GROUP SEQUENTIAL TESTING OF LONGITUDINAL DATA

David M. Reboussin
K. K. Gordon Lan
David L. DeMets

MADISON, WISCONSIN
Group Sequential Testing of Longitudinal Data

by

David M. Reboussin,

K. K. Gordon Lan

and

David L. DeMets*

August 10, 1992

* David M. Reboussin is an Assistant Professor in the Department of Public Health Sciences, Bowman Gray School of Medicine, Winston-Salem NC, 27157. K. K. Gordon Lan is a Professor in the Department of Statistics / Computer and Information Systems, The George Washington University, Rockville, MD 20852. David L. DeMets is Professor in the Departments of Statistics and Biostatistics, University of Wisconsin, Madison, WI 53706. David Reboussin and David DeMets’ work was supported by NIH grants T32-CA-09-565 and R01-CA-18-32. Gordon Lan’s work was supported by a grant from the National Cancer Institute (NCI1-R01-CA-55-098). The authors wish to thank Jeff Albert for his helpful comments.
Abstract

Recently, the Lan–DeMets procedure for group sequential testing (Lan & DeMets, 1983) has been extended to clinical trials collecting longitudinal data (Lee & DeMets, 1991; Wu & Lan, 1992). The extension depends on identifying the covariance structure of the estimated treatment differences computed at each interim analysis. Lee & DeMets (1991) proposed a group sequential procedure for comparing rates of change under a linear mixed effects model described by Laird & Ware (1982). Application of the procedure they suggest appears to require the numerical integration of general multivariate normal densities. However, the covariance in this case has a very simple structure which is not immediately apparent, and which allows the computation of appropriate group sequential bounds to be done using existing software (Reboussin, DeMets, Lan & Kim, 1992). We provide the algebraic details for this simplification and discuss the group sequential test.

KEY WORDS: Clinical Trials; Lan–DeMets Procedure; Laird–Ware Model; Repeated Measures Data.

Group sequential testing in clinical trials is commonplace, and as its popularity spreads the demand for application to a wider variety of experimental data increases. In particular, many clinical trials involve longitudinal data; that is, a series of measurements taken over time on each patient in the
study. This has been the subject of several recent articles (Lee & DeMets, 1991; Wu & Lan, 1992; Lan, Reboussin & DeMets, 1992). Lee & DeMets (1991) proposed a group sequential procedure for comparing rates of change under a linear mixed effects model described by Laird & Ware (1982). Application of the procedure appears to require numerical integration of a general multivariate normal density, but in fact the covariance matrix presented in that paper can be greatly simplified. If information is defined as the inverse of the variance of these estimates, application of the Lan–DeMets procedure (Lan & DeMets, 1983) for computing group sequential boundaries is no more difficult than for the single independent normal response case.

We will provide the practical and algebraic details for this claim. In Section 1 we review the Laird and Ware model, considering first a simple case in order to fix ideas. For the model most useful in clinical trials, the variance of the estimated treatment difference has a useful structure in this simple case, and an analogous structure can be found in the more general mixed effects model. Section 2 shows how the special structure of the covariance can be used to simplify sequential testing.

1 MODELING LONGITUDINAL DATA

In the simplest case, the response is modeled as a constant, plus a slope
over time, plus a random error. For patient \( i \) at time \( t \), we have \( y_i = \alpha_i + \beta_i t + e \) where \( e \sim N(0, \sigma_e^2) \) and \( \beta_i \sim N(B, \sigma_B^2) \), all independent. This model can also be written \( y_i = \alpha_i + B t + \beta_i^* t + e \) where \( B \) is fixed and \( \beta_i^* \sim N(0, \sigma_B^2) \). Given \( (\alpha_i, \beta_i) \) and \( n \geq 2 \), the model for a series of measurements \( y_{i1}, y_{i2}, \ldots, y_{im} \) on patient \( i \) is \( y_{ij} = \alpha_i + \beta_i t_{ij} + e_{ij} \). The number of measurements \( n \) may depend on \( i \), but we do not reflect this in the notation. For this series, the least squares estimate of slope is \( \hat{\beta}_i = \frac{\sum_{j=1}^{n}(t_{ij} - \bar{t})(y_{ij} - \bar{y})}{\sum_{j=1}^{n}(t_{ij} - \bar{t})^2} \). Conditional on \( \beta_i \), we have \( \text{Var} \hat{\beta}_i|\beta_i = \sigma_e^2 / \sum_{j=1}^{n}(t_{ij} - \bar{t})^2 = \sigma_w^2. \) As an estimate of \( B \), we do not condition on \( \beta_i \), so

