ANALYSIS OF MULTIPLE TIME SERIES BY

BAYESIAN AND EMPIRICAL BAYESIAN ANALYSIS

S. Weerahandi, J.V. Zidek

Tech. Report No. 96

SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY

HEALTH SCIENCES PROGRAM
RAND CORPORATION

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY

DEPARTMENT OF STATISTICS
UNIVERSITY OF BRITISH COLUMBIA
ANALYSIS OF MULTIPLE TIME SERIES BY BAYESIAN 
AND EMPIRICAL BAYESIAN NONPARAMETRIC METHODS

by

S. Weerahandi, University of Wisconsin

J. V. Zidek, University of British Columbia

Tech. Report # 96

May 1986

Prepared in part under support to SIMS from the United States Environmental Protection Agency
Department of Statistics, University of British Columbia, 2021 West Mall,
Vancouver, B.C. V6T 1W5
Analysis of Multiple Time Series by Bayesian and
Empirical Bayesian Nonparametric Methods

by

S. Weerahandi  J.V. Zidek

and

Univ. of Wisconsin  Univ. of British Columbia

May 1986
ABSTRACT

The problem considered here is an extended version of the problem of growth and dose response curves. While our methods are nonparametric both frequency theory-like and Bayesian solutions are proposed. As a consequence of the simplicity of the approach, proposed here, it is possible to construct exact credibility bands for growth curves using only one-dimensional numerical integrations.

Key words: Growth curves, Taylor series, Exchangeability, Noninformative priors.
INTRODUCTION: This report is about the statistical analysis of multiple time series, $S_i(t)$, $-\infty < t < \infty$, $i=1, \ldots, m+1$. Data $s_{ij} = S_i(t_{ij})$, $j = 1, \ldots, n_i$ for all $i$ are available and inference is about $S_{m+1}(t_{m+1}) = H$, say, and parameters of its distribution. Such series might represent, for example, the growth curves of the weights of randomly sampled biological monitors.

The method presented here is an extension of that of Weerantani and Zidek (1985). It exploits two important features of the situation. First, since the organisms are randomly sampled, the $\{S_i(t)\}$ are exchangeable random processes. Thus the data from any one series, $s_i = (s_{i1}, \ldots, s_{in})^T$ with $n = n_i$, carries some information about every other series; "borrowing from strength," in the terminology of empirical Bayes, is feasible. Second, each time series is relatively smooth and $s_{ij} = S_i(t_{ij})$ carries information about $S_i(t_{m+1})$ and hence, through exchangeability, about $H$, if $|x_{ij}|$ is small, where

$$x_{ij} = t_{ij} - t_{m+1}.$$ 

for all $i$ and $j$.

The method to be described in this report is nonparametric. No functional form for the $S_i(t)$ is prescribed and instead its smoothness is exploited to achieve a locally parametric form through a Taylor series representation of the $\{S_i\}$. If the $S$'s are regarded as having $p+1$ derivatives

$$s_{ij} = \beta_{i0} + x_{ij}\beta_{i1} + \cdots + x_{ij}^p\beta_{ip}/p! + \varepsilon_{ij} \quad (1.1)$$

where

$$\beta_{ik} = DS_i(t_{m+1}), \quad k = 0, \ldots, p$$
$$\epsilon_{ij} = D^{p+1} s_i(x^*) x_{ij}^{p+1}/(p+1)!$$

and $x^*$ is a point in the interval joining $t_{ij}$ and $t_{m+1}$. Let $n = n_i$,

$$\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{in})^T$$

and

$$X_i = (x_0^{i}, \cdots, x_p^{i})$$

where $X_i$ is an $n_i \times (p+1)$ matrix whose $j$-th row is

$$(1, x_{ij}^{1}, \cdots, x_{ij}^{p}/p!)$$

for all $i$. Let

$$\beta_i = (\beta_{i0}, \cdots, \beta_{ip})^T.$$

Then

$$s_i = X_i \beta_i + \epsilon_i, i=1, \cdots, m,$$  \hspace{1cm} (1.2)

a locally linear model at $t = t_{m+1}$ which may be analyzed in any of the conventional ways, by generalized least squares, for example.

As $t = t_{m+1}$ varies so would the $\beta$'s, $H = S_{m+1}(t_{m+1})$ and the corresponding inferences about these and related quantities. The analysis of the model in (1.2) and inferences about $H$ are the subject of Section 2. An application to the analysis of growth curves is given in Section 3.

There are a number of advantages of the proposed method over conventional growth curve and time series methods. Our method unlike its counterparts in growth curve analysis is non-parametric. Unlike time series analysis, almost any prior knowledge about these series can be accommodated through joint prior distributions for unspecified quantities like $H$ and the
\(\beta's\), for example. Alternatively, no prior knowledge is needed or introduced if empirical Bayes or generalized least squares approaches are adopted. The proposed method yields inference about not only \(H\) but its derivatives as well. Credibility (i.e. error) intervals and bands are readily found. The method may be adapted to allow finite parameter models as additive components of \(S_{m+1}\); this would enable discontinuous, for example point impact, functions, as well as trends, seasonability and so on to be incorporated in the analysis. Finally covariates can be included in the argument of \(S_{m+1}\); then multivariate versions of the Taylor expansion lead to an extension of the model in equation (1.2).

2. **INFEERENCE FOR MULTIPLE TIME-SERIES.** Uncertainty about the \([\beta_i]\) and \([\epsilon_i]\) in Model (1.2) may be represented by a joint probability distribution. It is assumed that the \([(\beta_i, \epsilon_i), i=1, \ldots, m]\) are independent. It is shown by Weerahandi and Zidek (ibid) that locally, for small values of \(|x_{ij}| = |t_{ij} - t_{m+1}|\), \(\epsilon_{ij}\) and \((\beta_{i0}, \ldots, \beta_{ip})\) are approximately independent and as they point out, for large values of \(|x_{ij}|\) the associated data is essentially "windowed out". Thus, to a first approximation, the assumed independence of \(\beta_i\) and \(\epsilon_i\) would seem to be justified. Since the a priori uncertainty about \(S_i(t_{ij})\) is bounded, \(\text{Var}(\epsilon_{ij}) = x_{ij}^{2p+2}h\) where \(h = 0[x_{ij}^{-(2p+2)}]\). However, for simplicity, it is assumed here that \(h = \sigma^2\). As Weerahandi and Zidek (ibid) point out, because data obtained at times remote from \(t_{m+1}\) are "windowed-out", this simplifying assumption has little effect on the inferred value of \(H\). In summary, \(\text{Cov}(\epsilon_i) = \sigma^2 D_1\) where here the \(n_1 \times n_1\) matrix \(D_1 = D\), say, is assumed to be diagonal with \(j\)-th diagonal element \(x_{ij}^{2p+2}\). In the application of Section 3, but perhaps not in general, this seems reasonable.
Prior opinion about $S_i$'s higher derivatives is bound to be vague. So we partition $\beta_i = \beta$ and $X_i = X$ conformably as

$$\beta = \begin{pmatrix} \gamma \\ \xi \end{pmatrix}, \quad X = (A,B)$$

where $\gamma$ and $\xi$ are (respectively), the vector of $S_i$'s $(q + 1)$ lowest and $(p - q)$ highest order derivatives (among the first $p + 1$) evaluated at $t = t_{m+1}$ and for simplicity the subscripts, $i$, have been deleted. It follows that

$$X_i \beta_i = A_i \gamma_i + B_i \xi_i.$$

It is assumed that the $\{\gamma_i\}$ are exchangeable, a reasonable assumption in the absence of special knowledge about particular sample items. In the spirit of Lindley and Smith (1972), the $\gamma$'s might well be regarded as a conditionally independent sample from a population with a $(q + 1)$ dimensional mean vector $\gamma_o$ and a covariance matrix $\Lambda$. Weerahandi and Zidek (1985) discuss the problem of modelling $\Lambda$ and show that when $q = 0$ or 1, $\Lambda = \text{diag}[\lambda_o^2, \ldots, \lambda_q^2]$ would be a reasonable assumption. This assumption might be true for $q \geq 2$ but this would need to be decided on a case-by-case basis.

