Statistica Sinica **5**(1995), 97-107

EM FOR LONGITUDINAL DATA FROM MULTIPLE STUDIES

Gilbert W. Fellingham and Morgan D. Wise

Brigham Young University

Abstract. The two-stage random-effects model (Harville (1977), Laird and Ware (1982)) offers a powerful and flexible tool for the analysis of longitudinal data. This method assumes individual response may be modeled as the sum of an overall population effect, a random individual deviation, and random error. When individuals are nested within families or companies, this source of variability should also be considered. Building a model for hearing loss for minimally noise-exposed workers with data compiled from multiple sources motivated extending the two-stage model to include a nested random effect. Computational methods to compute population effects, random effects, and variance components in this more general setting using the EM algorithm are given.

Key words and phrases: Random effects models, EM algorithm, restricted maximum likelihood, mixed models, hierarchical models.

1. Introduction

Longitudinal studies are typically designed to investigate changes over time in a characteristic that is measured repeatedly for each study participant. When all data are present and measurements are made at the same time point for each of many individuals, standard multivariate methods may be used to analyze such data. When measurements are missing, when there is considerable variation in the number or timing of observations, or when samples are small, estimation using standard methods is impossible.

Harville (1977), Dempster, Rubin and Tsutakawa (1981) and Laird and Ware (1982), among others, have proposed and investigated the use of two-stage random effects models in this setting. These models assume a patterned covariance structure (block diagonal), which allows for estimation of variance components. The EM algorithm may be invoked to provide maximum likelihood (or restricted maximum likelihood) estimates of the variance components. These estimated variance components are then typically considered known to provide empirical Bayes estimates of fixed and random effects.

This method, however, does not incorporate nested random effects as when individuals are nested within families or companies. When combining data from multiple studies, random variability between data sources is important. For example, suppose individuals being monitored for hearing loss are nested within different companies, then estimation of both company-to-company and individual-to-individual variability is necessary. We will focus on an example where data come from multiple companies to extend current methodology to allow for estimation of company-to-company variability as well as subject-to-subject variability. With this more general model, appropriate estimates of fixed effects and their variances will be possible.

2. The Problem

The Occupational Safety and Health Administration (OSHA) mandates audiometric testing of employees exposed to time-weighted average (TWA) industrial noise levels greater than 85 Db. Various methods have been proposed to analyze such data (Melnick (1984), Royster and Royster (1986)), but as yet no model which incorporates both longitudinal and cross-sectional data exists.

The problem for our analysis was to describe hearing loss in a population with minimal noise exposure. Although companies are legally liable for hearing loss in employees from industrial noise, there is no well-accepted model for hearing loss where there is no noise exposure. Thus, establishing that industrial noise is the cause of specific hearing loss is difficult. With a baseline model established, companies could compare their employee experience with what would be expected with minimum noise exposure. If employee hearing loss exceeded that expected in a control setting, the need for remedial measures would be established. Such a baseline model could also be used by regulatory agencies to monitor company performance with regard to employee hearing loss.

The National Technical Information Service (NTIS) of the U.S. Department of Commerce has a data tape available which was prepared for the National Institute for Occupational Safety and Health (NIOSH) by Environmental Noise Consultants, Inc. (1987). This tape contains audiometric data on over 45,000 individuals from 22 companies in the U.S. and Canada. Over 15,000 of the individuals had at least 4 audiograms and roughly 4,000 individuals had at least 8 audiograms.

Data were prepared for this analysis by eliminating all individuals for whom birth date, race, gender or noise exposure was not known. Also, since this analysis was to focus on subjects with minimal noise exposure, once an individual's TWA noise exposure was 75 Db or greater, any subsequent measurements on that individual were eliminated. This left 3,562 individuals from 6 companies with 8,768 observations for analysis. The dependent variable was left and right ear average hearing level at 4000 Hz. (An increased hearing level indicates hearing deterioration.) Explanatory variables considered in the analysis were gender, race (white, black, Hispanic), age at first test, and time from first test.