\[
\text{Var} \hat{\beta}_i = \sigma_B^2 + \frac{\sigma_e^2}{\sum_{j=1}^{n}(t_{ij} - \bar{t})^2} = \sigma_B^2 \left[ 1 + \frac{R}{\sum_{j=1}^{n}(t_{ij} - \bar{t})^2} \right]
\]

where \( R = \sigma_e^2 / \sigma_B^2 \). We will treat \( \sigma_e^2 \) and \( \sigma_B^2 \) as known throughout.

For two groups of patients, we are generally interested in estimating the mean slope in each group. We will assume that patients are independent with respect to the measurements and underlying parameters \( \beta_i \). For \( g = 1, 2 \) and group sizes \( m_1 \) and \( m_2 \), the summary statistic for group \( g \) is a weighted sum of the individual slope estimates, where the weights are the inverse of the variance. Thus

\[
\bar{B}_g = \bar{\beta}_g = \frac{\sum_{i=1}^{m_g} \left( \text{Var} \hat{\beta}_i \right)^{-1} \hat{\beta}_i}{\sum_{i=1}^{m_g} \left( \text{Var} \hat{\beta}_i \right)^{-1}} \tag{1}
\]
\[
= \left[ \sum_{i=1}^{n} \left( 1 + \frac{R}{\sum_{j=1}^{n} (t_{ij} - \bar{t}_i)^2} \right)^{-1} \right]^{-1} \sum_{i=1}^{n} \left( 1 + \frac{R}{\sum_{j=1}^{n} (t_{ij} - \bar{t}_i)^2} \right)^{-1} \beta_i
\]

Models more general than the single slope case would be useful if a straight line regression for each patient is less appropriate than, say, a polynomial model or a model including covariates. For the \(i^{th}\) patient let \(y_i\) now denote the \(n\)-vector of responses. Then \(y_i = X_i \beta_i + \epsilon_i\) where we take \(\epsilon_i\) to be an \(n\)-vector, \(X_i\) to be a \(n \times p\) matrix, and \(\beta_i\) to be a \(p\)-vector. The covariate matrix \(X_i\) is assumed to include an intercept. We also assume \(\beta_i \sim N_p(B, \sigma_\beta^2 I_p)\) and \(\epsilon_i \sim N_n(0, \sigma_\epsilon^2 I_n)\) independently, where \(I_d\) is the \(d \times d\) identity matrix. Then

\[
\text{Var} \ y_i = V_i = \sigma_\beta^2 X_i'X_i + \sigma_\epsilon^2 I_n
\]  

(2)

This is a special case of the Laird and Ware model.

To estimate \(\beta\) using a single patient we use \(\hat{\beta}_i = (X'_iX_i)^{-1}X'_iy_i\). We will proceed as if all needed inverses exist, but similar arguments involving generalized inverses could be made if this assumption were dropped. Conditional on \(\beta_i\), \(\text{Var} \ \hat{\beta}_i | \beta_i = \sigma_\epsilon^2 (X'_iX_i)^{-1}\). Using \(\hat{\beta}_i\) as an estimate of \(B\), the unconditional variance is

\[
\text{Var} \ \hat{\beta}_i = \sigma_\beta^2 I_p + \sigma_\epsilon^2 (X'_iX_i)^{-1} = \sigma_\beta^2 \left[ I_p + R(X'_iX_i)^{-1} \right]
\]
\[ (X_i'V_i^{-1}X_i)^{-1}. \] (3)