This paper, an initial contribution to the possible development of a general nonparametric, Bayes or empirical Bayes theory of growth curve analysis, assumes all distributions have a multivariate normal form. Thus

$$\gamma_i | \gamma_o \sim N(\gamma_o, \Lambda)$$

$$\varepsilon_i \sim N(0, \delta D_i),$$

$$\varepsilon_i^2 \sim N(0, \delta D_i),$$

(2.1)
for all \( i = 1, \ldots, m \) and all of these random variables are conditionally independent given the covariance parameters. As well, \( H = \beta_{m+1,0} = \gamma_{m+1,0} \) is distributed as \( N(\gamma_{oo}, \lambda_o^2) \) independently of the other variables in (2.1).

Our assumptions imply that a priori given \( \gamma_o \) and \( \xi \),

\[
S_i = S = A\gamma_o + B\xi + \varepsilon_i, \quad i=1, \ldots, m
\]  

(2.2)

where \( \varepsilon_i = \varepsilon_i^1 \sim N(0, \delta^2D + \Lambda \Lambda^T) \) are independently distributed.

Two different modes of analysis are discussed below. In the first, diffuse priors are attached to \( \gamma_o \), the \( \zeta \)'s, \( \delta^2 \) and \( \Lambda \) and an (improper) Bayesian approach is developed. In the second approach, estimates of the unspecified parameters are obtained for the model in equation (2.2) by maximum likelihood and the required distribution theory is then obtained directly from (2.2).

2.1. **BAYESIAN RESULTS.** In implementing the first of these two approaches, the \( \{ \xi_i \} \), which are nuisance parameters, are eliminated. Suppose \( \xi_i \sim N(0,C) \) are i.i.d. random vectors where \( C \) is "large". Then marginally the \( S_i \)'s are independent and

\[
S_i \sim N(A_i\gamma_o, G_i)
\]

(2.3)

where

\[
G_i = \delta^2D_i + B_iCB_i^T + A_i\Lambda\Lambda_i^T.
\]

(2.4)

The covariance parameters, \( \delta^2 \) and \( \Lambda \), enter the model (2.3) through a complex structure and their uninformative prior distribution is by no means
obvious. In this report their Jeffrey's prior is adopted; it is computed in
the Appendix. This prior is defined in equations (A.1) and (A.18), and
evaluated with the help of equations (A.15) – (A.17). In general there is no
explicit expression for the prior. But for the application in Section 3 in
which the observation times are identical in all records and q = 0, this prior is

\[ p(\delta^2, \lambda_o^2) = \delta^{-2}(\delta^2 + \lambda_o^2 A^T V A)^{-1}. \]

Where \( A^T V A = I_D^{-1} - (I_D^{-1} B)(B_D^{-1} B)^{-1}(B_D^{-1} I) \).

**INFEERENCE ABOUT \( \gamma_o \).** An approximate posterior density for \((\gamma_o, \delta^2, \Lambda)\),
will now be found for large C by using equations (A.18) and, in equation
(2.3), the approximations (A.8) – (A.10). From equations (2.3), ignoring
irrelevant multiplicative factors,

\[
p(\gamma_o, \delta^2, \Lambda | \text{data}) \propto p(\delta^2, \Lambda) \times
\]

\[
(\prod |G|)^{-1/2} \exp \left[- \frac{1}{2} \sum (s - A\gamma_o)^T G^{-1} (s - A\gamma_o) \right]
\]

where, for simplicity, the product index, index of summation, and subscripts,
i, on s, G and A have been dropped in equation (2.6).

**LEMMA 2.1.** As \( C \to \infty \), \( (s - A\gamma_o)^T G^{-1} (s - A\gamma_o) = \delta^{-2} \left\{ (s - A\breve{\gamma})^T V (s - A\breve{\gamma}) +
\right\}

\[
(\gamma - \gamma_o)^T F_o^{-1} (\gamma - \gamma_o),
\]

where \( F_o = \delta^{-2} \Lambda + (A^T V A)^{-1}, \)

\[
\breve{\gamma} = (A^T V A)^{-1} (A^T V s), \ i=1, \ldots, m,
\]

and \( V = D^{-1} - D^{-1} B (B_D^{-1} B)^{-1} B_D^{-1} \).

(2.7)

(2.8)
PROOF.

Equation (A.9) asserts that
\[ L^{-1} = \delta^{-2}V. \]  
(2.9)

Thus, by equation (A.8),
\[ G^{-1} = \delta^{-2}[V - \delta^{-2}VAF^{-1}A^{T}V] \]
(2.10)

where
\[ F = (\delta^{-2}A^{T}VA + \Lambda^{-1}). \]
(2.11)

It follows that
\[ A^{T}G^{-1}A = \delta^{-2}[A^{T}VA - \delta^{-2}A^{T}VAF^{-1}A^{T}VA] \]
\[ = \delta^{-2}\Lambda^{-1}F^{-1}A^{T}VA, \]
i.e.
\[ A^{T}G^{-1}A = \delta^{-2}[\zeta + (A^{T}VA)^{-1}]^{-1} = \delta^{-2}F_{o}^{-1} \]
(2.12)
say, where \( \zeta = \delta^{-2}\Lambda = \text{diag}[\zeta_{o}^{2}, \ldots, \zeta_{q}^{2}] \) and \( F_{o} = \zeta + (A^{T}VA)^{-1} \). At the same time
\[ A^{T}G^{-1}s = \delta^{-2}[A^{T}Vs - \delta^{-2}A^{T}VAF^{-1}A^{T}Vs] \]
\[ = \delta^{-2}\Lambda^{-1}F^{-1}A^{T}Vs \]
\[ = \delta^{-2}\Lambda^{-1}F^{-1}A^{T}VA \]
i.e.
\[ A^{T}G^{-1}s = \delta^{-2}F_{o}^{-1}\gamma \]
(2.13)

Finally
\[ s^{T}G^{-1}s = \delta^{-2}[s^{T}Vs - \delta^{-2}s^{T}VAF^{-1}A^{T}Vs] \]
\[ = \delta^{-2}[s^{T}Vs - \delta^{-2}\gamma^{T}A^{T}VAF^{-1}A^{T}VA\gamma] \]
\[ = \delta^{-2}[s^{T}Vs - \gamma^{T}(I - \Lambda^{-1}F^{-1})A^{T}VA\gamma] \]
\[ = \delta^{-2}[s^{T}Vs - \gamma^{T}A^{T}VA\gamma + \gamma^{T}F_{o}^{-1}\gamma] \]
i.e.
\[ s^{T}G^{-1}s = \delta^{-2}[(s - A^{T})^{T}V(s - A^{T}) + \gamma^{T}F_{o}^{-1}\gamma] \]
(2.14)

By combining equations (2.12) - (2.14), we obtain
\[
(s - \hat{A} \gamma_o)^T \delta^{-2} (s - \hat{A} \gamma_o) + \delta^{-2} (\hat{\gamma} - \gamma_o) F_{\delta} \gamma_o
\]

thus proving the Lemma.