In an environment where noise exposure is limited, cross-sectional hearing loss (i.e. loss across many individuals associated with age) should be the same as longitudinal hearing loss (i.e. loss on a single individual associated with time). Thus, it is necessary to have both cross-sectional and longitudinal data included in the analysis. In our formulation, cross-sectional loss may be estimated using the covariate "age at first test" while longitudinal loss is associated with "time from first test".

3. The Model

In the standard formulation, the response is modeled as the sum of fixed effects (unknown population parameters) and random effects (individual deviations). In our more general formulation, individual effects are nested within company effects which are assumed to be random.

In general, let $\boldsymbol{y}_{ij}$ denote the responses for the jth individual in the ith data set. Let α denote a $p_\alpha \times 1$ vector of unknown population parameters, and $\boldsymbol{X}_{ij}$ be a known $n_{ij} \times p_\alpha$ design matrix linking α to the responses $\boldsymbol{y}_{ij}$. Let $\boldsymbol{c}_i$ denote a $p_c \times 1$ vector of unknown company random effects and $\boldsymbol{W}_{ij}$ be a known $n_{ij} \times p_c$ design matrix linking $\boldsymbol{c}_i$ to $\boldsymbol{y}_{ij}$. Finally, let $\boldsymbol{b}_{ij}$ denote a $p_b \times 1$ vector of unknown individual random effects for individuals nested within companies, and $\boldsymbol{Z}_{ij}$ be a known $n_{ij} \times p_b$ design matrix linking $\boldsymbol{b}_{ij}$ to $\boldsymbol{y}_{ij}$. We propose the following model for multivariate normal data. For each individual unit, ij,

$$\boldsymbol{y}_{ij} = \boldsymbol{X}_{ij}\alpha + \boldsymbol{W}_{ij}\boldsymbol{c}_i + \boldsymbol{Z}_{ij}\boldsymbol{b}_{ij} + \boldsymbol{e}_{ij}, \tag{1}$$

where $\boldsymbol{e}_{ij} \sim \mathrm{MVN}(0, \sigma^2 \boldsymbol{I}_{n_{ij}}), \boldsymbol{c}_i \sim \mathrm{MVN}(0, \boldsymbol{E})$ and $\boldsymbol{b}_{ij} \sim \mathrm{MVN}(0, \boldsymbol{D})$, all independent of each other. $\boldsymbol{E}$ is a $p_c \times p_c$ positive-definite covariance matrix, and $\boldsymbol{D}$ is a $p_b \times p_b$ positive-definite covariance matrix. In the Bayesian formulation, α is also considered to be normally and independently distributed with $\mathrm{Var}(\alpha) = \boldsymbol{\Gamma}$.

In our problem, hearing loss was assumed to be a linear function of time (a term defining a quadratic function of time was found to be insignificant). The random effects, both company and individual, were mean and slope of hearing loss based on time from first test. Illustrative design matrices, $\boldsymbol{Z}_{ij}$ and $\boldsymbol{W}_{ij}$, for these random effects for an imaginary individual are shown below. Population or fixed effects were the overall mean, overall slope of hearing loss as a function of time from first test (the longitudinal effect of time) (time from first test was centered by subtracting 2.7), covariates for age (centered by subtracting 40) at first test (the cross-sectional effect of time), gender, race (black vs. non-black) (a second degree of freedom comparing white to Hispanic was found to be nonsignificant), and an age at first test by gender interaction. A hypothetical design matrix, $\boldsymbol{X}_{ij}$, for these effects for an imaginary individual is shown below.

These matrices describe an individual who was 42.0 years old at the time of his first test, (2nd column is age -40) who was a black male (3rd column is gender, 0 = female, 1 = male; 4th column is race, 0 = non-black, 1 = black) and who has had 5 tests over a span of 5 years. The test for the second year after the initial test was missing. (The 6th column is time since first test -2.7.) Column 5 is an age at first test by gender interaction.

$$\boldsymbol{X}_{ij} = \begin{bmatrix} 1 & 2 & 1 & 1 & 2 & -2.7 \\ 1 & 2 & 1 & 1 & 2 & -1.6 \\ 1 & 2 & 1 & 1 & 2 & 0.2 \\ 1 & 2 & 1 & 1 & 2 & 1.5 \\ 1 & 2 & 1 & 1 & 2 & 2.4 \end{bmatrix}, \qquad \boldsymbol{Z}_{ij} = \boldsymbol{W}_{ij} = \begin{bmatrix} 1 & -2.7 \\ 1 & -1.6 \\ 1 & 0.2 \\ 1 & 1.5 \\ 1 & 2.4 \end{bmatrix}.$$

In general, let k_i be the number of individuals in the ith company, m be the number of companies, $\sum_{i=1}^{m} k_i = k$, and $\sum_{i=1}^{m} \sum_{j=1}^{k_i} n_{ij} = n$, be the total number of individuals and total number of observations, respectively.