Details for the last equality are in Appendix A. The estimate for \( B \) in group \( g \), denoted \( \hat{B}_g \), is a weighted sum of the \( \hat{\beta}_i \) for that group:

\[
\hat{B}_g = \left[ \sum_{i=1}^{m_g} \left( \text{Var } \hat{\beta}_i \right)^{-1} \right]^{-1} \sum_{i=1}^{m_g} \left( \text{Var } \hat{\beta}_i \right)^{-1} \hat{\beta}_i
\]

\[
= \left[ \sum_{i=1}^{m_g} X_i'V_i^{-1}X_i \right]^{-1} \sum_{i=1}^{m_g} \left( X_i'V_i^{-1}X_i \right) \left[ (X_i'X_i)^{-1} X_i'y_i \right]
\]

\[
= \left[ \sum_{i=1}^{m_g} X_i'V_i^{-1}X_i \right]^{-1} \sum_{i=1}^{m_g} X_iV_i^{-1} \left[ X_i (X_i'X_i)^{-1} X_i'y_i \right] y_i
\]

\[
= \left[ \sum_{i=1}^{m_g} X_i'V_i^{-1}X_i \right]^{-1} \sum_{i=1}^{m_g} X_iV_i^{-1} y_i
\] (4)

where the last equality follows because \( X_iV_i^{-1} \) is in the range space of \( X_i \).

The estimate in equation (4) is the usual mixed model estimate for a fixed effect (Laird & Ware, 1982; Lee & DeMets, 1991).

Finally, we let \( U \) denote the variance of \( \hat{B}_g \):

\[
\text{Var } \hat{B}_g = \left[ \sum_{i=1}^{m_g} X_i'V_i^{-1}X_i \right]^{-1} \left[ \sum_{i=1}^{m_g} X_i'V_i^{-1} (\text{Var } y_i) V_i^{-1}X_i' \right] \left[ \sum_{i=1}^{m_g} X_i'V_i^{-1}X_i \right]^{-1}
\]

\[
= \left[ \sum_{i=1}^{m_g} X_i'V_i^{-1}X_i \right]^{-1}
\]

\[
= U
\]
2 SEQUENTIAL ESTIMATES AND INFORMATION

In a clinical trial, monitoring proceeds by computing estimates and their variances at each interim analysis. To employ sequential tests, the covariance of the estimates at different interim analyses is needed. We denote the estimate for \( B \) computed at the \( k^{th} \) interim analysis as \( \hat{B}^{(k)} \) and its variance as \( U^{(k)} \). Then the distribution of the estimates at interim analyses 1, \ldots, \( K \) is multivariate normal with covariance

\[
\begin{bmatrix}
U^{(1)} & U^{(2)} & U^{(3)} & \cdots & U^{(K)} \\
U^{(2)} & U^{(2)} & U^{(3)} & \cdots & U^{(K)} \\
U^{(3)} & U^{(3)} & U^{(3)} & \cdots & U^{(K)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
U^{(K)} & U^{(K)} & U^{(K)} & \cdots & U^{(K)}
\end{bmatrix}.
\]  

Details for this result are in Appendix B.

We have followed Lee & DeMets (1991) in developing the model for a \( p \)-dimensional design matrix, but the decision to stop a trial early is typically based on the estimate of a single parameter: the treatment difference. Let us suppose that the treatment difference is defined either as some fixed linear combination of several parameters or as a single parameter adjusted for several factors. For example, suppose the effect of wearing filter masks
during times of high air pollution is to be tested in two groups of children over several years. The response is pulmonary function measured on a continuous scale. Body size has a strong influence on pulmonary function, so that the model should include it as a covariate. Sequential testing focuses on the estimated effect of the treatment adjusted for this covariate.

$\bar{B}_g$ and $U$ are thus one dimensional. The estimated difference between the two groups is $\Delta = \bar{B}_1 - \bar{B}_2$, and it has variance $U_1 + U_2$. The sequential estimate at interim analysis $k$ is denoted $\Delta^{(k)} = \bar{B}_1^{(k)} - \bar{B}_2^{(k)}$ with variance $U_1^{(k)} + U_2^{(k)}$. The information corresponding to $\Delta^{(k)}$ is $1/(U_1^{(k)} + U_2^{(k)})$. For sequential testing, we use the standardized differences $Z^{(k)} = \Delta^{(k)}/\sqrt{U_1^{(k)} + U_2^{(k)}}$.

We will show that by using the standardized differences and the information corresponding to each estimate, standard computations for sequential bounds developed for models in which all observations are independent can be applied by rote. For trials in which each patient yields a single, normally distributed response, Lan & DeMets (1983) related the distribution of test statistics in group sequential monitoring to Brownian Motion. The Brownian Motion process associated with sequential estimates from repeated measures data has been discussed by Wu & Lan (1992) and Lan & Zucker (1992). We will briefly demonstrate the main points of this argument.