In view of this result define

\[
T = \sum_i (s_i - A \hat{\gamma}_i)^T V_i (s_i - A \hat{\gamma}_i)
\]

and

\[
S_i = (\hat{\gamma}_i - \gamma_o) F_{\delta} \gamma_o (\hat{\gamma}_i - \gamma_o).
\]

To further simplify equation (2.6) observe that

\[
|G|^{-1} = |(L + BCB^T)|^{-1}
\]

where \(L = (\delta^2 D + A^T A A^T)\). Thus with \(M = B^T L^{-1} B + C^{-1}\)

\[
|G|^{-1} = |L^{-1} - B^{-1} B^{-1} B^T L^{-1}|
\]

\[
= |L^{-1}| |I - L^{-\frac{1}{2}} B^{-1} B^T L^{-\frac{1}{2}}|
\]

\[
= |L^{-1}| |I - M^{-\frac{1}{2}} B^T L^{-1} B M^{-\frac{1}{2}}|
\]

\[
= |L^{-1}| |M^{-1} M - B^T L^{-1} B|.
\]

Thus

\[
|G|^{-1} = |L^{-1}| |M|^{-1} |C^{-1}|.
\]

But for large \(C\), \(|M|^{-1} = |B^T L^{-1} B|^{-1}\) approximately. Now \(L^{-1} = \delta^{-2} (D + A \tau A^T)^{-1} = \delta^{-2} (D^{-1} - D^{-1} A \hat{M}_1 A^T D^{-1})\) say where \(\hat{M}_1 = (A^T D^{-1} A + \zeta^{-1})\). Thus

\[
|G|^{-1} = (\delta^{-2})^n -(p-q) |D^{-1}| |I - M^{-\frac{1}{2}} A^T D^{-1} A^{-1} M^{-\frac{1}{2}}|
\]

\[
\times |B^T D^{-1} B|^{-1} |I - M^{-\frac{1}{2}} B^T D^{-1} (B^T D^{-1} B)^{-1} B D^{-1} A M^{-\frac{1}{2}}|^{-1}.
\]

where \(n = n_1\). Thus

\[
|G|^{-1} = (\delta^{-2})^n -(p-q) |D^{-1}| |B^T D^{-1} B|^{-1} |A^T V A|^{-1} |\zeta + (A^T V A)^{-1}|^{-1} |C^{-1}|
\]

(2.16)

In the calculation of the approximate posterior density function, the factor, \(|C^{-1}|\), in equation (2.16) cancels from numerator and denominator. It,
along with the factors of (2.16) which do not depend on the hyperparameters, are irrelevant so we may write

\[ |G^{-1}| \propto (\delta^{-2})_n \cdot (p-q) |\zeta + (A^T V A)^{-1}| \]  

(2.17)

Equations (2.16) and (2.17) imply that equation (2.6) may be rewritten as

\[
p(\gamma_o, \delta^2, \Lambda | \text{data}) \propto p(\delta^2, \Lambda) \delta^{-\frac{M}{2}} \prod_i |F_{oi}|^{-\frac{1}{2}} \times \exp[-\delta^{-2}(T + S_1)/2)] 
\]

(2.18)

where \( M = N - (p - q)m \), and \( N = \sum_i n_i \). The change of variables, \( \zeta = \delta^{-2} \Lambda \) further simplifies the result and gives

\[
p(\gamma_o, \delta^2, \zeta | \text{data}) \propto \delta^{-(M+2)} g(\zeta) \prod_i |F_{oi}|^{-\frac{1}{2}} \times \exp(-\delta^{-2}(T + S_1)/2),
\]

(2.19)

where \( g(\zeta) = p(1, \zeta) \) as indicated following equation (A.18).

The marginal posterior density of \((\gamma_o, \zeta)\), obtained from equation (2.19), is

\[
p(\gamma_o, \zeta | \text{data}) \propto g(\zeta) \prod_i |F_{oi}|^{-\frac{1}{2}} (T + S)^{-M/2}.
\]

(2.20)

Observe that

\[
S = \sum_i (\hat{\gamma}_i - \gamma_o) F_{oi}^{-1} (\hat{\gamma}_i - \gamma_o)
\]

\[
= (\gamma_o - \hat{\gamma}_o)^T \Gamma (\hat{\gamma}_o - \gamma_o) + S_o
\]

where

\[
\Gamma = \sum_i F_{oi}^{-1}
\]

\[
S_o = \sum_i (\hat{\gamma}_i - \hat{\gamma}_o) F_{oi}^{-1} (\hat{\gamma}_i - \hat{\gamma}_o),
\]
and

\[ \hat{\gamma}_o = (\sum F_{oi})^{-1} (\sum F_{oi}^{-1} \hat{\gamma}_i). \]

Consequently equation (2.20) may be rewritten as

\[
p(\gamma_o, \zeta | \text{data}) = g(\zeta) \pi |F_o|^{-\frac{N}{2}} (T + S_o)^{-N/2} \times [1 + (M-q-1)^{-1} (\gamma_o - \hat{\gamma}_o)^T \Gamma_o (\gamma_o - \hat{\gamma}_o)]^{-M/2}
\]

where \( \Gamma_o = (M - q - 1)(T + S_o)^{-1} \Gamma \) and \( F_{oi} = \zeta^T A_{i1} A_{i1}^{-1} \). Equation (2.21) says in particular that a posteriori, \( \gamma_o \) has the multivariate t distribution with \( M - q - 1 = N - pm + (m - 1)q - 1 \) degrees of freedom, location vector \( \hat{\gamma}_o \) and precision matrix \( \Gamma_o \). In other words,

\[
\Gamma_o(\gamma_o - \hat{\gamma}_o) \sim t_{M-q-1}
\]

independently of \( \zeta \) which then has the marginal posterior density

\[
p(\zeta|\text{data}) = g(\zeta) \pi |F_o|^{-\frac{N}{2}} (T + S_o)^{-N/2} |\Gamma_o|^{-\frac{N}{2}}.
\]

Equations (2.22) and (2.23) provides a basis for inference about \( \gamma_o \).

**INFERENC ON H.** Of more central interest is \( H = S_{m+1}(t_{m+1}) \). Recall that

\( H|\gamma_o, \delta^2, \zeta \sim N(\gamma_o, \delta^2 \zeta_o) \) a priori. Thus

\[
p(H, \gamma_o, \delta^2, \zeta | \text{data}) = p(H|\gamma_o, \delta^2, \zeta) p(\gamma_o, \delta^2, \zeta | \text{data})\]
since $H$ is independent of the data given $\gamma_o$, $\delta^2$, and $\zeta$.