The model for all individuals within all data sets may now be written as:

$$\boldsymbol{y} = \boldsymbol{X\alpha} + \boldsymbol{Wc} + \boldsymbol{Zb} + \boldsymbol{e}, \tag{2}$$

with

$$\boldsymbol{y}' = (\boldsymbol{y}'_{11}, \boldsymbol{y}'_{12}, \ldots, \boldsymbol{y}'_{mk_m}),\ \ \boldsymbol{X}' = (\boldsymbol{X}'_{11}, \boldsymbol{X}'_{12}, \ldots, \boldsymbol{X}'_{mk_m}),\ \ \boldsymbol{e}' = (\boldsymbol{e}'_{11}, \boldsymbol{e}'_{12}, \ldots, \boldsymbol{e}'_{mk_m}),$$

$$\boldsymbol{W} = \mathrm{diag}(\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_m) \text{ where } \boldsymbol{W}_i{}' = (\boldsymbol{W}_{i1}{}', \boldsymbol{W}_{i2}{}', \ldots, \boldsymbol{W}_{ik_i}{}'),$$

$$\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{E}, \ldots, \boldsymbol{E}), \quad \boldsymbol{Z} = \mathrm{diag}(\boldsymbol{Z}_{11}, \ldots, \boldsymbol{Z}_{mk_m}), \quad \Delta = \mathrm{diag}(\boldsymbol{D}, \ldots, \boldsymbol{D}),$$

$$\boldsymbol{c}' = (\boldsymbol{c}'_1, \ldots, \boldsymbol{c}'_m) \quad \text{and} \quad \boldsymbol{b} = (\boldsymbol{b}'_{11}, \boldsymbol{b}'_{12}, \ldots, \boldsymbol{b}'_{mk_m}),$$

which extends the standard formulation by inserting the $\boldsymbol{Wc}$ term.

4. Estimation Given Hyperparameters

If $\sigma^2, \boldsymbol{D}, \boldsymbol{E}$, and $\boldsymbol{\Gamma}$ were known, Bayesian estimates for $\alpha, \boldsymbol{b}, \boldsymbol{c}$ could be obtained as their posterior expectations, given $\boldsymbol{y}, \sigma^2, \boldsymbol{D}, \boldsymbol{E}$, and $\boldsymbol{\Gamma}$. With $\sigma^2, \boldsymbol{D}, \boldsymbol{E}$, and $\boldsymbol{\Gamma}$ unknown, an empirical Bayes approach (Deely and Lindley (1981), Efron and Morris (1975), and Dempster, Laird and Rubin (1977)) replaces $\sigma^2, \boldsymbol{D}, \boldsymbol{E}$, and $\boldsymbol{\Gamma}$ with estimates of these variance components (see Section 5). These empirical Bayes estimates, also called shrinkage or Stein estimates (James and Stein (1961)), essentially "borrow strength" from the entire data set to estimate any particular effect. The formulation implies that there is information about a particular company effect contained in all the companies, so that a naive estimate of a company effect may be improved by borrowing strength from the other companies.

We let $\boldsymbol{\Gamma}^{-1} = 0$, indicating vague prior information about $\boldsymbol{\alpha}$, and use estimates of $\sigma^2, \boldsymbol{D}$ and $\boldsymbol{E}$ obtained by maximizing the limiting (as $\boldsymbol{\Gamma}^{-1} \to 0$) marginal normal likelihood of $\sigma^2, \boldsymbol{D}$ and $\boldsymbol{E}$ given $\boldsymbol{y}$ after integrating over $\boldsymbol{\alpha}, \boldsymbol{b}$ and $\boldsymbol{c}$ (see Section 5). These estimates are the restricted maximum likelihood (REML) estimates (Harville (1977) and Dempster, Rubin and Tsutakawa (1981)).

