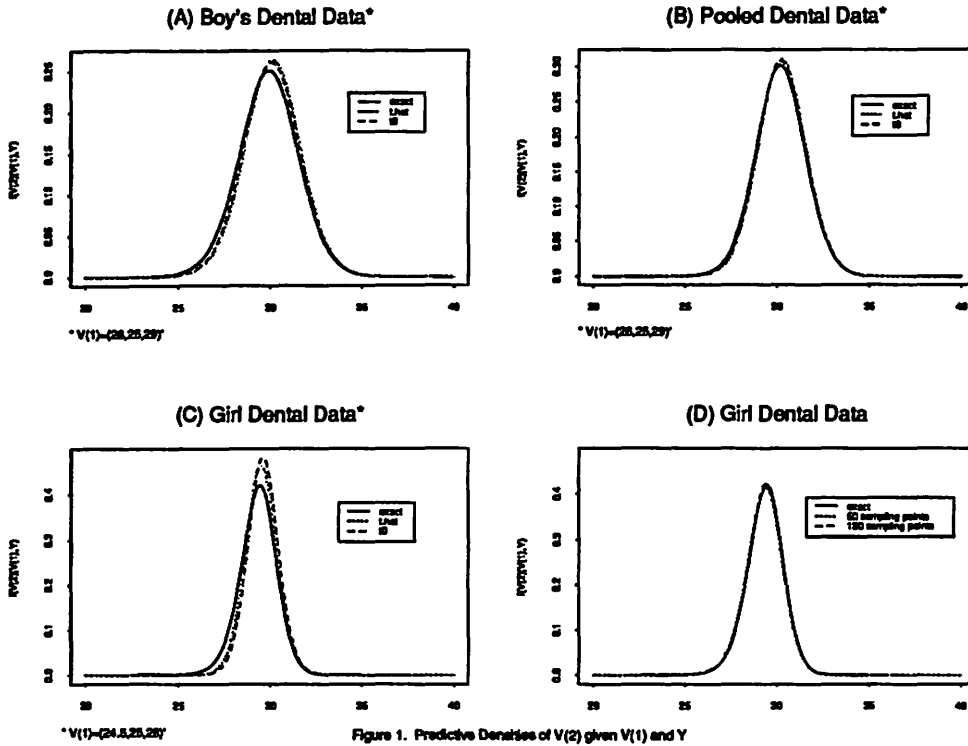


**(A) Boy's Dental Data\***

**(B) Pooled Dental Data\***

**(C) Girl Dental Data\***

**(D) Girl Dental Data**

Figure 1. Predictive Densities of V(2) given V(1) and Y

and $Y$ are shown in Figure 1. The difference between Figures 1(A) and 1(C) is in the treatment of the covariance matrix. In Figure 1(B) the dental data for both girls and boys are assumed to have an identical covariance matrix while in Figure 1(A) girls and boys are assumed to have different covariance matrices for their dental measurements. From Figures 1(A)-1(C) it is clear that the approximations given by (5.11) and (5.21) are quite adequate, at least as far as the predictive region is concerned. Meanwhile, even for the worst situation as shown in Figure 1(C), vast improvement can be accomplished via (5.22) as evidenced in Figure 1(D), which is a tremendous improvement over Figure 1(C) even for $L = 50$. As noted earlier, however, that an average of $L$ multivariate Student $t$ densities is no longer a multivariate Student $t$ density. Hence, an approximation such as (5.11) or (5.21) still has its place in our development unless the random variable of interest is one-dimensional in which (5.22) should be preferred.