Define $E^{(k)} = \Delta^{(k)}/(U_1^{(k)} + U_2^{(k)})$. Dividing $E^{(k)}$ by its variance $1/(U_1^{(k)} + U_2^{(k)})$ yields $Z^{(k)}$. We are considering the $U_g^{(k)}$ to be known, so $Z^{(k)}$ and $E^{(k)}$
are normally distributed. To confirm the Brownian Motion structure, we
must further show that for \(i < j\) the covariance of \(E^{(i)}\) and \(E^{(j)}\) is the variance
of \(E^{(i)}\), and that they have uncorrelated increments. From equation (5), we
have \(\text{Cov} \left( \Delta^{(i)}, \Delta^{(j)} \right) = \text{Cov} \left( \overline{B}_1^{(i)}, \overline{B}_1^{(j)} \right) + \text{Cov} \left( \overline{B}_2^{(i)}, \overline{B}_2^{(j)} \right) = U_1^{(j)} + U_2^{(j)}\) and
so the covariance of \(E^{(i)}\) and \(E^{(j)}\) is

\[
\text{Cov} \left( \frac{\Delta^{(i)}}{U_1^{(i)} + U_2^{(i)}}, \frac{\Delta^{(j)}}{U_1^{(j)} + U_2^{(j)}} \right) = \frac{1}{U_1^{(i)} + U_2^{(i)}} \text{Cov} \left( \Delta^{(i)}, \Delta^{(j)} \right) \frac{1}{U_1^{(j)} + U_2^{(j)}} = \frac{1}{U_1^{(i)} + U_2^{(i)}}
\]

which is the minimum of the two variances. Now take \(i < j \leq k < l\). The
covariance of the two increments \(E^{(j)} - E^{(i)}\) and \(E^{(l)} - E^{(k)}\) is

\[
\text{Cov} \left( \frac{\Delta^{(j)}}{U_1^{(j)} + U_2^{(j)}}, \frac{\Delta^{(i)}}{U_1^{(i)} + U_2^{(i)}}, \frac{\Delta^{(l)}}{U_1^{(l)} + U_2^{(l)}}, \frac{\Delta^{(k)}}{U_1^{(k)} + U_2^{(k)}} \right)
= \text{Cov} \left( \frac{\Delta^{(i)}}{U_1^{(i)} + U_2^{(i)}}, \frac{\Delta^{(l)}}{U_1^{(l)} + U_2^{(l)}} \right) - \text{Cov} \left( \frac{\Delta^{(j)}}{U_1^{(j)} + U_2^{(j)}}, \frac{\Delta^{(k)}}{U_1^{(k)} + U_2^{(k)}} \right) - \text{Cov} \left( \frac{\Delta^{(j)}}{U_1^{(j)} + U_2^{(j)}}, \frac{\Delta^{(l)}}{U_1^{(l)} + U_2^{(l)}} \right) + \text{Cov} \left( \frac{\Delta^{(i)}}{U_1^{(i)} + U_2^{(i)}}, \frac{\Delta^{(k)}}{U_1^{(k)} + U_2^{(k)}} \right)
= \frac{1}{U_1^{(i)} + U_2^{(i)}} - \frac{1}{U_1^{(j)} + U_2^{(j)}} - \frac{1}{U_1^{(l)} + U_2^{(l)}} + \frac{1}{U_1^{(i)} + U_2^{(i)}} + \frac{1}{U_1^{(j)} + U_2^{(j)}} + \frac{1}{U_1^{(l)} + U_2^{(l)}} - \frac{1}{U_1^{(k)} + U_2^{(k)}}
= 0
\]

Thus we have uncorrelated increments.
To apply this result to clinical trial data, suppose that we are interested in testing the difference in mean slopes in two groups, possibly with some covariate adjustment. Further suppose that the properly weighted mean slopes have been computed as in equation (2) and standardized. The computations for the spending function approach of Lan & DeMets (1983) are the same as expressed by Reboussin et al. (1992) in the documentation for FORTRAN programs which implement the general method. In some cases, sufficient accuracy is possible by using only the standardized estimates and the percentage of elapsed calendar time. More accuracy can be obtained by including the information for each sequential estimate. The role of calendar time as opposed to "information time" has been discussed by Lan, DeMets & Halperin (1984), Lan & DeMets (1989) and Lan et al. (1992).