Thus

$$p(H, \gamma_o, \delta^2, \zeta | \text{data}) = g(\zeta) \zeta_o^{-\frac{1}{2}} \prod_i \left| F_{oi} \right|^{-\frac{1}{2}} (\delta^2)^{(M+3)/2}$$

and

$$X \exp[-\frac{1}{2}\delta^{-2}(T + S_o + T_o)]$$

where

$$T_o = \zeta_o^{-1} (H - \gamma_{oo})^2 + (\gamma_o - \hat{\gamma}_o)^T \Gamma (\gamma_o - \hat{\gamma}_o).$$

Let $\Gamma_1^\perp = \zeta_o^{-1} \Delta_1^{\perp}$, where $\Delta_1$, denotes the $(q + 1) \times (q + 1)$ matrix all of whose elements are zero except for the first diagonal element which is 1. And let $a$ be any $(q + 1)$ dimension column vector whose first element is 1. Then using the formula for combining quadratic forms (c.f. Box and Tiou 1973, p. 418)

$$T_o = (\gamma_o - Ha)^T \Gamma_1 (\gamma_o - Ha) + (\gamma_o - \hat{\gamma}_o)^T \Gamma (\gamma_o - \hat{\gamma}_o)$$

$$= (\gamma_o - \hat{\gamma}_o)^T (\Gamma + \Gamma_1^\perp) (\gamma_o - \hat{\gamma}_o) + Q_o$$

where

$$\hat{\gamma}_o = (\Gamma + \Gamma_1^\perp)^{-1} (H \Gamma_1 a + \Gamma_1^\perp \gamma_o).$$

and

$$Q_o = (Ha - \hat{\gamma}_o)^T \Gamma_1 (\Gamma + \Gamma_1^\perp)^{-1} \Gamma (Ha - \hat{\gamma}_o).$$

But

$$(\Gamma + \Gamma_1^\perp)^{-1} \Gamma = (I + \Gamma_1^{-1} \Gamma_1)^{-1}$$

and

$$= \begin{bmatrix}
(G^{11} + \zeta_o)^{-1} \zeta_o^{-1} & 0 \\
-(G^{11} + \zeta_o)^{-1} \Gamma (2) & I
\end{bmatrix}$$
where $[\Gamma_{(2)}]^{T} = (\Gamma_{21}, \ldots, \Gamma_{q+1,1})$. Thus

$$Q = (\Omega - \hat{\Omega}_{o})^2(\Gamma_{11} + \zeta_{o}^{-1})^{-1},$$  \hspace{1cm} (2.24)

and

$$T = (\Omega - \hat{\Omega}_{o})^{T}(\Gamma + \Gamma_{1})(\Omega - \hat{\Omega}_{o}) + (\Omega - \hat{\Omega}_{o})^2(\Gamma_{11} + \zeta_{o}^{-1})^{-1} \hspace{1cm} (2.25)$$

The marginal density of $(H, \delta^2, \zeta)$ is now easily found, since $\gamma_{o}$ is assumed to have a uniform (improper) distribution; it is

$$p(H, \delta^2, \zeta|\text{data}) \propto g(\zeta)\gamma_{o}^{-\frac{N}{2}}|F_{o}|^{-\frac{N}{2}}(\delta^{-2})(M+3)^{1/2}
\times \Gamma + \Gamma_{1} \Gamma_{I}\frac{1}{2}
\times \exp[-\frac{1}{2}\delta^{-2}(T + S_{o} + (\Omega - \hat{\Omega}_{o})^2(\Gamma_{11} + \zeta_{o}^{-1})^{-1})].$$

Since $|\delta^{-2}(\Gamma + \Gamma_{1})|^{1/2} = (\delta^{-2})^{-(q+1)/2}|\Gamma + \Gamma_{1}|^{-1/2}$, the marginal distribution of $(H, \zeta)$ is in turn

$$p(H, \zeta|\text{data}) \propto g(\zeta)\gamma_{o}^{-\frac{N}{2}}|F_{o}|^{-\frac{N}{2}}|\Gamma + \Gamma_{1}|^{-\frac{N}{2}}
\times \exp[-\frac{1}{2}(T + S_{o} + (\Omega - \hat{\Omega}_{o})^2(\Gamma_{11} + \zeta_{o}^{-1})^{-1})(M-q)^{1/2}] \hspace{1cm} (2.26)$$

It follows that $H$ has the student's $t$ distribution with $M - q - 1$ degrees of freedom, location $\mu_{H} = \hat{\Omega}_{o}$ and precision $\sigma_{H}^{2} = (T + S_{o})(M - q - 1)^{-1}(\Gamma_{11} + \zeta_{o}^{-1})$ i.e.

$$\frac{(H - \mu_{H})}{\sigma_{H}} \sim t_{M-q-1},$$ \hspace{1cm} (2.27)

independently of the nuisance parameters, $\zeta$. Equation (2.27) together with (2.23) serves as a basis for inferences about $H$. 
2.2. FREQUENCY THEORY RESULTS.

In the problem under consideration, a frequency theory-like approach can be taken by assuming the mixed model,

\[ s_i = A_i \gamma_i + \beta_i \xi_i + \epsilon_i, \quad i = 1, 2, \ldots, m, \]  

(2.28)

where \( \gamma_i \sim N(\gamma_0, \Lambda) \) and \( \epsilon_i \sim N(0, \delta^2 D_i) \), as the starting point.

Eliminating the unobservable parameters, \( \gamma_i \), from the model we get

\[ s_i = A_i \gamma_0 + B_i \xi_i + \epsilon_i^1, \quad i = 1, 2, \ldots, m, \]  

(2.29)

where \( \epsilon_i^1 \sim N(0, \delta^2 D_i + A_i \Lambda A_i^T) \).

MAXIMUM LIKELIHOOD ESTIMATES. In order to find MLE's of the unknown parameters, first fix \( \delta^2 \) and \( \Lambda \). Then the log likelihood function can be written as

\[ L^* = k - \frac{1}{2\delta^2} \sum_i (s_i - A_i \gamma_0 - B_i \xi_i)^T H_i (s_i - A_i \gamma_0 - B_i \xi_i), \]

where \( H_i = (D_i + A_i \zeta \Lambda_i^T)^{-1} \), \( \zeta = \delta^{-2} \Lambda \), and \( k = \frac{1}{2} \sum \ln|H_i| - m \ln \delta + c, \) \( c \) being a constant. Notice that the MLE of each nuisance parameter, \( \xi \), can be expressed in terms of other parameters as \( \xi = (B_i^T H_i B_i)^{-1} B_i^T H_i (s_i - A_i \gamma_0) \). At this MLE, the log likelihood function of the other parameters reduces to

\[ L^* = k - \frac{1}{2\delta^2} \sum_i (s_i - A_i \gamma_0)^T H_i (s_i - A_i \gamma_0), \]

where \( H_i = H_i - H_i B_i (B_i^T H_i B_i)^{-1} B_i^T H_i. \) But

\[ (s - A \gamma_0)^T H (s - A \gamma_0) = (s - A \gamma_0)^T [V^{-1} + A \Lambda A^T]^{-1} (s - A \gamma_0), \]  

(2.30)

where \( V_i = D_i^{-1} - D_i^{-1} B_i (B_i^T D_i^{-1} B_i)^{-1} B_i^T D_i^{-1} \). This identity can be easily established, for instance by integrating the function \( \exp\{-(s - A x - B y)^T D_i^{-1} (s - A x - B y) - (x - \gamma_0)^T \zeta^{-1} (x - \gamma_0)\} \) w.r.t. \((x, y)\) and \((y, x)\) respectively.

