ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation Grant DMS-8601314 and National Institutes of Health Grant GM-25271.

ADDITIONAL REFERENCES

FEARN, T. (1975). A Bayesian approach to growth curves. *Biometrika* **62** 89–100.

GEISSER, S. (1970). Bayesian analysis of growth curves. *Sankhyā Ser. A* **32** 53–64.

GEISSER, S. (1974). A predictive approach to the random effect model. *Biometrika* **61** 101–107.

GEISSER, S. (1975b). A new approach to the fundamental problem of applied statistics. *Sankhyā Ser. B* **37** 385–397.

GEISSER, S. (1980a). Growth curve analysis. In *Multivariate Analysis Handbook. I. Analysis of Variance* (P. R. Krishnaiah, ed.) 89–115. North-Holland, Amsterdam.

GEISSER, S. (1980b). Predictive sample reuse techniques for censored data (with discussion). In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) 430–468. University Press, Valencia.

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

Rejoinder

C. Radhakrishna Rao

For ready reference, the problem considered in the paper is the following. We have observations (U_i, W_i), where U_i is a p vector of measurements taken at p time points and W_i is the measurement taken at a future $(p+1)$th time point, on $i = 1, \cdots, n$ individuals drawn from a population S. Another individual drawn from S provides the first p measurements U_c, and the problem is to predict the $(p+1)$th measurement W_c on the individual.

What is relevant in a problem of this kind is the conditional (predictive) distribution of W_c given U_c,

$$P_{\text{pred}}(W_c \mid U_c, \psi), \tag{1}$$

with respect to some *reference* population, where ψ is a parameter specific to the reference population. One choice of the reference population is S itself. However, when ψ is unknown, we have two possibilities. We may estimate ψ by $\hat{\psi}$ from the available data

$$(U_i, W_i), \quad i = 1, \cdots, n, \quad \text{and} \quad U_c \tag{2}$$

and consider an estimate of (1),

$$P_{\text{empred}}(W_c \mid U_c, \hat{\psi}), \tag{3}$$

as the basic conditional distribution. An alternative is to consider S as a member of a super population generated by a prior distribution on ψ, in which case the relevant distribution is

$$P_{\text{Baypred}}(W_c \mid U_c) \tag{4}$$

obtained by integrating (1) with respect to the posterior distribution of ψ given the observed data (2). On the other hand, we may wish to consider the current individual's observations (U_c, W_c) as arising from a stochastic process *specific* to the individual. In such a case the empred (3) is defined in terms of $\hat{\psi}$ estimated from U_c alone and the Baypred (4) is obtained by choosing a prior on ψ and computing the posterior distribution based on U_c alone. The second possibility of considering an individual separately is specially recommended when on the basis of an initial examination of data, the measurements U_c are found to have an unusual pattern different from those of $U_1, \cdots, U_n$.

The theory as developed in Section 2 of the paper and outlined above is complete in itself although its practical applications involves various issues that I would like to discuss on the basis of the comments made by the discussants of my paper.

DATA AND CROSS-EXAMINATION OF DATA

For illustrative purposes I have chosen three real data sets, which are well documented and which have been studied by a number of authors for predictive purposes. I thank Izenman for giving some details about the mice data that will be helpful to future investigators. I have made the necessary corrections regarding the original source of the dental data based on his comments. In my analysis of the mice data, I omitted the measurements on one mouse (not reported in Table 2, but can be found in Izenman's comments), which looked different from the others and whose weight actually decreased at the end. Izenman asks what effect it would have had on my results if this mouse had been retained in the data set. I have deliberately chosen my reference population as the set of mice that generally exhibit an increase in growth at all time points and derived the appropriate prediction

formula. Perhaps this is more relevant from a practical point of view. A large difference between the observed value and the predicted value would then indicate some disturbance in the growth process. In theory, the abnormal mouse could have been included but then the reference population and the object of prediction would be different.

Draper thought that I have not made use of the "realness" of the data sets and perhaps simulated data might have been used to illustrate my methods. I believe there is some challenge in using real data sets where the underlying stochastic mechanism is highly complex and unknown and that exhibit peculiarities of various types. The preanalysis (or cross-examination) of data, not fully reported in the paper, consisted of drawing graphs, looking for clusters, outliers and possible errors in recording, etc. The decision to omit the abnormal mouse was based on such an analysis.