The information in a single individual is \( \left[1 + R / \sum_{j=1}^{n} (t_j - \bar{t})^2 \right]^{-1} \). This may be defined at any sequential analysis to include all available measurements by suitable adjustment to the notation. Similarly, the information in each group of \( m_g \) subjects at the \( k^{th} \) interim analysis may be denoted \( i_k = \sum_{i=1}^{n_g} \left[1 + R / \sum_{j=1}^{n} (t_j - \bar{t})^2 \right]^{-1} \) if the range of the sums are suitably restricted. The FORTRAN program described by Reboussin et al. (1992) accepts as input the standardized differences and \( i_1, i_2, \ldots, i_k \), and produces as output the appropriate group sequential boundaries.
APPENDIX A: THE VARIANCE OF $\hat{\beta}_i$

Recall that $V_i = \sigma_B^2 X_i X_i' + \sigma_e^2 I_n$. We show $\sigma_B^2 I_p + \sigma_e^2 (X_i' X_i)^{-1} = (X_i' V_i^{-1} X_i)^{-1}$.

First, rewrite $V_i$ as $X_i (\sigma_B^2 I_p) X_i' + \sigma_e^2 I_n$. Then

$$V_i^{-1} = \sigma_e^{-2} I_n - \sigma_e^{-2} I_n X_i \left( X_i' X_i + \frac{\sigma_e^2}{\sigma_B^2} I_p \right)^{-1} X_i' \sigma_e^{-2} I_n$$

$$= \sigma_e^{-2} \left( I_n - X_i \left( X_i' X_i + \frac{\sigma_e^2}{\sigma_B^2} I_p \right)^{-1} X_i' \right)$$

and so

$$X_i' V_i^{-1} X_i = \sigma_e^{-2} \left[ X_i' X_i - X_i' X_i \left( X_i' X_i + \frac{\sigma_e^2}{\sigma_B^2} I_p \right)^{-1} X_i' X_i \right]$$

Let $A = X_i' X_i$. Application of a standard result for matrix inverses (Rao, 1973, page 33) gives

$$(X_i' V_i^{-1} X_i)^{-1} = \sigma_e^2 \left[ A^{-1} - A^{-1} A \left( A + \frac{\sigma_e^2}{\sigma_B^2} I_p \right)^{-1} A^{-1} A \right]$$

$$= \sigma_e^2 \left[ (X_i' X_i)^{-1} - \left( (X_i' X_i) - (X_i' X_i) - \frac{\sigma_e^2}{\sigma_B^2} I_p \right)^{-1} \right]$$

$$= \sigma_e^2 \left[ (X_i' X_i)^{-1} + \frac{\sigma_B^2}{\sigma_e^2} I_p \right]$$

$$= \sigma_e^2 (X_i' X_i)^{-1} + \sigma_B^2 I_p$$

which is the result.
APPENDIX B: THE COVARIANCE FOR SEQUENTIAL ESTIMATES

$X_i^{(k)}$ denotes the quantity $X$ for the $i^{th}$ patient at the $k^{th}$ analysis. $n_i^{(k)}$ will denote the number of observations on the $i^{th}$ patient accumulated by the $k^{th}$ interim analysis. The actual covariate values on the $i^{th}$ patient accumulated by the $k^{th}$ interim analysis are denoted by the $n_i^{(k)} \times p$ matrix $X_i^{(k)}$, where $p$ is the number of covariates measured (plus one for the intercept term). The matrix of covariate values measured on the $i^{th}$ patient at the end of the trial can be denoted simply $X_i$, but note that this is the same as $X_i^{(K)}$ if there are a total of $K$ interim analyses conducted. $X_i$ or $X_i^{(K)}$ has dimension $n_i^{(K)} \times p$, or $n \times p$, where $n$ denotes the total number of observations on the $i^{th}$ patient at the end of the trial.