If the variance components are known, it is possible to show that the posterior distributions of $\boldsymbol{\alpha}, \boldsymbol{b}, \boldsymbol{c}$ are normal with mean and variance

$$\hat{\boldsymbol{\alpha}} = E[\boldsymbol{\alpha}|\sigma^2, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{E}] = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y} \tag{3}$$

and

$$\mathrm{Var}(\boldsymbol{\alpha}|\sigma^2, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{E}) = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}, \tag{4}$$

where $\boldsymbol{\Sigma} = \boldsymbol{W}\boldsymbol{\Lambda}\boldsymbol{W}' + \boldsymbol{Z}\boldsymbol{\Delta}\boldsymbol{Z}' + \sigma^2\boldsymbol{I}_n$. Also,

$$\hat{\boldsymbol{b}} = E[\boldsymbol{b}|\sigma^2, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{E}] = (\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{Z} + \sigma^2\boldsymbol{\Delta}^{-1})^{-1}\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{y} \tag{5}$$

and

$$\mathrm{Var}[\boldsymbol{b}|\sigma^2, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{E}] = \sigma^2(\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{Z} + \sigma^2\boldsymbol{\Delta}^{-1})^{-1}, \tag{6}$$

where

$$\boldsymbol{G} = \boldsymbol{H} - \boldsymbol{H}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{H}\boldsymbol{W})^{-}\boldsymbol{W}'\boldsymbol{H} + \boldsymbol{H}\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{H}\boldsymbol{W} + \sigma^2\boldsymbol{\Lambda}^{-1})^{-1}\sigma^2\boldsymbol{\Lambda}^{-1}(\boldsymbol{W}'\boldsymbol{H}\boldsymbol{W})^{-}\boldsymbol{W}'\boldsymbol{H}$$

and $\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, with "$-$" indicating a generalized inverse. These results reduce to those of Laird and Ware (1982) in the case where there is no company effect.

Also,

$$\hat{\boldsymbol{c}} = E[\boldsymbol{c}|\sigma^2, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{E}] = (\boldsymbol{W}'\boldsymbol{F}\boldsymbol{W} + \sigma^2\boldsymbol{\Lambda}^{-1})^{-1}\boldsymbol{W}'\boldsymbol{F}\boldsymbol{y} \tag{7}$$

and

$$\mathrm{Var}[\boldsymbol{c}|\sigma^2, \boldsymbol{y}, \boldsymbol{D}, \boldsymbol{E}] = \sigma^2(\boldsymbol{W}'\boldsymbol{F}\boldsymbol{W} + \sigma^2\boldsymbol{\Lambda}^{-1})^{-1}, \tag{8}$$

where

$$\boldsymbol{F} = \boldsymbol{H} - \boldsymbol{H}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{Z})^{-}\boldsymbol{Z}'\boldsymbol{H} + \boldsymbol{H}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{Z} + \sigma^2\boldsymbol{\Delta}^{-1})^{-1}\sigma^2\boldsymbol{\Delta}^{-1}(\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{Z})^{-}\boldsymbol{Z}'\boldsymbol{H}.$$

5. Finding MLE's of Hyperparameters Using EM

To apply the results of Section 4, it is necessary to have estimates of $\sigma^2, \boldsymbol{D}$ and $\boldsymbol{E}$. We estimate the variance components using the EM algorithm (Dempster, et al. (1977)). These estimated variance components are then used in the estimation of the other model parameters.