Of course, if clusters could be identified as with the dental data on boys and girls, they should be treated separately and I fully agree with Draper on this issue. I was not happy with the dental data in the first instance as the measurements exhibit some departures from a regular growth process. I have included it for historical reasons as this was chosen for illustration by Potthoff and Roy (1964) while introducing growth curve models. Further, the data were used by Lee and Geisser (1975) without making the sex distinction to illustrate some predictive techniques developed by them. I have used the data without omitting the aberrant observations and taking my reference population as the mixture of boys and girls for the sole purpose of comparing my results with those of Lee and Geisser.

LEAVE-ONE-OUT (LOO) METHOD

I have used the LOO method for cross-validation in choosing a model for prediction or more precisely in choosing a prediction function. I believe that cross-validation assessment error (CVAE) has wider applicability than other criteria such as AIC, C_p, S_p and A_p. I am glad to see from Akaike's comments that some of my results on selection of variables in estimating regression without assuming any covariance structure can be obtained through the use of AIC. But suppose that one wishes to compare the relative efficiencies of linear predictors estimated by the least squares and ridge methods using the same number of independent variables. The CVAE criterion can be easily used in such cases, but not the others without some modification. Also, let us consider the problem of choosing the degree of the polynomial in time to be fitted to the first p measurements for the prediction of future values by extrapolation. The use of CVAE for this purpose is quite straightforward and is illustrated in the paper. It is not clear how AIC, C_p, S_p and A_p can be used in this problem. These criteria are useful in examining the goodness of fit of a model to the measurements in a given time span, but not in assessing the adequacy of a model that can be fitted to measurements in a given time range and used to predict measurements outside this time range. It would be useful to derive criteria similar to AIC, etc., in such cases. Some results obtained in this direction will be reported elsewhere.

Both Izenman and Geisser raised the question of the reliability of the CVAE as an estimate of the prediction error in growth models. I have used the CVAE only as a comparative measure for selecting between alternative models. For actual estimates of prediction errors, methods such as those proposed by Efron may have to be used, as mentioned in Izenman's comments. I believe graphical representation of data is extremely valuable in model building, scrutiny of data and in looking for a suitable transformation of data to simplify statistical analysis. A graph can suggest some property of the data, but confirmation is needed by appropriate statistical analysis. Similarly, what is inferred from purely statistical analysis could be examined through graphical representation for further illumination and explanation. Graphical and inferential data analyses should go hand in hand. One is not a substitute for the other. Draper says that my conclusion that "the best procedure for prediction is to draw a line through the points corresponding to y_{p-1} and y_p and extrapolate for y_{p+1}" could have been drawn by *staring* at the graphs of the growth curves. This is perhaps a misleading statement if Draper meant "that staring at graphs and looking at data" are the main tools of inference.

I have some fascination for the graphs of growth curves. In fact, my earlier papers had graphs of growth curves. I did not include the graphs in my paper as the complete data sets are given in tables. I am glad that the graphs of the present data will appear at least in the comments section.

SELECTION OF VARIABLES

It is a common experience that in predicting a variable (such as y_{p+1}) by its estimated regression on a number of independent variables (such as $y_1, \cdots, y_p$ in our problem), the efficiency of prediction goes down as the number of independent variables increases beyond a certain set. Thus, the selection of independent variables is of paramount importance in regression, and there is a vast amount of literature on the subject. Traditionally, selection of variables means choosing a subset of the independent variables. In my paper I have considered variable selection in a more general way as that of selecting a subspace of the vector space $\{y_1, \cdots, y_p\}$ generated by all linear combinations of

the variables $y_1, \cdots, y_p$. (This concept seems to be new and some further elaboration may be useful.) The number of such subspaces is uncountable. However, in any given problem, one can narrow down the choice to a few subspaces depending on the nature of the problem. I have considered the following subspaces:

$$\{y_i, \cdots, y_p\}, \quad i = 1, \cdots, p, \tag{5}$$

$$\{b_0, b_1, \cdots, b_k\}, \quad k = 1, 2, \cdots, \tag{6}$$

$$\{b_0\psi_0(p+1) + \cdots + b_k\psi_k(p+1)\}, \quad k = 1, 2, \cdots, \tag{7}$$

