Comment

Nan Laird and Nick Lange

One can scarcely think of any problem in growth curve estimation without bringing to mind the extensive contributions of C. R. Rao, which span some thirty years of research in this area. With his present paper, Professor Rao demonstrates once again his capacity for stimulating his audience to think in new ways about an important problem, in this case the prediction of future observations.

We choose to confine our discussion to the use of least squares and empirical Bayes procedures for the polynomial growth curve models discussed in Sections 4.1 to 4.4, primarily because of the connection to our previous work (Lange and Laird, 1986). As Rao notes, there are many ways to approach the empirical Bayes prediction problem, and these choices depend upon what data one employs when estimating the parameters of the prior model. In Rao's notation, one can either use only the past data values $\{U_i, X_i\}_{i=1}^{n+1}$ for all $n + 1$ individuals (Method I) or all data values $\{U_i, W_i, X_i, x_i\}_{i=1}^{n}$ for only n individuals (Method II). Both methods retain balance in the data, and thus either choice allows one to obtain closed-form estimates for $(\gamma, \sigma^2, \Gamma)$, and also for the predicted $\hat{W}_c$. There are also many numerical methods available for estimating these parameters with unbalanced data (cf. Goldstein, 1986; Laird, Lange and Stram, 1987; Longford, 1987), and hence one could in principle use all available data on n individuals, $\{U_i, W_i, X_i, x_i\}_{i=1}^{n}$, and only past data $\{U_c, X_c\}$ on the cth individual (Method III). Reinsel (1985) also discusses noniterative estimators, similar to Rao's and suitable for unbalanced data. Figure 1 illustrates how our three methods compare in their use of data available for estimating the prior and for prediction. Each method obviously gives rise to a different empirical Bayes predictor.

Rao (Table 7) considers only a Method I empirical Bayes predictor. Although such an approach yields a minimum CVAE under a polynomial growth curve model for the ramus and dental data, its performance with the mice data is very poor when compared to Rao's calibrated predictor.

Nan Laird is Professor, Department of Biostatistics, School of Public Health, Harvard University, Boston, Massachusetts 02115. Nick Lange is Assistant Professor, Department of Community Health, Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912. The comments of the second author were written while he was with the Department of Mathematics, Massachusetts Institute of Technology.

On intuitive grounds, we would expect that a Method I predictor might perform poorly in a variety of situations. Method I makes no use of $\{W_i, x_i\}_{i=1}^{n}$ in obtaining $\hat{W}_c$, and thus is essentially extrapolating the prior model beyond the range of the data used for the fit. The same is true for Rao's individual regression predictor, which performs slightly worse than his empirical Bayes predictor. Rao uses the calibrated predictor to fix up the individual regression predictor. One can also calibrate the Method I predictor in exactly the same way. Alternately, one could use Method II empirical Bayes, which includes the $\{W_i, x_i\}_{i=1}^{n}$ to fit the prior at the expense of not using any data on the cth individual. Method III would appear to be the ideal empirical Bayes predictor, because it makes maximum use of available data to estimate prior parameters.

We compare the four empirical Bayes predictors (Methods I, II and III and Method I calibrated) on the three example data sets in Table 1. Rao's individual regression and calibrated predictors are given in parentheses next to Method I and Method I calibrated, respectively. (The CVAE for our Method I predictor differ slightly from those in Table 7 of Rao, due apparently to computational issues. All of our calculations were obtained through use of the S system running on a Sun 3/160 workstation.)

For the mice data, Method II shows marked improvement over Method I, although the calibrated Method I is clearly superior to any other. The calibrated empirical Bayes predictor appears more stable than its least squares counterpart for the mice data. Method III's performance with these data is disappointing, showing hardly any improvement over Method I. An ordinary Method I empirical Bayes remains the best predictor for both the ramus and dental data sets, with neither the calibrated Method I nor Methods II and III showing any improvement.