Application of the EM algorithm in the current setting is done by considering the complete data to consist of $\boldsymbol{y}, \boldsymbol{c}, \boldsymbol{b}$ and $\boldsymbol{e}$. If we were to observe $\boldsymbol{c}, \boldsymbol{b}$ and $\boldsymbol{e}$ as well as $\boldsymbol{y}$, we could find closed form maximum likelihood estimates of $\sigma^2, \boldsymbol{D}$ and $\boldsymbol{E}$ based on quadratic forms in $\boldsymbol{e}, \boldsymbol{b}$ and $\boldsymbol{c}$. We would use

$$\hat{\sigma}^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n} = \frac{t_1}{n}, \tag{9}$$

$$\hat{\boldsymbol{E}} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{c}_i\boldsymbol{c}_i' = \frac{\boldsymbol{t}_2}{m}, \tag{10}$$

and

$$\hat{\boldsymbol{D}} = \frac{1}{k}\sum_{i=1}^{m}\sum_{j=1}^{k_i} \boldsymbol{b}_{ij}\boldsymbol{b}_{ij}' = \frac{\boldsymbol{t}_3}{k}, \tag{11}$$

the sufficient statistics for $\sigma^2, \boldsymbol{E}$ and $\boldsymbol{D}$ being t_1 and the non-redundant components of $\boldsymbol{t}_2$ and $\boldsymbol{t}_3$.

If estimates of $\sigma^2, \boldsymbol{D}$ and $\boldsymbol{E}$ are available (call these estimates $\hat{\boldsymbol{\omega}}$ and estimates of $\boldsymbol{b}$ and $\boldsymbol{c}$ based on $\hat{\boldsymbol{\omega}}$ call $\hat{\boldsymbol{b}}(\hat{\boldsymbol{\omega}})$ and $\hat{\boldsymbol{c}}(\hat{\boldsymbol{\omega}})$), we use them to calculate estimates of the missing sufficient statistics by setting them equal to their expectations, conditional on $\boldsymbol{y}$ and $\hat{\boldsymbol{\omega}}$. In this case the estimates are

$$\hat{t}_1 = E(\boldsymbol{e}'\boldsymbol{e}|\boldsymbol{y}, \hat{\boldsymbol{\omega}}) = \hat{\boldsymbol{e}}(\hat{\boldsymbol{\omega}})'\hat{\boldsymbol{e}}(\hat{\boldsymbol{\omega}}) + \operatorname{tr}\operatorname{Var}(\boldsymbol{e}|\boldsymbol{y}, \hat{\boldsymbol{\omega}}), \tag{12}$$

$$\hat{\boldsymbol{t}}_2 = E\left[\sum_{i=1}^{m} \boldsymbol{c}_i\boldsymbol{c}_i'|\boldsymbol{y}, \hat{\boldsymbol{\omega}}\right] = \sum_{i=1}^{m} \hat{\boldsymbol{c}}_i(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{c}}_i(\hat{\boldsymbol{\omega}})' + \operatorname{Var}(\boldsymbol{c}_i|\boldsymbol{y}_i, \hat{\boldsymbol{\omega}}), \tag{13}$$

$$\hat{\boldsymbol{t}}_3 = E\left[\sum_{i=1}^{m}\sum_{j=1}^{k_i} \boldsymbol{b}_{ij}\boldsymbol{b}_{ij}'|\boldsymbol{y}, \hat{\boldsymbol{\omega}}\right] = \sum_{i=1}^{m}\sum_{j=1}^{k_i} \hat{\boldsymbol{b}}_{ij}(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{b}}_{ij}(\hat{\boldsymbol{\omega}})' + \operatorname{Var}(\boldsymbol{b}_{ij}|\boldsymbol{y}_{ij}, \hat{\boldsymbol{\omega}}). \tag{14}$$

After choosing suitable starting values for $\sigma^2, \boldsymbol{D}$ and $\boldsymbol{E}$, we iterate between (12), (13) and (14) (which define the E-step) and (9), (10) and (11) (which define the M-step) until convergence. At convergence, we obtain not only the variance components, but also estimates for $\alpha, \boldsymbol{c}, \boldsymbol{b}$, and their variances based on the estimated variance components (see Section 4). Since for even a few individuals, the formulas for estimating effects and their variances are quite unwieldy, we present equivalent computational formulas in the Appendix.

6. Analysis

It is important to reiterate that the current analysis uses data from only six companies. Thus, estimation of the variance components matrix $\boldsymbol{E}$ might be called into question. However, our experience with the asymptotics even with small samples has been positive. We do feel extreme caution should be used if any of the random effect distributions appear to be skewed (Fellingham and Raghunathan (1994)). This was not the case in this analysis.