Using the identity \( (V^{-1} + A \Lambda A^T)^{-1} = V - V A (A^T V A + \zeta^{-1})^{-1} A^T V, \) (2.30) can be decomposed as
\[(s - A\gamma_o)^T(H(s - A\gamma_o) = (s - A\gamma)^T V(s - A\gamma) + (\gamma - \gamma_o)^T A^T (V^{-1} + A\zeta A^T)^{-1} A(\gamma - \gamma_o)
\]
\[= (s - A\gamma)^T V(s - A\gamma) + (\gamma - \gamma_o)^T [(A^T V A) + \zeta^{-1}](\gamma - \gamma_o), ~ (2.31)\]

where \(\gamma_i = (A^T V A)^{-1}(A^T V A s_i).\) Hence the log likelihood function of \(\alpha_o\) can be expressed as

\[L^* = k \cdot \frac{1}{2\delta^2} \{ T + \sum S_i \}, ~ (2.32)\]

where \(T = \sum (s_i - A_i \gamma_i)^T V_i (s_i - A_i \gamma_i), \quad F_{oi} = \zeta + (A^T i_i V A)^{-1}\) and \(s_i = (\gamma_i - \gamma_o)^T F_{oi}^{-1}(\gamma_i - \gamma_o).\) Now it is evident that the MLE of \(\gamma_o\) is \(F_{oi}\)

\[\gamma_o = (\sum F_{oi}^{-1})(\sum F_{oi}^{-1} \gamma_i). \quad (2.33)\]

Notice that \(\gamma_o\) is the same as the mean of the posterior distribution given in (2.23). It is also of interest to note that, when the observation times of the \(m\) subjects are identical, \(\gamma_o\) reduces to the simple average \(\gamma_o = \sum \gamma_i/m\) which is independent of unknown variances.

In order to estimate \(\delta^2\) and \(\zeta, L^*\) can be evaluated as

\[L^* = c - N \ln \delta - \frac{1}{2} \sum \ln |D_i + A_i \zeta A_i^T| - \frac{1}{2\delta^2} \{ T + S_o \}, \quad (2.34)\]

where \(S_o = \sum (\gamma_i - \gamma_o)^T F_{oi}^{-1}(\gamma_i - \gamma_o).\) The MLE's of \(\delta^2\) and \(\zeta\) can now be found by the numerical maximization of \(L^*.\) More explicit formulas for these MLE's for the important particular case \(q = 0\) will be given in Section 3.

When the MLE of \(\zeta\) is known, one can use (2.33) to find the MLE of \(\gamma_o\).

**UNBIASED ESTIMATES.** Since \(V_i B_i = 0\) and \(V_i D_i V_i = V_i\), it follows from (2.28) that

\[\hat{V}_i s_i = V_i A_i \gamma_i + u_i, \quad (2.35)\]

where \(u_i = V_i \epsilon_i \sim N(0, \delta^2 V_i).\) Therefore \(\gamma_i = (A^T i_i V A)^{-1} A_i^T V_i s_i\) can be expressed as

\[\gamma_i = \gamma_i + (A^T i_i V A)^{-1} A_i^T u_i, \quad (2.36)\]
so that

\[ \hat{\gamma}_i - \gamma_i \sim N(0, \delta^2 (A_i^T V_i A_i)^{-1}). \]  
(2.37)

In order to find an unbiased estimator for \( \delta^2 \), re-express \( e_i = s_i - A_i \hat{\gamma}_i \) as

\[ e_i = \beta_i \xi_i + M_i \epsilon_i, \]  
(2.38)

where \( M_i = I_n - A_i (A_i^T V_i A_i)^{-1} A_i^T V_i \). Notice that

\[ e_i^T V_i e_i = \epsilon_i^T M_i V_i M_i \epsilon_i \]
(2.39)

\[ = \text{tr}(M_i^T V_i M_i \epsilon_i \epsilon_i^T) \]

and hence \( \text{E}(e_i^T V_i e_i) = \delta^2 \text{tr}(V_i (M_i D_i M_i^T)) \).

But

\[ V(\text{MDM}^T) = VD - VA(A^T VA)^{-1} A^T VD \]  
(2.40)

is idempotent and has rank \( n_i - p - 1 \). Thus \( \text{E}(e_i^T V_i e_i) = \delta^2 (n_i - p - 1) \) and,

in turn, an unbiased estimate of \( \delta^2 \) is obtained as

\[ \hat{\delta}^2 = T / (N - pm - m), \]  
(2.41)

where \( T = \sum (s_i - A_i \hat{\gamma}_i)^T V_i (s_i - A_i \hat{\gamma}_i) \).

It immediately follows from (2.37) and the distribution of \( \gamma_i \) given in (2.28) that \( \hat{\gamma}_o \) given in (2.33) is in fact an unbiased estimate of \( \gamma_o \).

However \( \hat{\gamma}_o \) can be used to estimate \( \gamma_o \) only when it is independent of unknown parameters. This is, for instance, the case when the observation times of the \( m \) subjects are identical. In any case, a good unbiased estimator which one can always employ in the estimation of \( \gamma_o \) is

\[ \overline{\gamma}_o = \frac{1}{m} \sum \hat{\gamma}_i. \]  
(2.42)

Consequently, we can proceed to obtain a good unbiased estimator of \( \Lambda \) by using the identity
\[
\sum_{i} (\hat{\gamma}_i - \bar{\gamma})(\hat{\gamma}_i - \bar{\gamma})^T = \sum_{i} (\gamma_i - \gamma_o)(\gamma_i - \gamma_o)^T - m(\bar{\gamma} - \gamma_o)(\bar{\gamma} - \gamma_o)^T
\]

As is clear from (2.36), \( \hat{\gamma}_i \sim N(\gamma_o, \Lambda + \delta^2 (A_{iV}^T A_i)^{-1}) \), i = 1, \ldots, m are independent and hence the expected values of the terms in the above identity can be evaluated to get
\[
E(\sum_{i} (\gamma_i - \bar{\gamma})(\gamma_i - \bar{\gamma})^T) = m \Lambda + \delta^2 \sum_{i} (A_{iV}^T A_i)^{-1} (1 - 1/m)
\]
\[
= (m - 1) \Lambda + m^{-1} E(\delta^2) \sum_{i} (A_{iV}^T A_i)^{-1}.
\]

Hence an unbiased estimator of \( \Lambda \) is found as
\[
\hat{\Lambda} = (m - 1)^{-1} \bar{W} - [m^{-1} T/(N - pm - m)] C
\]
with
\[
\bar{W} = \sum (\hat{\gamma}_i - \bar{\gamma})(\hat{\gamma}_i - \bar{\gamma}) \quad \text{and} \quad C = \sum (A_{iV}^T A_i)^{-1}.
\]

**CONFIDENCE REGIONS FOR \( \gamma_o \).** It follows from (2.39) that
\[
e_{iV}^T e_i = \delta^2 \sum \lambda_{ij} Z_{ij}
\]
where \( Z_{ij} \) are i.i.d. \( N(0,1) \) random variables and \( \lambda_{ij}, j = 1, \ldots, n_i \) are the latent roots of \( D_{iV}^T D_{iV} M_{iV}^{-1} U_{iV} \) or of \( V_{iV} M_{iV}^{-1} \). It is seen in (2.40) that this quantity is idempotent and has trace \( n_i - p - 1 \). Hence \( e_{iV}^T e_i \sim \chi^2_{n_i - p - 1} \), \( i = 1, \ldots, m \) are independent and consequently we get the distribution of \( T \) as
\[
T/ \delta^2 \sim \chi^2_{N - pm - m}
\]

To get a result similar to the Bayesian one, let us first fix \( \zeta = \delta^{-2} \Lambda \). Then we can exploit the identity
\[
\sum_{i} (\gamma_i - \gamma_o) F_{oi}^{-1} (\gamma_i - \gamma_o) = \sum_{i} (\gamma_i - \gamma_o) F_{oi}^{-1} (\gamma_i - \gamma_o)
\]
\[
- \sum_{i} (\gamma_o - \gamma_o) F_{oi}^{-1} (\gamma_o - \gamma_o)
\]

and the fact that
\[
\delta^{-1} F_{oi}^{-1} (\gamma_i - \gamma_o) \sim N(0, I_{q+1})
\]
(2.45) to get the distribution of \( S_o = \sum (\gamma_i - \gamma_o) F_{oi}^{-1} (\gamma_i - \gamma_o) \) as
\[ S_{o} / \delta^2 \sim \chi^2_{(m-1)(q+1)} \]