A seemingly attractive feature of the Method I predictor is that Rao gives a formula (4.4.15) that expresses this predictor as a function only of $n^{-1}B = \hat{\Gamma} + \hat{\sigma}^2(X^TX)^{-1}$. Because B is always positive definite, the issue of nonpositive definite $\hat{\Gamma}$ never arises; in fact, Γ never need be estimated explicitly. Rao's expression (4.4.15) can be written alternately as

$$\hat{W}_c^{(B)} = x[b_c^{(l)} - g'\hat{Q}(b_c^{(l)} - \hat{\gamma})],$$

where

$$\hat{Q} = \hat{\sigma}^2(X^TX)^{-1}[\hat{\Gamma} + \hat{\sigma}^2(X^TX)^{-1}]^{-1}$$

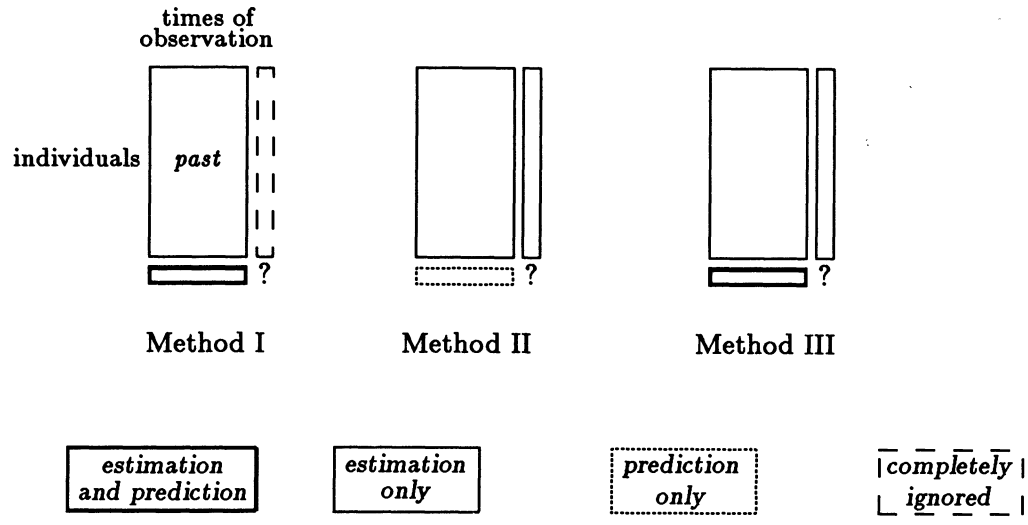

FIG. 1. *Three methods for obtaining an empirical Bayes predictor of a future observation for a current individual (indicated by ?). The various uses of observed data for the CVAE comparison of predictors under the different methods are denoted as follows. Thick-lined boxes, for estimation of parameters and for prediction; thin-lined boxes, for estimation of parameters only; dotted box, for prediction only; long-dashed box, data completely ignored.*

and

$$g' = \left[\frac{(n+1)(s-k-1)}{(n+1)(s-k-1)+2}\right]\left[\frac{n-k-3}{n}\right].$$

The matrix $\hat{Q}$ appears in Rao's expression (4.4.3) and clearly shows that whenever $\hat{\Gamma}$ has nonpositive diagonal elements, the eigenvalues of $\hat{Q}$ can be greater than 1. Such cases correspond to "shrinking beyond the prior" in the univariate dimension. However, such "over-shrinking" is probably compensated by the use of a $g' < 1$, and considerably so in some cases. Substituting $n-k-2$ for $n-k-3$ in g' would give a constant consistent with Reinsel's formula (1985, equation 6, page 643), and also consistent with the g in equation (4.4.9) of Rao.

