Comment

Leo Breiman

It is a pleasure to comment on papers where one finds much of merit together with major issues on which one can contend. This is the case with Ramsay's work. I have become impressed with the usefulness of splines over the past few years. They have nice theoretical properties as well as being a good practical tool in data analysis. Their use in statistics is not widespread, so I welcome this article because it may lead to increased interest and applications.

Now, as to my comments:

Ramsay's example on ACE will probably strike terror into the hearts of hundreds of contented ACE users. His ACE runs on the gasoline consumption data purport to show an extreme dependence on the order in which variables are entered into ACE.

We have used the same data (kindly published in the article) and are completely unable to replicate his results. We thought he might be using the monotone option in ACE, so we ran it this way. Then we ran it without the monotone option. We also ran ACE on the data with a 40% fixed window size, which (see below) is roughly equivalent to using one interior knot.

In all three cases we changed the order in which variables were entered and compared the results. These are given in Figure 1. The differences, due to the change in order of entering variables, is miniscule. I don't understand how Ramsay got the results he did, but my best guess is that some mistake was made in doing his ACE runs.

The ACE algorithm has been circulated and used far and wide over the last few years. In practice, it has proven a generally robust and illuminating data analysis tool.

I am very wary of the assumption of monotonicity. I was, for instance, against the inclusion of monotizing transformations as an option in the ACE algorithm. But my co-author, Jerome Friedman, argued me into it asserting that, whatever the reason, lots of statisticians like monotone transformations.

No one has decreed that phenomena in nature are inherently monotone. By restricting yourself to monotone transformations, you risk missing some important discoveries. For example, in the ACE paper (Breiman and Friedman, 1985) we give an example of ACE runs on an air pollution data set where one of the predictor variables is the pressure difference at two meteorological stations. One is in the Los Angeles basin and the other about 30 miles to the north.

Leo Breiman is Professor, Department of Statistics, University of California at Berkeley, Berkeley, California 94720.

In previous analyses of this data, many monotone transformations had been tried on this variable, and it always wound up as having very little predictive power. ACE produced a transformation on this variable that resembled $-|x|$. In this transformed mode, it became a strongly predictive variable. In retrospect, the reason is obvious. Any kind of pressure difference, either positive or negative, encourages a moving air mass, and reduces pollution in the Basin.

I only know of infrequent cases in which I would insist on monotone transformations. Finding nonmonotonicity can lead to interesting scientific discoveries. If the appropriate transformation is monotone, then the fitted spline functions (or ACE transformations) will produce close to a monotonic transformation. So it is hard to see what there is to gain in the imposition of monotonicity.

Ramsay claims that in practical applications a very small number of interior knots are sufficient. He cites two interior knots as being usually good enough. This is contrary to my experience.

The number of knots in a spline fit can be viewed as equivalent to the window size in a smoother. In fact, because spline fitting is a linear operation, one can compute the shape of the windows produced by spline fitting.

For one interior knot in a quadratic spline the windows are very broad, being equivalent (under a sensible definition) to a fixed window size smoother that uses about 35% of the data. For two interior knots, the equivalent window size is 29% of the data. For six interior knots the equivalent window size is 15%, and 10 interior knots drop it to 10% of the data.

Now a smoother with a 30% window will capture only the broadest gross features of the data. It will oversmooth the peaks and valleys and wipe out any salient fine features of the data. Unless one is willing to live with these oversmoothing characteristics, two knots are not sufficient.

The setting of an appropriate window size for the data is a critical and complex problem. In ACE it is done by using a smoother that selects a locally optimal window size ranging from 7% of the data to 30%. In smoothing splines it is done by choosing the optimal size of the penalty parameter through cross-validation. A large literature has developed in the theory and practice of choosing a good kernel sharpness