We actually analyzed a number of models before settling on the one which we present here. Early models included a term for quadratic hearing loss and additional degrees of freedom for race, as well as interaction terms. We also checked for a possible learning effect (i.e., the possibility that subjects perform better on later tests as they become familiar with the testing procedure). The model we present is used for two main reasons. One, it is parsimonious. Two, all terms included in the model are ones which other researchers in the area have suggested should be important (see Melnick (1984), Royster and Royster (1986)). The final model includes fixed effects for the intercept, age (the cross-sectional effect), race (non-black vs. black), gender, a gender by age interaction, and time since first test (the longitudinal effect). Random effects include the intercept and time since first test. The design matrices are as shown in more detail in Section 3.

The results of the analysis are given in Table 1.

Table 1. Results of nested random effects analysis of audiometric data from six companies

Variance Components

$\hat{\sigma}^2 = 27.79 \quad \hat{\sigma} = 5.27$	(within-individual)
$\hat{E} = \begin{bmatrix} 13.65 & 0.264 \\ & 0.052 \end{bmatrix}$	(between-companies)
$\hat{D} = \begin{bmatrix} 226.9 & 1.030 \\ & 0.697 \end{bmatrix}$	(between-individuals)

Population Effects

	Estimate	Standard deviation	Z-score	p-value
Mean	14.40	1.707	8.43	< 0.001
Age at first test (-40)	0.3792	0.047	8.07	< 0.001
Gender	17.36	0.819	21.2	< 0.001
Race (black vs. non-black)	-5.833	0.694	-8.41	< 0.001
Gender by age interaction	0.6243	0.057	11.1	< 0.001
Time since first test	0.1590	0.120	1.33	< 0.184

The within subject variability, 5.27 seems reasonable since the measurement in an audiometric test is typically accurate to 5 Db. Individual intercepts have a standard deviation on the order of 15 Db (variance = 226.9). Since audiometric tests typically cover a range from 0 to 99, and 6 standard deviations cover 90 Db, this variance component seems reasonable as well. Individual slopes have a standard deviation of .835 (variance = .697), which indicates a high degree of individual to individual variability. There is very little relationship between individual slopes and intercepts (covariance = 1.03, correlation = .08).

Between company variability is much lower than between individual variability. Company to company intercepts have a standard deviation of only 3.7, and slopes have a standard deviation of only 0.23. The correlation of intercepts and slopes is much greater within companies than within individuals (.31 to .08). It may be that differences in monitoring equipment, testing procedures, or testing personnel that are consistent within a company but not between companies are associated with this difference.

Population effects indicate that females hear better than males and blacks hear better than non-blacks. Mean hearing loss with age is on the order of .4dB per year. However, the cross-sectional hearing loss in this group is more severe for males than for females. That is, as individuals age, males lose hearing faster than do females.

Of major importance in this study is the relationship of the cross-sectional hearing loss (age at first test) to the longitudinal loss (time since first test). If there was no accelerated hearing loss, as there should not be in this control group, then cross-sectional and longitudinal components should be similar. Comparing longitudinal and cross-sectional components yields a z-score of 1.71 and a p-value of .087. Of course, in this case, we would be concerned only if the longitudinal loss exceeded cross-sectional loss, which is not the case here.

It is interesting to compare this result to an analysis which ignores the company effect. Laird and Ware's model was used to perform such an analysis. The results are given in Table 2.

Table 2. Results of random effects analysis of audiometric data when data set effect is ignored

Variance Components

$\hat{\sigma}^2 = 27.66 \quad \hat{\sigma} = 5.26$

$$\hat{D} = \begin{bmatrix} 228.8 & 1.053 \\ & 0.790 \end{bmatrix}$$

Population Effects

	Estimate	Standard deviation	Z-score	p-value
Mean	15.39	0.618	24.9	< 0.001
Age at first test (−40)	0.3388	0.045	7.58	< 0.001
Gender	18.58	0.728	25.5	< 0.001
Race (black vs. non-black)	−5.902	0.696	−8.48	< 0.001
Gender by age interaction	0.6856	0.054	12.7	< 0.001
Time since first test (−2.7)	0.0944	0.041	2.29	< 0.022

Estimates for $\sigma^2, \boldsymbol{D}$, and the population effects are quite similar to those in Table 1. However, because we have failed to account for variability between companies, the population effects have been estimated with variances which are too small.