With Method II and III, however, Γ must be estimated explicitly. For Method II, Rao suggests zeroing out all rows and columns of $\hat{\Gamma}$ corresponding to any of its nonpositive diagonal elements. Alternately, one can obtain the restricted maximum likelihood (REML) estimator of Γ, as has been described by Lange and Laird (1986). Covariance component estimators are maximum likelihood (ML) or REML only if their solutions yield positive definite matrices. When such is not the case, the ML or REML solution occurs on a boundary of the parameter space, and the total amount of variation in the data can be apportioned differently between these estimators, as follows. From the original $\hat{\sigma}^2$ and $\hat{\Gamma}$, we have

$$\hat{\Gamma} = \hat{\sigma}^2 G^{-T}(\hat{\sigma}^{-2}\hat{\Lambda} - I)G^{-1},$$

where $\hat{\Lambda} = \hat{\sigma}^2 I_r + G^T\hat{\Gamma}G$ and G a $k+1$ square matrix satisfying $GG^T = X^TX$. Note that $G^{-T}\hat{\Lambda}G^{-1} = \hat{\Gamma} + \hat{\sigma}^2(X^TX)^{-1}$, and hence $\hat{\Lambda}$ is always positive definite for any $\hat{\Gamma}$. ($G^{-T}\hat{\Lambda}G^{-1}$ is Rao's $n^{-1}B$.) Let $\hat{E} = \text{diag}(\hat{e}_1, \cdots, \hat{e}_{k+1})$ be the diagonal matrix of eigenvalues of $\hat{\sigma}^{-2}\hat{\Lambda}$ and $\hat{L}$ be the orthogonal matrix of eigenvectors. When $\hat{\Gamma}$ is not positive definite, we may use $\hat{E}$ and $\hat{L}$ to extract the positive definite part of $\hat{\Gamma}$. Let $q \le k+1$ be the number of eigenvalues in $\hat{E}$ that are greater than or equal to 1. Also, let $\hat{E}^*$ be the version of $\hat{E}$ with $\hat{e}_i$ replaced by zero whenever $\hat{e}_i < 1$, for $i = 1, \cdots, k+1$, and let I^* be a version of the $k+1$ identity matrix with its corresponding diagonal elements replaced with zeroes. Lange and Laird (1986) showed that by adapting an approach given by Amemiya (1985, pages 116 and 117) to the present context, the REML solutions on the boundary are

$$\hat{\Gamma}_+ = \hat{\sigma}^2 G^{-T}\hat{L}(\hat{E}^* - I^*)\hat{L}^T G^{-1},$$

and

$$\hat{\sigma}^2_+ = h\hat{\sigma}^2 + (1-h)\sum_{i=q+1}^{k+1}\hat{e}_i,$$

where, for Method II,

$$h = \frac{n(s-k-1)}{n(s-k-1)+d(k+1-q)}.$$

For ML estimation, $d = n$, and $d = n-k-1$ for REML estimation.

We computed the Method II estimates in three ways. First, we simply used Rao's formula (4.4.9) with $\hat{\Gamma}$ as in (4.4.8), making no adjustment for nonpositive definite values, as this estimator would be comparable to Method I. We then also applied Rao's zeroing out suggestion. Last, we used Rao's formula (4.4.9) replacing any nonpositive definite value of $\hat{\Gamma}$ by the REML estimate. (We defined "nonpositive" on the computer

TABLE 1
CVAE for the three-example data sets under the four empirical Bayes prediction methods