27

We next compare the predictive accuracy for conditional prediction of $V^{(2)}$ given $V^{(1)}$ and $Y$ by the approximate means under four different covariance structures : Arbitrary p.d., RSS, Uniform and Serial. The approximate mean for arbitrary p.d. $\Sigma$ is obtained from (4.5) and (4.6) and the approximate means for RSS, Uniform and Serial structures are $\mu_{r2\cdot1}(\hat{t})$, $\mu_{u2\cdot1}$, and $\mu_{s2\cdot1}$, as given in (5.12), (6.3) and (7.4), respectively with $\hat{t}$ being the model of the $F$ distribution. The comparison is summarized in Table 1. The entries of this table are obtained from Lee and Geisser (1975) and Lee (1988). It appears that the serial structure is most appropriate for this data set.

Table 1. The MSD and MAD Between the Predicted and Actuals: Dental Data

|     | Arbitrary p.d. | RSS   | Uniform | Serial |
| --- | --- | --- | --- | --- |
| MSD | 2.128 | 2.035 | 2.196 | 1.354 |
| MAD | 1.144 | 1.127 | 1.206 | 0.940 |

# 13  Concluding Remarks

We have reviewed predictive methods for many of the growth curve models with varying covariance structures. It is noted that none of these models will represent a true description of the underlying process for most rather complex data. However, for a particular data set what is usually required is that model be adequate for predictive purposes. Hence, estimates of predictive error can be made by randomly dividing the data into a construction sample and a validation sample and predicting the validation sample from the construction sample for a large sample. For a small sample, one can delete an observation and predict it from the rest and calculate the error and then cycle through all the observations yielding an average predictive error. Although this will underestimate the actual predictive error, it should prove useful as a gauge of the adequacy of the particular model for prediction in any real problem and serve to discriminate between alternative models. It is also to be noted that in many of these complex situations we have suggested use of "plug in estimates" and large sample normal approximations. Since the multivariate Student $t$ distribution is somewhat more diffuse than the multivariate normal, it might be better to use the multivariate Student $t$ distribution when the sample size is not

overly large. Another inferential mode that may prove useful is the method of predictive likelihood, e.g., Butler (1986). However, it has not as yet been applied to growth curves.

# References

Anderson, T. W. (1984). An Introduction to Multivariate Statistical Analysis, 2nd edition. New York: Wiley.

BMDP Statistical Software Manual, Vol. 2 (1988). University of California Press, Berkeley, California.

Box, G.E.P. (1950). Problems in the analysis of growth and wear curves. Biometrics, 6, 362-89.

Butler, R. W. (1986). Predictive likelihood inference with applications (with Discussion). Jounral of the Royal Statistical Society, Series B, 48, 1-38.

Chi, E.M. and Reinsel, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors. Journal of the American Statistical Association, 84, 452-59.

Dawid, A. P. (1984). Present position and potential development: Some personal views. Journal of the Royal Statistical Society, Series A 147, 278-92.

Fearn, T. (1975). A Bayesian approach to growth curves. Biometrika 62, 89-100.

Geary, D. N. (1989). Modeling the covariance structure of repeated measurements, Biometircs,45, 1183-1195.

Geisser, S. and Cornfield, J.(1963). Posterior distributions for multivariate normal parameters. Journal of the Royal Statistical Society, Series B 25, 368-76.

Geisser, S. (1965). Bayesian estimation in multivariate analysis. Ann. Math. Statist. 36, 150-59.

Geisser, S. (1970). Bayesian analysis of growth curves. Sankhya, Series A 32, 53-64.

Geisser, S. (1974). A predictive approach to random effect model. Biometrika, 61, 101-7.

Geisser, S. (1975). The predictive sample reuse method with application. Journal of the American Statistical Association, 70, 320-8.

Geisser, S. and Eddy, W.F.(1979). A predictive approach to model selection . Journal of the American Statistical Association , 74, 153-160.

Geisser, S. (1980). Growth curve analysis. In Handbook of Statistics, Vol. 1,P.R. Krishnaiah(ed.), 89-115. Amsterdam: North-Holland.

Geisser, S. (1981). Sample reuse procedures for prediction of the unobserved portion of a partially observed vector. Biometrika, 68, 243-250.