Although the differences at first glance appear to be small, it is possible to see where analysis with the less general model could lead one to different conclusions. For example, if we compare cross-sectional to longitudinal components in this model, the z-score for the difference is 4.02 with a p-value of 0.000. With this analysis we would conclude that longitudinal loss is significantly less than cross-sectional loss. That is, hearing loss prior to the implementation of a testing program significantly exceeded hearing loss after a monitoring program was installed. We believe this rather odd conclusion to be the result not of inappropriate estimates of fixed effects, but of incorrect estimates of variance components. Estimates of the variance of the fixed effects mean (intercept) and time since first test (slope) are on the order of eight times smaller for the model which ignores company-to-company variance. Thus, inappropriate model specification could well lead to inappropriate conclusions.

7. Discussion and Further Remarks

There are other approaches which might be applied in the hierarchical setting. A considerable body of literature exists on analyzing hierarchical models using maximum likelihood methods (see e.g. Bryk and Raudenbush (1992)). Goldstein (1986, 1989) has shown that iterative generalized least-squares estimates used in the hierarchical linear model are equivalent to maximum likelihood estimates under multivariate normality. He also shows how to correct for bias in these estimates.

From the Bayesian point of view, there is an emerging body of literature on the use of sampling-based methods to estimate posterior densities. Such methods are quite general and implementation is possible in a variety of settings. Gelfand and Smith (1990) provide an extensive overview. These techniques also allow for the comparison of different models applied to the same data. (See e.g. Belin and Rubin (1994) who use these methods to compare various models of Schizophrenic reaction times.) Simulation-based methods are, however, somewhat controversial. Gelman and Rubin (1992a,b) point out that naive use of the Gibbs sampler (and other iterative simulation methods) can give falsely precise answers, and so should be applied with care.

In conclusion, we have developed methods to analyze longitudinal data where measures come at varying intervals and there is a nested random effect. The EM algorithm is implemented to estimate variance components from a restricted likelihood surface. Empirical Bayes methodology is used to estimate population and random effects given the estimated variance components. We conjecture that failure to account appropriately for all sources of variability may be leading analysts to conclude that certain model effects are significant when they are not.

Appendix

Computational Formulas:
(Note: In the summation, i ranges from 1 to m and j ranges from 1 to k_i)

$$\begin{aligned}
\text{Call} \qquad V_{ij} &= (Z'_{ij}Z_{ij} + \hat{\sigma}^2\hat{D}^{-1})^{-1},\\
T_{ij} &= I_{n_{ij}} - Z_{ij}V_{ij}Z'_{ij},\\
U_i &= (\sum_j W'_{ij}T_{ij}W_{ij} + \hat{\sigma}^2\hat{E}^{-1})^{-1},\\
\text{and} \qquad A &= \Big[\sum_i\sum_j X'_{ij}T_{ij}X_{ij} - X'_{ij}T_{ij}W_{ij}U_iW'_{ij}T_{ij}X_{ij}\Big]^{-1}.
\end{aligned}$$