Previous measurements used	Degree of polynomial fitted	Empirical Bayes predictors			
		Method I[a]	Method I calibrated[b]	Method II	Method III
		Mice data: prediction of Y_7, $n = 13$[c]			
$Y_1 - Y_{6or7}$[d]	5	— (7.472)	— (0.252)	3.483	
	4	0.400 (0.600)	0.186 (0.235)	0.254	
	3	0.142 (0.175)	0.070 (0.093)	0.098	
	2	0.088 (0.104)	0.033 (0.037)	0.067	
	1	0.194 (0.206)	0.036 (0.035)	0.131	0.182
$Y_2 - Y_{6or7}$	4	— (2.405)	— (0.235)	0.900	
	3	0.175 (0.241)	0.102 (0.141)	0.125	
	2	0.076 (0.095)	0.033 (0.040)	0.063	
	1	0.144 (0.158)	0.036 (0.035)	0.087	0.153
$Y_3 - Y_{6or7}$	3	— (0.757)	— (0.192)	0.258	
	2	0.070 (0.096)	0.032 (0.052)	0.045	
	1	0.097 (0.111)	0.035 (0.034)	0.062	0.089
$Y_4 - Y_{6or7}$	2	— (0.229)	— (0.094)	0.065	
	1	0.053 (0.066)	0.033 (0.034)	0.044	0.053
$Y_5 - Y_{6or7}$	1	— (0.055)	— (0.033)	0.036	0.046
		Ramus data: Prediction of Y_4, $n = 20$			
$Y_1 - Y_{3or4}$	2	— (2.989)	— (2.172)	0.751	
	1	0.533 (0.584)	0.616 (0.683)	0.558	0.602
$Y_2 - Y_{3or4}$	1	— (0.812)	— (0.751)	0.637	0.860
		Dental data: Prediction of Y_4, $n = 27$			
$Y_1 - Y_{3or4}$	2	— (47.398)	— (9.483)	6.143	
	1	2.266 (3.998)	2.496 (3.680)	2.722	2.767
$Y_2 - Y_{3or4}$	1	— (12.426)	— (8.358)	3.694	

[a] Values in parentheses are from individual regression predictions.
[b] Values in parentheses are from calibrated individual regression predictions.
[c] Entries are 13 times the actual values.
[d] Methods II, III and I calibrated use measurements taken on the last occasion of observation, but not for the current individual; Method I ignores these data.

as any diagonal value of $\hat{\Gamma}$ less than or equal to 5×10^{-5}.) The CVAE reported in Table 1 for Method II are based on REML estimates. The CVAE for the first two of our approaches were very similar, not uniformly smaller or larger than those given in the table.

The problem of adjustment on a boundary actually occurs frequently in calculations with these data. For example, under Method II, the fit of a linear growth curve model to the last three time observations of the dental data, deleting the data on the first individual (considered as the "current" individual), yields

$$\hat{\sigma}^2 = 2.15 \quad \text{and} \quad \hat{\Gamma} = \begin{bmatrix} 4.59 & 0.93 \\ 0.93 & -0.23 \end{bmatrix},$$

as one may anticipate by examining the plot of this data set given in Figure 2. We see considerably greater within-variation relative to between-variation for the dental data. (Such does not appear to be the case for the mice or ramus data.) Similar nonpositive definite $\hat{\Gamma}$ occur when every individual is deleted one-at-a-time for this case, as well as for several others in the other two example data sets. Application of our adjustment to the original $\hat{\sigma}^2$ and $\hat{\Gamma}$ yields

$$\hat{\sigma}_+^2 = 1.42 \quad \text{and} \quad \hat{\Gamma}_+ = \begin{bmatrix} 4.59 & 0.88 \\ 0.88 & 0.17 \end{bmatrix}.$$

For Method III, we used the REML estimate of σ^2 and Γ, calculated iteratively as described by Laird, Lange and Stram (1987). As noted by Reinsel (1985), it is not possible to give a closed form for the optimal value of the "shrinking constant" in obtaining empirical Bayes estimates in the unbalanced case. Hence, as an approximation, we used Reinsel's constant for the balanced case, which assumes all $n + 1$ individuals contribute $s + 1$ observations in estimating the prior. This yields