Geyer, C. J. and Thompson, E. A.(1992). Constrained Monte Carlo maximum likelihood for dependent data. Journal of the Royal Statistical Society, Series B 54,657-700.

Jenrich, R.I. and Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. Biometrics, 42, 805-20.

Keramidas, E.M. and Lee , J. C.(1988). Forecasting technological substitutions with concurrent short time series. Proceedings of the Business and Economic Statistics Section, American Statistical Association, 1-10.

Keramidas, E.M. and Lee, J. C.(1990). Forecasting technological substitutions with concurrent short time series. Journal of the American Statistical Association. 85, 625-632.

Keramidas, E.M. and Lee, J. C. (1995). Selection of a covariance structure for growth curves. Biometrical Journal(to appear)

Khatri, C. G. (1966). A note on MANOVA model applied to problems in growth curve. Annals of the Institute of Statistical Mathematics, 18, 75-86.

Khatri, C. G. (1973). Testing some covariance structures under a growth curve model. Journal of Multivariate Analysis, 3, 102-116.

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. Biometrics,38, 963-74.

Lee, J. C. and Geisser, S. (1972). Growth curve prediction. Sankhya, Series A 34, 393-412.

Lee, J. C. and Geisser, S. (1975). Applications of growth curve prediction. Sankhya, Series A 37, 239-56.

Lee, J. C. and Hu, L. (1995). On the distribution of linear functions of independent $F$ and $U$ variates. Statistics and Probability Letters. (to appear)

Lee, J. C. (1982). Classification of growth curves. In Handbook of Statistics, Vol. II, P. R. Krishnaiah and L. W. Kanal(eds.), 121-37. Amsteram: North-Holland.

Lee, J. C. and Tan, W.Y. (1984). On the degree of polynomial in a general linear model. Communications in Statistics, Series A 13, 781-790.

Lee, J. C. (1988). Prediction and estimation of growth curves with special covariance structures. Journal of the American Statistical Association, 83, 432-440.

Lee, J. C. (1991). Test and model selection for the general growth curve model. Biometrics, 47,147-159.

Liu, S. I. (1995). Bayesian multiperiod forecasts for ARX models. Annals of the Institute of Statistical Mathematics (In press)

Lyung, G. M. and Box, G.E.P. (1980). Analysis of variance with autocorrelated observations, scandinavian Journal of Statistics, 7, 172-180.

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. (1958). Some statistical methods for comparison of growth curves. Biometrics, 14,1-17.

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 analyses. In Multivariate Analyses, Vol. II, P. R. Krishnaish(ed.), 87-103. New York: Academic Press.

Rao, C. R.(1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. Proceedings of the Fifth Berkeley Sympoium on Mathematical Statistics and Probability, 1, 355-372.

Rao, C. R.(1977). Prediction of future observations with special reference to linear models. In Multivariate Analysis, Vol. IV, P. R. Krishnaiah(ed.), 193-208. Amsterdam: North-Holland.

Rao, C. R. (1984). Prediction of future observations in polynomial growth curve models. In Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions, 512-520. Calcutta: Indian Statistical Institute.

Rao, C. R.(1987). Prediction of future observations in growth curve models. Statistical Science, 2, 434-471.

Reinsel, G. (1984). Estimation and Prediction in a Multivariate Random Effects Generalized Linear Model, Journal of the American Statistical Association, 79, 406-414.

Wishart, J. (1938). Growth rate determinations in nutrition studies with the bacon pig, and their analysis. Biometrika, 30,16-28.

Zellner, A., Bauwens, L and Van Dijk, H. K. (1988). Bayesian specification analysis and estimation of simultaneous equation models using Monte Carlo methods. J. of Econometrics, 38, 39-72.

Zerbe, G. O. (1979). Randomization analysis of the completely randomized design extended to growth and response curves. Journal of the American Statistical Association, 74, 215-221.