Then

$$\begin{aligned}
\hat{\alpha} &= A\Sigma_i\Sigma_jX'_{ij}T_{ij}y_{ij} - X'_{ij}T_{ij}W_{ij}U_iW'_{ij}T_{ij}y_{ij},\\
\hat{\text{Var}}(\hat{\alpha}) &= \hat{\sigma}^2A,\\
\hat{c}_i &= U_i\Sigma_j(W'_{ij}T_{ij}y_{ij} - W_{ij}T_{ij}X_{ij}\hat{\alpha}),\\
\hat{\text{Var}}(\hat{c}_i) &= \hat{\sigma}^2U_i(I_{p_c} + \Sigma_jW'_{ij}T_{ij}X_{ij}AX'_{ij}T_{ij}W_{ij}U_i),\\
\hat{b}_{ij} &= -\Big[V_{ij}Z'_{ij}X_{ij} - V_{ij}Z'_{ij}W_{ij}U_i\sum_j W'_{ij}T_{ij}X_{ij}\Big]\hat{\alpha}\\
&\quad + V_{ij}Z'_{ij}y_{ij} - V_{ij}Z'_{ij}W_{ij}U_i\sum_j W'_{ij}T_{ij}y_{ij},\\
\hat{\text{Var}}(\hat{b}_{ij}) &= \hat{\sigma}^2\Big\{V_{ij} + V_{ij}Z'_{ij}\Big\{\Big[X_{ij} - W_{ij}U_i\sum_j W'_{ij}T_{ij}X_{ij}\Big]\\
&\quad \cdot A\Big[\sum_j X'_{ij}T_{ij}W_{ij}U_iW'_{ij} - X'_{ij}\Big] + W_{ij}U_iW'_{ij}\Big\}Z_{ij}V_{ij}\Big\},\\
\hat{\text{Cov}}(\hat{\alpha},\hat{c}_i) &= -\hat{\sigma}^2A\sum_j X'_{ij}T_{ij}W_{ij}U_i,\\
\hat{\text{Cov}}(\hat{\alpha},\hat{b}_{ij}) &= (-\hat{\sigma}^2AX'_{ij} + \hat{\text{Cov}}(\hat{\alpha},\hat{c}_i)W'_{ij})Z_{ij}V_{ij},\\
\hat{\text{Cov}}(\hat{c}_i,\hat{b}_{ij}) &= -\hat{\sigma}^2U_iW'_{ij}Z_{ij}V_{ij} - U_i\sum_j W'_{ij}T_{ij}X_{ij}\hat{\text{Cov}}(\hat{\alpha},\hat{b}_{ij}).
\end{aligned}$$

References

Belin, T. R. and Rubin, D. B. (1994). The Analysis of Repeated-Measures Data on Schizophrenic Reaction Times Using Mixture Models. Statist. Med. (in press).

Bryk, A. S. and Raudenbush, S. W. (1992). Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Publications, Newbury Park.

Deely, J. J. and Lindley, D. V. (1981). Bayes empirical Bayes. J. Amer. Statist. Assoc. **76**, 833-841.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. Roy. Statist. Soc. Ser.B **39**, 1-38.

Dempster, A. P., Rubin, D. B. and Tsutakawa, R. K. (1981). Estimation in covariance components models. J. Amer. Statist. Assoc. **76**, 341-353.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. J. Amer. Statist. Assoc. **70**, 311-319.

Environmental Noise Consultants, Inc. (1987). Industrial audiometric data – computer tape and documentation. U.S. Department of Commerce, National Technical Information Service.

Fellingham, G. W. and Raghunathan, T. E. (1994). Sensitivity of point and interval estimates to distributional assumptions in longitudinal data analysis. Comm. Statist. Simulation Comput. (in review).

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. J. Amer. Statist. Assoc. **85**, 398-409.

Gelman, A. and Rubin, D. B. (1992a). Inference from iterative simulation using multiple sequences. Statist. Sci. **7**, 457-511.

Gelman, A. and Rubin, D. B. (1992b). A single series from the Gibbs sampler provides a false sense of security. In Bayesian Statistics **4** (Edited by J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), 625-631. Oxford University Press, Oxford.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika **73**, 43-56.

Goldstein, H. (1989). Restricted unbiased iterative generalized least-squares estimation. Biometrika **76**, 622-623.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems (with comment). J. Amer. Statist. Assoc. **72**, 320-340.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In Proc. 4th Berkeley Symp. Math. Statist. Probab. **1**, 361-379. University of California Press, Berkeley.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. Biometrics **38**, 963-974.

Melnick, W. (1984). Evaluation of industrial hearing conservation programs: A review and analysis. Amer. Indust. Hygiene Assoc. J. **45**, 459-467.

Royster, J. D. and Royster, L. H. (1986). Audiometric data base analysis. In Noise and Hearing Conservation Manual (Edited by E. H. Berger, W. D. Ward, J. C. Morrill, and L. H. Royster), 4th edition. Akron, OH. Amer. Indust. Hygiene Assoc. J.

Department of Statistics, Brigham Young University, Provo, UT 84602, U.S.A.

(Received September 1993; accepted August 1994)