$$\hat{W}_c^{(B)} = x[b_c^{(l)} - g^* \hat{Q}(b_c^{(l)} - \hat{\gamma})],$$

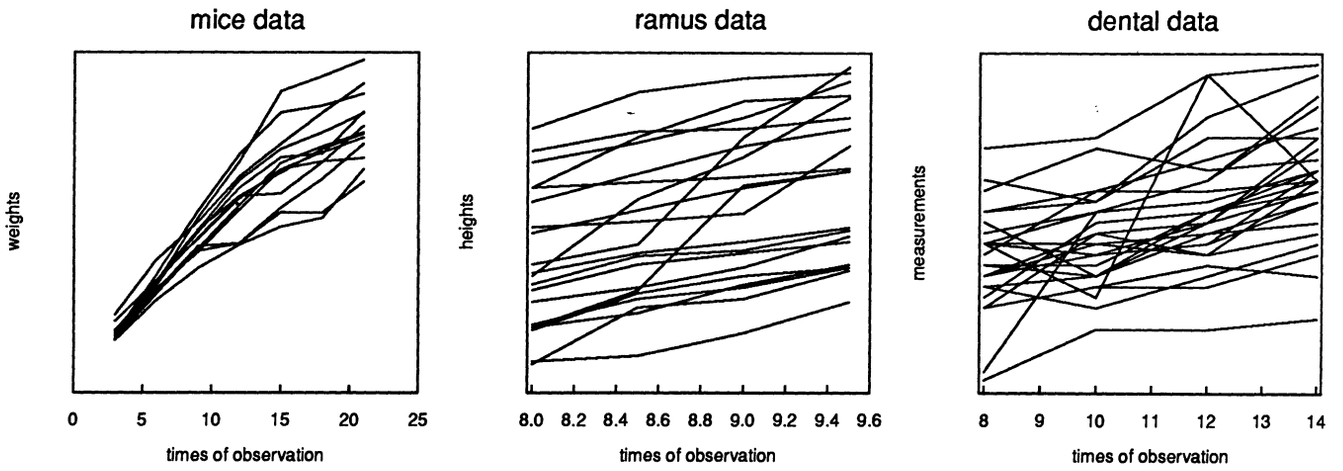

FIG. 2. *Plots of example data.*

where

$$g^* = \left[\frac{(n+1)(s-k)}{(n+1)(s-k)+2}\right]\left[\frac{n-k-2}{n}\right].$$

In conclusion, we agree with Professor Rao that his empirical Bayes predictor of future observations in growth curve models performs better than its least squares counterpart. We have also described several other empirical Bayes prediction methods. With the three example data sets, we have found our calibrated empirical Bayes predictor to yield smaller CVAE and to be more stable than its calibrated least squares counterpart.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health Grant GM29745. We also thank the Statistics Center, Massachusetts Institute of Technology, for the use of their computing resources.

ADDITIONAL REFERENCES

AMEMIYA, Y. (1985). What should be done when an estimated between-group covariance matrix is not nonnegative definite? *Amer. Statist.* **39** 112–117.

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

LAIRD, N., LANGE, N. and STRAM, D. (1987). Maximum likelihood computations with repeated measures: Application of the EM algorithm. *J. Amer. Statist. Assoc.* **82** 97–105.

LANGE, N. and LAIRD, N. M. (1986). Random-effects and growth-curve modeling for balanced and complete longitudinal data. Technical Report, Dept. Mathematics, Massachusetts Institute of Technology, Cambridge, Mass.

LONGFORD, N. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. Technical Report 87-77, Educational Testing Service, Princeton, N. J.

REINSEL, G. (1985). Mean squared error properties of empirical Bayes estimators in a multivariate random-effects generalized linear model. *J. Amer. Statist. Assoc.* **80** 642–650.

Comment: On Exchangeability Judgments in Predictive Modeling and the Role of Data in Statistical Research

David Draper

Professor Rao has shared with us some thought-provoking ideas on prediction in growth curve modeling. The paper has four basic attributes, two of which seem positive and two negative. On the positive side,

- the basic problem is predictive in nature, thereby emphasizing inference on observable quantities (future values of outcome variables of interest) rather than on unobservable quantities (parameters); and

David Draper is a member of the Statistical Research and Consulting Group, Department of Economics and Statistics, RAND Corporation, 1700 Main Street, Santa Monica, California 90406.