$$\{P_1, \cdots, P_k\}, \quad k = 1, 2, \cdots, \tag{8}$$

where $b_0, b_1, \cdots, b_k$ in (6) are the coefficients of the kth degree polynomial fitted to subsets in (5) as in Section 4.2 of the paper, $b_0\psi_0(p+1) + \cdots + b_k\psi_k(p+1)$ is the individual regression predictor of y_{p+1} obtained by extrapolating the kth degree polynomial fitted to subsets in (5) as in Section 4.3, and $P_1, \cdots, P_k$ in (8) are the principal components of the subsets in (5) as in Section 5.1. I have computed the regression of y_{p+1} on each of the subspaces in (5), (6), (7) and (8) and compared their relative efficiencies in predicting y_{p+1}. I hope the clarification I have given above will answer some of the questions and doubts raised by Draper regarding the methods discussed in Sections 4 and 5. Draper also wondered about my choice of the subsets in (5) instead of all possible subsets. To anyone working in time series, the choices in (5) are the obvious ones for determining how far back one has to go to capture all the dependence. Draper applied standard backward selection regression methods and suggested dropping the variables in a different order, and not in the order 1, 2, $\cdots$ I have considered. This is not surprising because the partial correlation of y_{p+1} with any y_i given y_p is extremely small and any method of backward selection should eliminate the variables in a random order. Traditional methods of selection of variables can be misleading as shown in Rao (1984) in the case of data on physiological measurements analyzed by Fisher (1938).

CALIBRATION

This appears to be a new technique. Generally speaking, an individual's future should be predicted by fitting an appropriate model to his past data and extrapolating for the future value. Such a predicted value may not be precise enough because of individual variations in the model and/or inadequacy of the model itself. Such deficiencies can be remedied to some extent by calibrating (fine tuning) the predicted value by using previous data. I am glad that Laird and Lange further illustrated this method by using it on empirical Bayes estimators also. The method is being tried in other projects currently under investigation at the Center for Multivariate Analysis of the University of Pittsburgh.

EMPIRICAL BAYES PREDICTOR

This is essentially fitting the regression of y_{p+1} on $y_i, \cdots, y_p$ (for any chosen value of i) using a structure for the mean value (μ) and covariance matrix (Σ) of $(y_{p+1}, y_p, \cdots, y_i)$, of the form

$$\mu = X\gamma, \qquad \Sigma = X\Gamma X' + \sigma^2 I, \tag{9}$$

where γ and Γ are unknown, and X is the matrix of coefficients of orthogonal polynomials in time. Naturally we expect some improvement over the usual regression method based on arbitrary μ and Σ. But much depends on to what extent the structure (9) holds, how γ and Γ are estimated and what part of the previous data is to be used. In my experience, the inclusion of the incomplete data on the current individual for estimating the unknowns (Method III in the comments by Laird and Lange) is counterproductive and is not recommended. As Brillinger suggests, some research is needed on the estimation of γ and Γ to realize the full potential of this method. Calibration of the Method I predictor as suggested by Laird and Lange is another possibility, although I feel that calibration of individual regression estimators as done in my paper can provide equally good estimators.

BORROWING STRENGTH

In my problem, I have clearly stated the objective as one of predicting values for future individuals drawn from the same population that provided the basic data, in which case the stochastic nature of an individual's parameters is implicit. But in the problem of Brillinger, one may have to be cautious. Perhaps the inclusion of some concomitants associated with locations in the model might be of some help.

GENERAL REMARKS

Draper has probably misunderstood my factor analytic model. This is an attempt to transform the time variable to simplify the shape of the growth trend. The graphs of the growth curves of mice on a suitably transformed time axis are reproduced in Figure 1. The shapes are more well defined than in the corresponding graphs based on actual time (see the graph in the comments by Draper). Further the general shape of the graphs over the whole period is linear, a phenomenon that I have noticed in all my previous studies on growth curves. I have used factor analytic techniques to estimate orthogonal functions whose linear combinations can accommodate trends of complex

