from the training sample as priors, construct a posterior predictive distribution for the unobserved time point given that individual's growth curve up to that point. The result of all this will be an analysis in which the data help make the necessary exchangeability judgments adaptively, and in which the posterior predictive variability captures all three sources of uncertainty above—structural, estimation and prediction.

I am grateful to Professor Rao for having written a paper that provoked a great deal of thought in me, and I look forward to comparing the results of this propagation of uncertainty analysis with those from his prediction methods and from other approaches to prediction in growth curve models.

ACKNOWLEDGMENTS

In preparing these remarks I have had the benefit of comments from James Hodges and John Rolph.

ADDITIONAL REFERENCES

ANDREWS, D. F. and HERZBERG, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker.* Springer, New York.

BRILLINGER, D. R. (1986). Comment on "Generalized additive models," by T. Hastie and R. Tibshirani. *Statist. Sci.* **1** 310–312.

DE FINETTI, B. (1974, 1975). *Theory of Probability*, **1, 2**. Wiley, New York.

DEGROOT, M. H. (1987). A conversation with C. R. Rao. *Statist. Sci.* **2** 53–67.

DRAPER, D., HODGES, J. S., LEAMER, E. E., MORRIS, C. N. and RUBIN, D. B. (1987). A research agenda for assessment and propagation of model uncertainty. The Rand Corporation, N-2683-RC, November 1987.

DUMOUCHEL, W. H. and HARRIS, J. E. (1983). Bayes methods for combining the results of cancer studies in humans and other species (with discussion). *J. Amer. Statist. Assoc.* **78** 293–315.

HARRISON, P. J. and STEVENS, C. F. (1971). A Bayesian approach to short-term forecasting. *Oper. Res. Quart.* **22** 341–362.

HARRISON, P. J. and STEVENS, C. F. (1976). Bayesian forecasting (with discussion). *J. Roy. Statist. Soc. Ser. B* **38** 205–247.

HOAGLIN, D. C., MOSTELLER, F. and TUKEY, J. W., eds. (1983). *Understanding Robust and Exploratory Data Analysis.* Wiley, New York.

HOAGLIN, D. C., MOSTELLER, F. and TUKEY, J. W., eds. (1985). *Exploring Data Tables, Trends, and Shapes.* Wiley, New York.

HODGES, J. S. (1987). Uncertainty, policy analysis and statistics (with discussion). *Statist. Sci.* **2** 259–291.

HODGES, J. S. and DRAPER, D. (1987). On the information content of models. In preparation.

LAGAKOS, S. W., WESSEN, B. J. and ZELEN, M. (1986). An analysis of contaminated well water and health effects in Woburn, Massachusetts (with discussion). *J. Amer. Statist. Assoc.* **81** 583–614.

LEAMER, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data.* Wiley, New York.

MOSTELLER, F. and WALLACE, D. L. (1984). *Applied Bayesian and Classical Inference: The Case of the Federalist Papers.* Springer, New York.

REINSEL, G. C. and TIAO, G. C. (1987). Impact of chlorofluoromethanes on stratospheric ozone. *J. Amer. Statist. Assoc.* **82** 20–30.

SEARLE, S. R. (1971). *Linear Models.* Wiley, New York.

SMITH, A. F. M. (1983). Bayesian approaches to outliers and robustness. In *Specifying Statistical Models: From Parametric to Nonparametric, Using Bayesian and Non-Bayesian Approaches* (J. P. Florens, M. Mouchart, J. P. Raoult, L. Simar, and A. F. M. Smith, eds.). Springer, New York.

VELLEMAN, P. F. and HOAGLIN, D. C. (1981). *Applications, Basics, and Computing of Exploratory Data Analysis.* Duxbury Press, Boston.

WEISBERG, S. (1986). A linear model approach to backcalculation of fish length. *J. Amer. Statist. Assoc.* **81** 922–929.

Comment

Alan Julian Izenman

1. INTRODUCTION

It gives me great pleasure to comment on this paper by Professor Rao. The central issues raised here are choice of a prediction model and assessment of associated prediction errors for growth curve data. Professor Rao has given us a number of different approaches to these problems. I offer a few general comments and some specific comments, mention alternative directions in growth curve modeling and prediction and also make some comments on the mice data used in this paper.