Define $P_i^{(k)} = [I_{n_i^{(k)}}: 0_{n \times (n - n_i^{(k)})}]_{n \times n}$ so that $X_i^{(k)} = P_i^{(k)} X_i$. Notice that $P_i^{(k)}$ "slices off" the rows of $X_i$ not observed by the $k^{th}$ interim analysis. It can also be used in denoting the variance of the responses for the $i^{th}$ patient observed by the $k^{th}$ analysis in terms of the variances of all responses observed by the end of the trial. Let $y_i = y_i^{(K)}$ denote the responses on the $i^{th}$ patient accumulated by the end of the trial ($K^{th}$ analysis), and $\text{Var}(y_i) = V_i$. Then $y_i^{(k)} = P_i^{(k)} y_i$ and $\text{Var}(y_i^{(k)}) = V_i^{(k)} = P_i^{(k)} V_i P_i^{(k)'}$. Letting $W_i = V_i^{-1}$, we also have $V_i^{(k)-1} = W_i^{(k)} = (P_i^{(k)} V_i P_i^{(k)'} )^{-1}$. 

11
Let \( \hat{B} \) be our (vector of) estimates at the end of the trial, and \( \hat{B}^{(k)} \) be our (vector of) estimates at the \( k^{th} \) interim analysis. Then

\[
\hat{B}^{(k)} = \left[ \sum_{i=1}^{m^{(k)}} X_i^{(k)'} W_i^{(k)} X_i^{(k)} \right]^{-1} \left[ (X_1^{(k)'} W_1^{(k)} P_1^{(k)} y_1), \ldots, (X_{m^{(k)}}^{(k)'} W_{m^{(k)}}^{(k)} P_{m^{(k)}}^{(k)} y_{m^{(k)}}) \right] 1_{m^{(k)} \times 1},
\]

where \( m^{(k)} \) is the total number of patients entered in the study by the \( k^{th} \) analysis. Denote

\[
U^{(k)} = \left[ \sum_{i=1}^{m^{(k)}} X_i^{(k)'} W_i^{(k)} X_i^{(k)} \right]^{-1}
\]

and modify the definition of \( M \) given by Lee & DeMets (1991):

\[
M(k)_{p \times \sum_{i=1}^{m} n_i} = U^{(k)} \left[ (X_1^{(k)'} W_1^{(k)} P_1^{(k)}), \ldots, (X_{m^{(k)}}^{(k)'} W_{m^{(k)}}^{(k)} P_{m^{(k)}}^{(k)}), 0, \ldots, 0 \right]
\]

where \( M(k) \) is the \( k^{th} \) block of \( p \) rows in the matrix \( M \), and \( m = m^{(K)} \) is the total number of patients at the end of the trial.

Readers may find that some of the following computations are easier to follow if \( p = 1 \), but it is not necessary to do so. We can write the matrix \( \text{Var} \ (\hat{B}^{(k)}) \) as \( M(k) \text{ diag}(V_1, V_2, \ldots, V_m) M(k)' \) and the matrix \( \text{Var} \ (\hat{B}^{(1)}, \hat{B}^{(2)}, \ldots, \hat{B}^{(K)}) \) as
\[
\begin{pmatrix}
M(1) \\
M(2) \\
\vdots \\
M(K)
\end{pmatrix} [\text{diag}(V_1, V_2, \ldots, V_n)] [M(1)', M(2)', \ldots, M(K)'].
\] (7)

This is the covariance matrix of the sequentially computed mixed model estimates, as in Lee & DeMets (1991). We will now simplify it.

As a notational convenience, let us define

\[
\begin{align*}
A_i^{(j)(k)} &= U^{(j)}(X_i^{(j)} W_i^{(j)} P_i^{(j)}) V_i (X_i^{(k)} W_i^{(k)} P_i^{(k)}) U^{(k)} \\
&= U^{(j)}(X_i^{(j)} W_i^{(j)} (P_i^{(j)} V_i P_i^{(k)}) W_i^{(k)} X_i^{(k)}) U^{(k)}.
\end{align*}
\] (8)