Since \( \hat{\gamma} \) and \((s_i - A_i \hat{\gamma})^T v_i (s_i - A_i \hat{\gamma})\) are independent for each \(i\), \(T\) and \(S_o\) are also independent and thus, we have

\[ (T + S_o) / \delta^2 \sim \chi^2_{M-q-1} \]

where \(M = N - pm - qm\). Moreover \(T + S_o\) is distributed independently of

\[ F^{1/2} (\hat{\gamma}_o - \gamma_o) / \delta \sim N(0, I_{q+1}) \]

where \(F = \sum \Gamma^{-1}_o\). From (2.47) and (2.48) it immediately follows that (c.f. Muirhead 1982, page 48)

\[ \Gamma_o (\hat{\gamma}_o - \gamma_o) \sim T_{M-q-1} \]

where \(\Gamma_o = (M-q-1) (T + S_o)^{-1} \Gamma\). This is the frequency theory counterpart of the Bayesian result given in (2.22). Since \(\zeta\) is in fact unknown, one can estimate \(\gamma_o\) and \(\Gamma_o\) by using \(\zeta = \delta^2 \hat{\gamma}\) in place of \(\zeta\). Hence (2.49) can be employed to construct approximate confidence regions for \(\gamma_o\). In particular, a result useful in constructing approximate confidence ellipsoids is

\[ (q+1)^{-1} (\hat{\gamma}_o - \gamma_o)^T \Gamma_o (\hat{\gamma}_o - \gamma_o) \sim F_{q+1, M-q-1} \]

CONFIDENCE INTERVALS FOR \(H\). Recall that the distribution of \(H = S_{m+1} / T_{m+1}\) is given by

\[ H \sim N(\gamma_{oo}, \delta^2 \zeta_o) \]

\(\gamma_{oo}\) and \(\zeta_o\) being the first components of \(\gamma_o\) and \(\zeta\). Moreover \(H\) is distributed independently of

\[ \hat{\gamma}_{oo} \sim N(\gamma_{oo}, \delta^2 \Gamma^{11}) \]

where \(\Gamma^{11}\) is the first diagonal element of \(\Gamma^{-1}\). Hence,

\[ \hat{\gamma}_{oo} - H \sim N(0, \delta^2 (\zeta_o + \Gamma^{11})) \]

and by eliminating \(\delta^2\) using (2.47) we get
\[(\hat{\gamma}_o - H)\sigma_H^{-1} \sim t_{N-q-1}\]  

where \(\sigma_H^2 = (T + S_0)(\zeta + I_{11}/(N-q-1)).\) This is the frequency theory counterpart of the Bayesian result given in (2.27). Approximate confidence intervals for \(H\) can thus be constructed.

3. EXAMPLE. Here attention is confined to growth curve analysis and the required computational procedures are illustrated by a simple example. In order to summarize the necessary results as applied to this particular case, assume that observations are available on the growth curves of \(m\) subjects, all at identical times \(t_1, t_2, \ldots, t_m.\) Let \(t_m+1\) be the time at which the future individual or representative is to be observed. In our illustrative example we shall make the reasonable assumption that the growth curves are locally linear so that (1.1) reduces to

\[s_{ij} = \gamma_i + \beta_j (t_j - t_{m+1}) + \epsilon_{ij},\]  

where \(\gamma_i = S_i(t_{m+1}), \beta_j = S_j(t_{m+1})\) and \(\epsilon_{ij} \sim N(0, \delta^2(t_j - t_{m+1})^2)\) as we have assumed. Hence the growth curve of \(i\)th subject is governed by the linear model

\[s_i = \gamma_i 1 + B\beta_i + \epsilon_i; \epsilon_i \sim N(0, \delta^2 D); i = 1, \ldots, m,\]  

where \(D\) is an nxn diagonal matrix, \(1\) is an nx1 vector of 1's, \(B\) is an nx1 vector in the above setup and more generally an nxp matrix when \(\beta_i\) is formed by \(p\) derivatives of \(S_i(t_{m+1}).\)

In the growth curve situation it is customary to make the exchangeability assumption only on \(S_i(t_{m+1});\) no assumption is made on the slope of growth curves, etc. (c.f. Fearn (1975)). Hence in this case (2.1) reduces to

\[\gamma_i \mid \gamma_o \sim N(\gamma_o, \lambda), i=1, \ldots, m+1,\]  

(3.3)
where \( \gamma_0 \) and \( \lambda \) are the unknown parameters of interest.

In this case the MLE of \( \gamma_0 \) given by (2.33), as well, the posterior mean of \( \gamma_0 \) given by (2.22) reduces to
\[
\hat{\gamma}_0 = \frac{\sum_{i=1}^{m} \hat{\gamma}_i}{m},
\]
where \( \gamma_i = 1^T V X_i / 1^T V 1 \) and \( V = D^{-1} - D^{-1} B (B^T D^{-1} B)^{-1} B^T D^{-1} \). The MLE of \( \zeta = \lambda / \delta^2 \) found by maximizing (2.34) is the positive solution of the quadratic equation
\[
(c_1 + \zeta)^2 T + (c_1 + \zeta) W = (c_2 + \zeta),
\]
where \( c_1 = 1^T V 1 \), \( c_2 = 1^T D^{-1} \), \( W = \sum (\hat{\gamma}_i - \hat{\gamma}_0)^2 \) and \( T = \sum (s_i - \hat{\gamma}_i)^T V (s_i - \hat{\gamma}_i) \). In turn, the MLE of \( \delta^2 \) is found as
\[
\delta^2 = \frac{T + W / (c_1 + \zeta)}{mn}
\]
It is seen from (2.22) and (2.49) that, in a frequency theory-like interpretation as well as in a Bayesian interpretation
\[
(\hat{\gamma}_0 - \gamma_0) [((m k) / (W + (c_1 + \zeta) T))^{1/2}] \zeta \sim t_k \]
where \( k = nm - pm - 1 \). The analogous distribution result for doing inference on \( H = S_{m+1} (t_{m+1}) \) is
\[
(\hat{\gamma}_0 - H) (c_1 + \zeta (m + 1))^{-1} (W + (c_1 + \zeta))^{1/2} \zeta \sim t_k
\]
Inference on the nuisance parameter \( \zeta \) appearing in above results can be done using the F distribution given by (2.44) and (2.46). Namely,
\[
[((c_1 + \zeta)^T m^{-1} (n-p-1)^{-1}) / (W (m-1)^{-1})] \sim F_{m (n-p-1), m-1}
\]
Notice that (3.7) and (3.9) provides a complete distribution theory for doing inference on \( (\gamma_0, \zeta) \). But, we have not been able to construct exact confidence intervals for \( \gamma_0 \). So, in our numerical illustration we shall construct, exact Bayesian credibility intervals with noninformative reference priors, rather than resorting to approximate confidence intervals. The result which enables us to do this is the exact posterior distribution of \( \zeta \) given by
(2.23) as
\[ P(\zeta \mid \text{data}) \propto (c_1 + \zeta)^{r/2-1}(T + W/(c_1 + \zeta))^{-k/2}, \zeta > 0 \quad (3.10) \]

where \( r = m(n-p-1) \) and \( k = mn - mp - 1 \). As a comparison of the frequency theory-like approach and the Bayesian approach it is of interest to note that, the posterior mode of \( \zeta \) is

\[ \tilde{\zeta} = (W/T)(r-2)(m+1)^{-1} - c_1 \quad (3.11) \]

whereas the likelihood function of \( \zeta \) given by (3.9) alone is maximized by

\[ \zeta = (W/T)(r-2)(m-3)^{-1} - c_1 \quad (3.12) \]

NUMERICAL ILLUSTRATION.