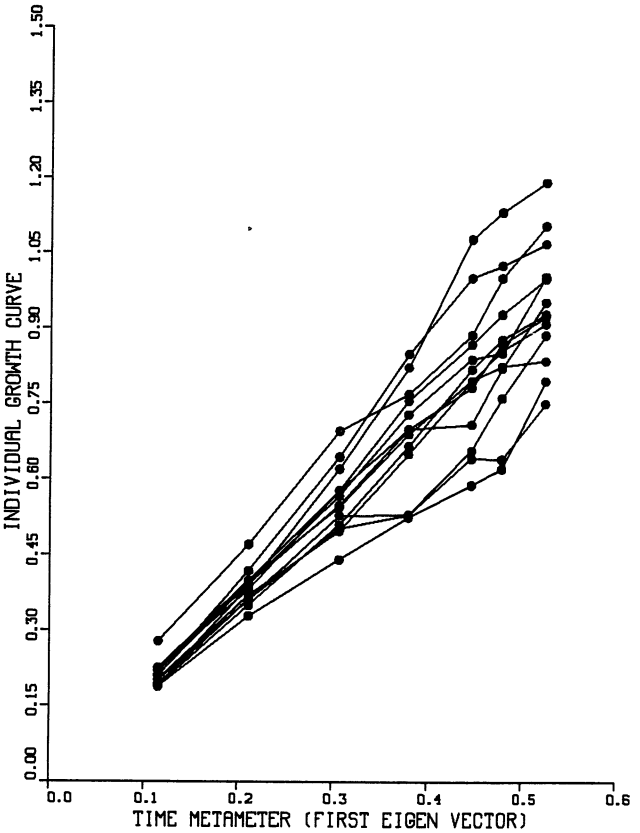

FIG. 1. *The mice data of Table 2.*

shape. It is my modest suggestion that Draper should consider plotting growth curves against a suitable time metameter in one or two dimensions to reduce the number of families of growth curves in his analysis. (The term time metameter does not seem to be in Draper's dictionary.) Using uniform linearity in transformed time, I find that low levels of CVAE can be achieved using all the previous measurements. In the earlier methods some of the previous measurements had to be dropped to reduce the CVAE. I have not explored the full potentialities of the factor analytic approach due to the computational difficulties involved. I hope to complete this project at a future date.

I am not sure whether it is a good idea to consider a discrete number of shapes as suggested by Draper instead of a continuously varying type of shape (as implicit in my formulation of a polynomial trend with varying coefficients or as in Young (1977) restricting the variation of the coefficients to some extent). The data are not usually numerous enough to be viewed through a complex model. One has to look for variables that carry much of the information relevant to the problem at hand and try simple models. Further, in problems of the type we are considering, it is hard to determine what type of curve is appropriate for a given individual and a uniform prescription for all individuals may be necessary. (The Bayesian approach also results in a uniform prescription once the posterior is determined.) However, there can arise individual cases who may have to be treated differently, especially when they exhibit an abnormal pattern of previous growth not represented in the training samples. In such a case the individual's empred and Baypred become relevant.

Draper says that in fitting the regression of y_{p+1} on subsets of $y_1, \cdots, y_p$ in Section 3 of my paper I have not used a growth curve model. This is true. But in all the later sections I have exploited the structure of the data using a growth curve model. The purpose of the exercise is to compare various methods. The fact remains that the use of the structure did not help much. The direct regression of y_{p+1} on y_p or on (y_p, y_{p-1}) is the winner, and I would not have discovered this phenomenon but for my trying the least obvious methods. It appears that any modeling involving the estimation of more than two or three parameters introduces too much noise.

Although I have reported the total or the average CVAE in various tables of my paper, the error for each individual by each method was examined. Such lists of errors would enable the use of other criteria such as mean deviation or minimax. In some practical situations minimax may be relevant. There was always one irritant, the error in predicting the dental measurement for the boy with number 20, which was large compared to the rest because of the substantial decrease in the last measurement below the previous one. If the results on the dental data are somewhat different from that of the others, it may be due to this aberrant observation.

I wish to thank all the discussants for the interest shown in my paper, their further suggestions for improvement and their constructive criticism. The problem is an interesting one, which involves an unusual interplay between classical and Bayesian methods and which may also provide material for examining the limitations in using a purely classical or purely Bayesian approach. (I have noticed in Draper's proposed analysis which is mostly Bayesian in character, he estimates the number of possible shapes of growth curves using the data.)

I eagerly look forward to Draper's proposed analysis of the growth data. I hope he will consider the measurements on the mouse I omitted from my analysis and suggest how they can be incorporated in his analysis. I would have appreciated it if Draper, instead of making lengthy comments running into several pages, had given a half page table giving the predictions resulting from his analysis for comparison with observed values.

I am aware that statisticians as other scientists cannot avoid making subjective judgments in analyz-

ing data. But as statisticians thrive on meddling in other people's business, they have the social responsibility of establishing the validity of their results in an objective manner, using the same data if possible, or applying them on fresh data when they become available.

ADDITIONAL REFERENCES

FISHER, R. A. (1938). On the statistical treatment of the relation between sea-level characteristics and high altitude acclimatization. *Proc. Roy. Soc. Ser. B* **126** 25–29.

RAO, C. R. (1984). Discussion of the paper by Robin Sibson. *J. Roy. Statist. Soc. Ser. A* **147** 205–207.