Since patients are independent, equation (7) can be written

\[
\begin{bmatrix}
\sum_{i=1}^{m^{(1)}} A_i^{(1)(1)} & \sum_{i=1}^{m^{(1)}} A_i^{(1)(2)} & \cdots & \sum_{i=1}^{m^{(1)}} A_i^{(1)(K)} \\
\sum_{i=1}^{m^{(2)}} A_i^{(2)(1)} & \sum_{i=1}^{m^{(2)}} A_i^{(2)(2)} & \cdots & \sum_{i=1}^{m^{(2)}} A_i^{(2)(K)} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{i=1}^{m^{(K)}} A_i^{(K)(1)} & \sum_{i=1}^{m^{(K)}} A_i^{(K)(2)} & \cdots & \sum_{i=1}^{m^{(K)}} A_i^{(K)(K)}
\end{bmatrix}.
\]

What follows will establish that \(\sum_{i=1}^{m^{(j)}} A_i^{(j)(k)} = \sum_{i=1}^{m^{(k)}} A_i^{(k)(k)} = U^{(k)}\) where \(j \leq k\).

We now simplify the terms in equation (8). Since \(P_i^{(k)} V_i P_i^{(k)} = V_i^{(k)}\), \(A_i^{(k)(k)}\) becomes

\[
U^{(k)} X_i^{(k)} (W_i^{(k)} V_i^{(k)} W_i^{(k)}) X_i^{(k)} U^{(k)} = U^{(k)} X_i^{(k)} W_i^{(k)} X_i^{(k)} U^{(k)}
\]
so that, recalling the definition of $U^{(k)}$ in equation (6),

$$\sum_{i=1}^{m^{(k)}} A^{(k)(k)}_i = U^{(k)} \left[ \sum_{i=1}^{m^{(k)}} X^{(k)^{\top}}_i W^{(k)}_i X^{(k)}_i \right] U^{(k)} = U^{(k)}.$$  

Now consider $A^{(j)(k)}_i$. We have

$$U^{(j)} X^{(j)^{\top}}_i \left[ W^{(j)}_i P^{(j)}_i V_i P^{(k)^{\top}}_i W^{(k)}_i \right] X^{(k)^{\top}}_i U^{(k)}$$

$$= U^{(j)} X^{(j)^{\top}}_i \left[ W^{(j)}_i \theta_{n_i^{(j)}} X^{(k)^{\top}}_i \right] X^{(k)^{\top}}_i U^{(k)}$$

$$= U^{(j)} \left[ X^{(j)^{\top}}_i W^{(j)}_i X^{(j)}_i \right] U^{(k)},$$

so that

$$\sum_{i=1}^{m^{(j)}} A^{(j)(k)}_i = U^{(j)} \left[ \sum_{i=1}^{m^{(j)}} X^{(j)^{\top}}_i W^{(j)}_i X^{(j)}_i \right] U^{(k)} = U^{(k)}.$$  

The crucial step from (9) to (10) can be seen by rewriting

$$\left[ W^{(j)}_i P^{(j)}_i V_i P^{(k)^{\top}}_i W^{(k)}_i \right] = \left( P^{(j)}_i V_i P^{(j)^{\top}}_i \right)^{-1} \left( P^{(j)}_i V_i P^{(k)^{\top}}_i \right) \left( P^{(k)}_i V_i P^{(k)^{\top}}_i \right)^{-1}. \quad (11)$$

Define $\left( P^{(j)}_i V_i P^{(j)^{\top}}_i \right) = Q$. Using the symmetry of $V_i$, the definitions of $P^{(j)}_i$ and $P^{(k)^{\top}}_i$, and defining $R$ and $S$ as needed, we have

$$\left( P^{(j)}_i V_i P^{(j)^{\top}}_i \right) = [Q : R], \quad \left( P^{(k)}_i V_i P^{(k)^{\top}}_i \right) = \begin{bmatrix} Q & R \\ R' & S \end{bmatrix}.$$  

Rewriting (11) in these terms and applying a standard formula for inverting
partitioned matrices (Rao, 1973, page 33),

\[ Q^{-1}[Q : R] \begin{bmatrix} Q & R \\ R' & S \end{bmatrix}^{-1} = [I_{n_i^{(p)}} : Q^{-1}R] \begin{bmatrix} Q & R \\ R' & S \end{bmatrix}^{-1} = [Q^{-1} : 0_{n_i^{(p)} \times (n_i^{(p)} - n_i^{(p)})}] \]

Thus the variance-covariance matrix for the sequential estimates of the fixed effects coefficients may be written as in equation (5).

REFERENCES