2. SOME GENERAL COMMENTS

For a statistician, context should always play a role in the modeling process. Too often, data are analyzed without regard to the original purpose of their collection. This can be especially true when modeling a growth process where biological reasoning may help in the modeling and subsequent interpretation of results. The first thing I noticed about this paper is that there is no clear description of the three data sets used as illustrations or why they are even interesting for prediction purposes. (Does anyone understand what

Alan Julian Izenman is Associate Professor, Department of Statistics, Temple University, Philadelphia, Pennsylvania 19122.

"the ascending part of the mandible" means as a definition of ramus height? What was actually measured for the dental data?) For the record, the ramus data (Table 3) originally appeared in Elston and Grizzle (1962) and the dental data (Table 4) in Potthoff and Roy (1964). Contrary to the legend of Table 4, the dental data were not mentioned in Grizzle and Allen (1969).

For assessing prediction error and thereby finding a reasonable predictor, Rao follows Stone (1974) and Geisser (1975a) in using a cross-validation (CV) criterion computed via a "leave-one-out" algorithm. The CV criterion is used in this paper to choose an optimal prediction function from a given class of such functions (ranging from linear regression functions to polynomials to factor analytic type functions) by selecting the best subset of available predictor variables. Thus, for linear regression growth models, it reduces to selecting a subset of the immediately previous measurements; for polynomial growth models, the choice involves a trade-off between the degree of the polynomial and the number of immediately previous measurements; and for factor analytical type growth models, the choice is the number of factors. This use of cross-validation is similar in spirit to that used by several authors for optimally choosing a bandwidth in nonparametric density estimation, regression and spline smoothing contexts. However, in this paper, the only insight we get into the CV procedure is its application to three very similar data sets. Because there appears to be nothing special in the way these data sets were treated throughout the paper, Rao could just as well have used a single (or even a simulated) data set to illustrate his points. I think that a more carefully selected set of examples might have been used with greater effect here.

It would be interesting to have had some discussion in this paper of the reliability of the CV method for assessing prediction error in growth models. Recently, work has been carried out regarding the bias of the CV estimate (2.3.5) of the true error rate (2.3.2) associated with the predictions of future observations in a variety of regression situations. Efron (1982, 1983, 1986) and Gong (1986) studied a variety of estimates of the true prediction error rate, including CV, bootstrap and jackknife estimates. Under certain conditions, the bootstrap estimate was found to yield a substantial improvement over the CV estimate. A related paper is Bunke and Droge (1984). Has Professor Rao any comments on or experience of such bias for his growth curve prediction methods?

An interesting theoretical point that was not pursued in this paper is the relationship, if any, for the linear prediction model, of the cross-validation assessment error (CVAE) to the number of immediately previous observations in the subset. There is numerical evidence in Table 5, for example, that the relationship is not quite monotonic; yet, contrary to the usual variable selection results in multiple linear regression, it appears that as a general rule CVAE gets *smaller* when there are fewer predictor variables in the model. Can Professor Rao explain why this should be true? Does it have anything to do with the correlation structure of the predictor variables?

3. THE MICE DATA

The mice data (Table 2) used to illustrate the methods of this paper are presented without description or comment. However, certain features of the data need to be explained here. The data are actually a portion of those given in Williams and Izenman (1981). The latter were obtained by randomly selecting 35 litters of mice from the same generation of a well established control line, and then randomly choosing a single male offspring from each selected litter. Of those 35 mice, 33 were weighed every third day according to the following schedule: 11 mice (Group 1) were weighed at ages $t = 0, 3, 6, 9, 12, 15, 18$ days after birth; 10 mice (Group 2) were weighed at ages $t = 1, 4, 7, 10, 13, 16, 19$ days after birth and 12 mice (Group 3) were weighed at ages $t = 2, 5, 8, 11, 14, 17, 20$ days after birth. The remaining two mice (D1 and D2) were weighed daily from $t = 0$ through $t = 20$ days after birth. Thus, the data in Table 2 consist of the following: the first eleven mice are the twelve Group 3 mice *with the third mouse omitted*, and the last two mice (numbers 12 and 13) are the D1 and D2 weighings, respectively, at those days after birth.

Three minor points should be noted regarding Table 2. First, the column headings are in error by one day. Second, the weight of 0.640 marked by an asterisk is correct; in fact, there are several instances in the complete mice data of declining weight from one time point to the next. Third, there is a transcription error in Table 2: the weight of mouse number 10 on day 3 is incorrect and should be 0.225 rather than 0.255. This error was, unfortunately, carried through all the paper's statistical computations.

