Comment

Randy Eubank

The author is to be congratulated on his insightful and far-reaching article. This article represents a substantial contribution to the field of nonparametric function estimation due in particular to the many new application areas it will introduce to statisticians who are concerned with developing such methodology. The collection of examples Professor Ramsay presents pose challenging problems, that are also of practical importance, for study by nonparametric data modelers. It will be interesting to see the alternative solutions that are sure to be developed as a result of this article. I am delighted to get one of the first shots at this and will suggest some other possible approaches in Section 2 below.

I found myself in agreement with much of what Professor Ramsay has said. My few points of disagreement are methodological, rather than philosophical, and therefore minor. My primary concerns are related to the problems of knot selection and inference in spline fitting. Here my experiences appear to be almost the opposite of the author's. I will elaborate further on this in the next section.

It seems to me that nonparametric and parametric estimation methods are too often viewed as competitors to one another. There is no reason that these methods cannot or should not be used in tandem. Indeed, it is foolish to do otherwise when conducting exploratory data analyses. I would therefore like to expand on Professor Ramsay's point concerning his example in Section 4.1. He notes that although his nonparametric approach did not lead to new results, it provided us with reassurance concerning a parametric fit. This is an illustration of how nonparametric procedures can provide diagnostic checks concerning the validity of parametric models. Whereas no difficulties were uncovered with the parametric model for the data in question, this need not always be the case. Serious parametric modeling deficiencies can be uncovered by comparison with nonparametric fits. An excellent illustration of this is the growth curve analysis conducted by Gasser, Müller, Köhler, Molinari and Prader (1984).

1. CHOOSING KNOTS, INFERENCE AND RELATED ISSUES

The transformations Professor Ramsay employs are splines of some specified order k with n knots having locations contained in a set t_n. In practice k, n and t_n must all be selected in some fashion. I agree with the author that the choice of k is generally not crucial; with $k = 4$ (i.e., cubic splines) being statisfactory for most purposes (provided the number and locations of the knots have been chosen correctly). However, he also indicates that good choices of n and t_n can be easily made and that the shape of a spline function is robust with respect to these choices. This view is inconsistent with my experiences and those of many others.

To illustrate my point, consider the data in Figure 1 which represent a property, Y, of titanium as a function of heat, X. (See de Boor (1978), page 222 for the actual data.) Three cubic splines have been fitted to the data via least squares, the first uses five uniformly spaced knots whereas the second has five knots that have been selected more carefully. Notice that although both fits have the same order and number of knots, they are not even remotely similar in shape. The spline based on uniformly spaced knots is also a woeful fit being very little (if at all) better than fits obtained using polynomials.

This example illustrates that knot placement can be crucial for both the shape and quality of a spline fit. Some ad hoc rules for good knot placement can be found in Wold (1974). They are motivated by the same considerations which led the author to propose his two guidelines for this purpose in his Section 3.

A more objective knot selection can be accomplished by optimizing the estimation criterion of interest with respect to both the knot set t_n and the vector a of spline basis coefficients. For example, both t_n and a could have been chosen to maximize the sample likelihood in the examples discussed in Sections 4.1 and 4.2. This idea is by no means new and has even been suggested by the author (Winsberg and Ramsay, 1980). The second fit in Figure 1 was obtained by optimizing the knot locations and therefore illustrates the gains to be realized from optimal knot selection.

Once the knot locations have been optimized, n can be chosen using various model selection techniques. A criterion such as that of Akaike (1974) would be easy to use with likelihood-based fitting methods.

Unfortunately, I am not optimistic about the practicality of methods for optimal knot selection. In the context of least squares estimation this is a very poorly behaved nonlinear optimization problem (Jupp, 1978). It is unlikely that matters will improve for more general likelihood-based methods.

An alternative approach to selecting knots is to

Randy Eubank is Professor, Department of Statistics, Texas A&M University, College Station, Texas 77843-3143.