As an example of a growth curve situation, we use the ramus height data used by Grizzle and Allen (1969) in their frequency theory solution and Fearn (1975) in his Bayesian solution to the growth curve problem. These data consist of the ramus heights, measured in m.m., of 20 boys at 8, 8 1/2, 9 and 9 1/2 years of age. The data were collected to establish a normal growth curve for the use of orthodontists.

The observed values of individual growth curves and the estimated mean growth curve are shown in Figure 3.1. The latter is obtained by computing \( \gamma_o(t_{n+1}) \) for a range of values of \( t_{n+1} \) using the formula given in (3.4). The 95\% point-wise Bayesian confidence band of \( H \) shown in the same graph was obtained with the aid of the equations (3.8) and (3.10).

In the computation of the exact half length of 95\% credibility intervals, an estimate was first obtained using (3.8) with \( \zeta \) replaced by \( \tilde{\zeta} \). Having found a neighbourhood of the 97.5th percentile of \( H \) in this manner, the exact percentile \( \gamma_o + z \) is found by solving the well behaved nonlinear equation.

\[ E[T_k(Z/SE(\zeta))] = .975 \quad (3.13) \]
where $T_k$ is the c.d.f. of student's $t_k$ distribution, $SE(\zeta) = ((c + \zeta(1+m))(W + (c + \zeta))/[mk(c + \zeta)])^{1/2}$ and the expectation is with respect to the posterior density of $\zeta$ given by (3.10). For instance, this equation can conveniently be solved using the IMSL subroutine ZFALSE. This requires only a very few evaluations of the expectation appearing in (3.13). In our computation, this one dimensional integration was facilitated by the transformation $V = c_1/(c_1 + \zeta)$ so that $V$ is distributed on $(0,1)$ as

$$p(V) \propto V^{(m-3)/2}(1 + VW/c_1)^{-k/2}$$

(3.14)

The use of IMSL subroutine DCADE in this numerical integration turned out to be straightforward.

It is of interest to notice that, although we have not assumed a parametric linear model, the estimated normal growth curve of ramus heights has turned out to be almost linear, somewhat concave to be more precise, over the range where observations are available. Hence it seems that the model assumed by Fearn (1975) is a reasonable one. However because of inherent technical difficulties of his approach, the construction of exact credibility intervals proved to be exceedingly difficult. Moreover our approach is capable of handling even nonlinear growth curve situations such as those treated by Berkey (1982) without resorting to polynomial models or intrinsically nonlinear models.

ACKNOWLEDGEMENTS

This work was partially supported by the Environmental Protection Agency through a co-operative research agreement with SIAM’s Institute for Mathematics and Society (SIMS).
REFERENCES


APPENDIX. THE JEFFREY'S PRIOR

The Jeffrey's prior density function for the model in equation (2.3) will be derived in this appendix. Initially $\beta_o$ will be fixed but the result will be found to be independent of $\beta_o$, i.e. unconditionally true as well. The density is given by

$$ p(\delta^2, \Lambda) = |I|^{\frac{1}{2}} $$

(A.1)

where $I = (I_{ij})_{i,j=0,\ldots,q}$ is the Fisher information matrix (with $\beta_o$ fixed; cf. Box and Taio 1973). The density in (A.1) is found here $C \to \infty$.

Assume $\Lambda = \text{diag}(\lambda_0^2, \ldots, \lambda_q^2)$. As pointed out in Section 2 this assumption would seem to be completely valid when $q = 0$ or $1$. It may well be approximately true when $q \geq 2$. In any case, the analysis of this appendix is, for the most part, the same whatever be $q$ provided $\Lambda$ is diagonal and so will be presented for the general case.

To simplify our analysis we adopt the following notation:

$$ Z = Z_i = S_i(t_{ij}) - A_i Y_i $$
$$ L = \delta^2 D + BCB^T $$
$$ G = L + A\Lambda A^T $$
$$ V = D^{-1} - D^{-1}B(B^T D^{-1}B)^{-1}B^T D^{-1} $$
$$ F = \delta^{-2}A^T VA + \Lambda^{-1} $$
$$ F^{-1} = (F_{ij})_{i,j=0,\ldots,q} $$

With this notation Model (2.3) may be expressed as $Z \sim N(0, G)$. Since the $\{S_i\}$ are assumed to be independent the logarithm of the likelihood, $L$, of the data is given by

$$ -2\ln L = \sum \ln |G| + \sum Z_i^T G^{-1} Z $$

(A.1)

where for simplicity the index of summation, $i = 1, \ldots, m$, has been suppressed in (A.1).
Observe that under a change, $\Delta G$, in $G$, 
\[ \Delta G^{-1} = G^{-1}(\Delta G)G^{-1} \]
and
\[ \Delta \lvert G \rvert = \lvert G \rvert \; \text{tr} \; G^{-1}\Delta G. \]

Thus
\[ \Delta(-2\ln L) = \Sigma \text{tr} G^{-1}(\Delta G) - \Sigma Z^T G^{-1}(\Delta G)G^{-1} Z, \quad (A.2) \]
\[ \delta(-2\ln L)/\delta^2 = \Sigma \text{tr} G^{-1}D - \Sigma Z^T G^{-1}(D)G^{-1} Z \quad (A.3) \]
and
\[ \delta(-2\ln L)/\delta \lambda_k^2 = \Sigma \text{tr} G^{-1}A \lambda_k A^T - \Sigma Z^T G^{-1}A \lambda_k A^T G^{-1} Z \quad (A.4) \]
since, in obtaining (A.4), the change $\Lambda \rightarrow \Lambda + h \Delta_k$.

Since $E(ZZ^T) = G$, equations (A.3) and (A.4) readily imply
\[ E\delta^2(-2\ln L)/\delta(\delta^2)^2 = \Sigma \text{tr} G^{-1}DG^{-1} = 2I_{00} \quad (A.5) \]
\[ E\delta^2(-2\ln L)/\delta \lambda_k^2 \delta \lambda_l^2 = \Sigma \text{tr} G^{-1}A \lambda_k A^T G^{-1}D = 2I_{0,k+1} \quad (A.6) \]
and
\[ E\delta^2(-2\ln L)/\delta \lambda_k^2 \delta \lambda_l^2 = \Sigma \text{tr} G^{-1}A \lambda_k A^T G^{-1}A \lambda_l A^T = 2I_{k+1,l+1} \quad (A.7) \]
for $k, l = 0, \ldots, q$.