A much more interesting situation was created when Rao constructed Table 2 by omitting one of the Group 3 mice. This mouse weighed noticeably less than the remainder of his group, which was probably the reason for his omission; his series of seven weighings was as follows: 0.141, 0.260, 0.472, 0.662, 0.760, 0.885, 0.878. (He was not the lightest mouse in the complete mice data set, however.) This naturally raises the following question: if this mouse had been retained in the data set used in this paper, what effect would his measurement record have had on the prediction results? More generally, we find ourselves interested in the detection and identification of deviating growth records, because all growth records enter into computations for modeling and prediction. A theory of influence functionals for growth curves is, therefore, needed here, just as Martin and Yohai (1986) developed such a theory for time series analysis. A comprehensive survey of the influence function approach may be found in Hampel, Ronchetti, Rousseeuw and Stahel (1986); similar techniques would be appropriate for developing influence functionals for growth curves.

4. SOME SPECIFIC COMMENTS

I have a few specific comments about this paper. First, the decompositions of the multiple squared correlation coefficient of Y_{p+1} on $Y_p, \cdots, Y_1$ for the three data sets appear to be the only evidence in the paper that "the squared correlation between Y_{p+1} and Y_p dominates $\cdots$ indicating that no improvement can be expected by using other measurements, except perhaps Y_{p-1}." This is an interesting result. However, could not some growth curve data be constructed so that these conclusions do not hold?

Second, Rao obtains certain conclusions that do not necessarily follow from the empirical results of this paper. For example, in Section 4.1, the claim is made that "in all the cases studied, the best procedure is to fit a straight line to just the two previous measurements, Y_p, Y_{p-1}, and extrapolate to predict Y_{p+1}." This conclusion was reached from the numerical results in column (3) of Table 7. Yet, assuming the computations as they appear in Table 7 are correct, then for both the ramus and the dental data sets, such a conclusion clearly does not follow. Similarly, in Section 4.2, Rao's statement that "the values in column (4) are smaller than those in column (3)" of Table 7 is not true for the ramus data.

Third, in Section 4, Rao remarks that "models which provide an adequate description of the past observations may not necessarily be suitable for predicting future observations." Perhaps this may be true, but Rao offers no justification (or reference) for such a remark in this paper. What would happen, for example, if any of the four nonlinear regression models listed at the beginning of this section was used for prediction purposes with the three data sets?

5. RELATED WORK

Interest has been shown recently in a nonparametric approach to modeling the dynamics of growth. It has been argued that the traditional parametric approach to fitting models to longitudinal data sometimes fails to account for sufficient relevant structure. Thus, in modeling height measurements over time of children, Gasser, Müller, Kohler, Molinari and Prader (1984) remarked that "the midgrowth spurt—not being part of the parametric models—disappeared from the literature when statistics and computing came into common use in growth studies," and thereby systematically distorted the estimation of onset of the pubertal spurt. This attitude led to nonparametric kernel regression analysis of individual (Gasser, Müller, Kohler, Molinari and Prader, 1984) and average (Hart and Wehrly, 1986) growth curves, and clearly constitutes an important new direction in growth curve modeling with possibilities also for prediction. It should be noted that similar remarks would hold for a maximum penalized likelihood approach to growth curves.

ACKNOWLEDGMENTS

This research was prepared using computer equipment supported in part by National Science Foundation Grant DMS-8504988 and by a Digital Equipment Corporation Research Agreement awarded to the Department of Statistics at Temple University.

ADDITIONAL REFERENCES

BUNKE, O. and DROGE, B. (1984). Bootstrap and cross-validation estimates of the prediction error for linear regression models. *Ann. Statist.* **12** 1400–1424.

EFRON, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans.* SIAM, Philadelphia.

EFRON, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *J. Amer. Statist. Assoc.* **78** 316–331.

EFRON, B. (1986). How biased is the apparent error rate of a prediction rule? *J. Amer. Statist. Assoc.* **81** 461–470.

GASSER, T., MÜLLER, H.-G., KOHLER, W., MOLINARI, L. and PRADER, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.* **12** 210–229.

GONG, G. (1986). Cross-validation, the jackknife, and the bootstrap: Excess error estimation in forward logistic regression. *J. Amer. Statist. Assoc.* **81** 108–113.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

HART, D. H. and WEHRLY, T. E. (1986). Kernel regression estimation using repeated measurements data. *J. Amer. Statist. Assoc.* **81** 1080–1088.

MARTIN, R. D. and YOHAI, V. J. (1986). Influence functions for time series (with discussion). *Ann. Statist.* **14** 781–855.