On letting $G \rightarrow \infty$ in (A.5) – (A.7) a simpler approximation to $I$ is obtained. In the ensuing analysis the following familiar matrix identities are used several times:
\[ (a + bcb^T)^{-1} = a^{-1} - a^{-1}b(b^Ta^{-1}b + c^{-1})^{-1}b^Ta^{-1} \]
and
\[ a(a + b)^{-1} = I - b(a + b)^{-1}. \]

Since $G = L + A \Delta A^T$ with $L = \delta^2D + BCB^T$,
\[ G^{-1} = L^{-1} - L^{-1} A(L^{-1}A + \Lambda^{-1})^{-1}A^T L^{-1} \quad (A.8) \]
But
\[ L^{-1} = \delta^{2}D^{-1} - \delta^{-4}D^{-1}B(C^{-1} + \delta^{-2}B_{D}^{-1}B)^{-1}B_{D}^{-1}. \]

Since \( C \) is large, we let \( C \to \infty \) and obtain in the limit
\[
L^{-1} = \delta^{2}[D^{-1} - D^{-1}B(B_{D}^{-1}B)^{-1}B_{D}^{-1}] = \delta^{2}V
\]
(A.9)

So using notation introduced at the beginning of this appendix, equation (A.9) implies
\[
A_{L}^{-1} A = \delta^{-2}A_{VA}^{T}. \quad (A.10)
\]

Observe that
\[
L^{-1}D = \delta^{-2}(I_{n} - D^{-1}B(B_{D}^{-1}B)^{-1}B^{T})
\]
so
\[
L^{-1}DL^{-1} = \delta^{2}L^{-1}
\]
(A.11)
and
\[
(L^{-1}D)^{2} = \delta^{2}L^{-1}D. \quad (A.12)
\]

From equations (A.8) and (A.10) it follows that
\[
G^{-1}D = L^{-1}D - L^{-1}AF^{-1}A_{L}^{T}L^{-1}D
\]
and
\[
(G^{-1}D)^{2} = \delta^{2}L^{-1}D - 2\delta^{2}L^{-1}AF^{-1}A_{L}^{T}L^{-1}D
\]
\[+ L^{-1}AF^{-1}A_{L}^{T}DL^{-1}AF^{-1}A_{L}^{T}L^{-1}D. \quad (A.13)\]

On applying successively (A.11) and then (A.10), the last term on the right hand side of equation (A.13) becomes
\[
\delta^{-4}L^{-1}AF^{-1}(A_{VA}^{T}F^{-1}A_{L}^{T}L^{-1}D. \quad (A.14)
\]

For convenience let \( U = \delta^{-2}(A_{VA}^{T}). \)

To compute \( I_{00} \), the trace of the quantity in equation (A.13) is required.

Since
\[
\text{tr}L^{-1}D = \delta^{-2}[\text{tr}I_{n} - \text{tr}I_{p-q}'],
\]
this trace is readily found to be
\[ \text{tr}(G^{-1}D)^2 = \delta^{-4} \{ n - p + q - 2\text{tr } U F^{-1} + \text{tr}(U F^{-1})^2 \} \]

This last result entails (see equation (A.5)),

\[ 2I_{0,0} = \delta^{-4} \{ n - p - 1 + \text{tr} \Lambda^{-2} F^{-2} \}. \]

(A.15)

To compute \( I_{0,k+1} \) from equation (A.6) observe that (A.8), (A.10) and (A.11) imply that

\[ G^{-1}D G^{-1} = \delta^{-2} L^{-1} - 2\delta^{-2} L^{-1} A F^{-1} A^T L^{-1} \]

\[ + \delta^{-2} L^{-1} A F^{-1} U F^{-1} A^T L^{-1}. \]

Thus

\[ A^T G^{-1} D G^{-1} A = \delta^{-2} (U - 2 U F^{-1} U + U F^{-1} U F^{-1} U) \]

\[ = \delta^{-2} U (I - F^{-1} U)^2 \]

\[ = \delta^{-2} (U F^{-1}) F^{-1} \Lambda^{-2} \]

\[ = \delta^{-2} (I - \Lambda^{-1} F^{-1}) F^{-1} \Lambda^{-2} \]

Thus

\[ \text{tr} \Delta_k A^T G^{-1} D G^{-1} A = \delta^{-2} (\text{tr} \Lambda^{-2} \Delta_k F^{-1} - \text{tr} \Lambda^{-2} \Delta_k F^{-2}) \]

\[ = \delta^{-2} (\lambda^{-4}_k \text{tr} \Delta_k F^{-1} - \lambda^{-6}_k \text{tr} \Delta_k F^{-2}) \]

Finally

\[ 2I_{0,k+1} = \delta^{-2} \lambda^{-4}_k \sum (F^{kk})^2 - \lambda^{-2}_k \sum \left( \sum (F^{kj})^2 \right) \]

(A.16)

The remaining equation, (A.7) for \( I_{k+1,k+1} \) is now readily simplified since

\[ A^T G^{-1} A = U F^{-1} \Lambda^{-1} = \Lambda^{-1} (I - F^{-1} \Lambda^{-1}). \]

Thus if \( \delta_{kl} \) denotes Kronecker's delta, i.e. 1 or 0 according as \( k=l \) or \( k \neq l \),

\[ 2I_{k+1,k+1} = \sum \lambda^{-4}_k \lambda^{-4}_l \{ (F^{kl})^2 + \delta_{kl} (\lambda^{-4}_k - 2\lambda^{-2}_k F^{kk}) \} \]

(A.17)

The noninformative prior density for the unspecified covariance parameters can now be computed in principle from equation (A.1). In all but the simplest cases, however, this density will not have a simple explicit
form. An inspection of equations (A.15) - (A.16) reveals that, if \( \zeta^2 = \zeta^{-2}\Lambda = \text{diag} \{ \zeta_o^2, \ldots, \zeta_q^2 \} \), the Jeffrey's prior in equation (A.1) may alternatively be represented (approximately when \( C \to \infty \)) as

\[
p(\delta^2, \Lambda) \, d\delta^2 \, d\Lambda^2 \cdots d\Lambda^q = \delta^{-2} \, g(\zeta^2) \, d\delta^2 \, d\zeta_o^2 \cdots d\zeta_q^2
\]

where \( g(\zeta^2) = p(1, \zeta) \).

The case of particular interest for the application in Section 3, is that in which \( q = 0 \) so that \( \Lambda = \lambda_o^{-2} \) is a scalar along with \( F = \delta^{-2}A^TVA + \lambda_o^{-2} \). Then equations (A.15) - (A.17) simplify considerably and yield

\[
2I_{oo} = \delta^{-4} \left[ N - m(p-1) \right] + \sum (\lambda_o^2 A_{i1}^T V_i A_i + \delta^2)^{-2}
\]  

(A.19)

\[
2I_{o1} = \sum (A_{i1}^T V_i A_i) \left( \lambda_o^2 A_{i1}^T V_i A_i + \delta^2 \right)^{-2}
\]  

(A.20)

\[
2I_{11} = \sum (A_{i1}^T V_i A_i)^2 \left( \lambda_o^2 A_{i1}^T V_i A_i + \delta^2 \right)^{-2}
\]  

(A.21)

where \( N = \sum n_i \) is the total number of observations.

In this case, after discarding irrelevant multiplicative constant, equation (A.1) becomes, with \( f = f_i = \lambda_o^2 A_{i1}^T V_i A_i + \delta^2 \),

\[
p(\delta^2, \lambda_o^2) = \left( \delta^{-4} \left[ N - m(p-1) \right] + \sum f_i^{-2} \right) \left[ \sum (A^TVA)^2 f_i^{-2} \right]^{-W}
\]

This expression simplifies somewhat to give

\[
p(\delta^2, \lambda_o^2) = \delta^{-4} \left[ N - m(p-1) \right] \left[ \sum (A_{i1}^T V_i A_i)^2 f_i^{-2} \right]^{-W} + \sum \sum (A_{i1}^T V_i A_i - A_{j1}^T V_j A_j)^2 f_i^{-2} f_j^{-2} \}
\]

(A.22)

In particular when the observation times, \( t_{ij} \), are identical in all records so that the \( n_i, A_i, B_i, D_i \) are identical, equation (A.22) gives

\[
p(\delta^2, \lambda_o^2) = \delta^{-2} f_i^{-1}
\]

(A.22)

This latter density is used in the application of Section 3.
Figure 3.1. Normal growth curve of ramus heights and confidence band.