*Statistics: 8th Annual Symposium on the Interface*, 413–418. Health Sciences Computer Facility, UCLA.
STONE, C. J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.* **5** 595–645.
STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.
STONE, C. J. and KOO, C.-Y. (1986b). *Function Estimates.* AMS Contemporary Math. Ser., Amer. Math. Soc., Providence, R. I.

# Comment

**Peter McCullagh**

Hastie and Tibshirani are to be congratulated for presenting the theory and methodology of generalized additive models in a form that keeps incidental mathematical details at an acceptably low level. I have little to add and my single comment is therefore brief.

The whole thrust of the authors' development seems to be based implicitly on the following assumption, here reduced to the bare essentials: zero interaction is fundamentally more plausible than componentwise linearity in the covariates. Has there been any attempt to justify this point of view, either philosophically or empirically by examining a large number of examples or by any other means? A closely related question concerning statistical strategy is the following: at what stage of analysis does the assumption of zero interaction come under scrutiny?

*Peter McCullagh is Professor of Statistics, Department of Statistics, The University of Chicago, 5734 University Avenue, Chicago, Illinois 60637.*

# Rejoinder

**Trevor Hastie and Robert Tibshirani**

## 1. THE GENERAL PROBLEM

In Section 5 of the paper, we motivated the local scoring and local likelihood estimation procedures as empirical methods for maximizing $E(l(\eta(X),\ Y))$. In the two procedures, the maximization problem is approached in different ways. In the *local likelihood* method, an estimate of $E((l(\eta(X),\ Y)\mid X = x))$ is constructed (for each $x$) and this has the form $(1/k_n)\sum_{j\in N_i} l(\eta(x_j),\ y_j)$ given in (26) of the paper. As Brillinger notes (his Section 2), one can generalize this and hence include robust estimates and many others.

On the other hand, the local scoring procedure maximizes $E(l(\eta(X),\ Y))$ by estimating the quantities in the update expressions (22) and (36). Note, however, that this procedure is not expressible as a maximation of the kind that Brillinger describes, i.e., a maximization of a function of the form $\sum_i \rho(Y_i\mid\hat{\eta})W_{ni}(X)$. However, it is possible to write down a finite sample justification of local scoring (to answer a question of Brillinger's) based on the notion of penalized likelihood. This justification applies only in the special case in which the local scoring algorithm uses linear smoothers. Recall that a linear smoother is one for which the result of smoothing a vector $\mathbf{z}$ can be written simply as $\hat{\mathbf{z}} = S\mathbf{z}$, for some matrix $S$, called a "smoother matrix." Now suppose we have data $(y_1,\ x_{11},\ x_{12},\ \cdots,\ x_{1p}),\ \cdots,\ (y_n,\ x_{n1},\ x_{n2},\ \cdots,\ x_{np})$ and let $S_j$ be the smoother matrix for the $j$th variable. Let $\mathbf{s}_j = (s_1(x_{1j}),\ s_2(x_{2j}),\ \cdots,\ s_n(x_{nj}))^t$, $j = 1,\ 2,\ \cdots,\ p$ and consider the following problem. Find $\mathbf{s}_1,\ \mathbf{s}_2,\ \cdots,\ \mathbf{s}_p$ to maximize

$$l(\boldsymbol{\eta}) - \frac{1}{2}\sum_1^p \mathbf{s}_j^t(S_j^- - I)\mathbf{s}_j \qquad (1)$$

where $\boldsymbol{\eta} = \alpha + \sum_1^p \mathbf{s}_j$ and $S_j^-$ is a generalized inverse of $S_j$. Then it is easy to show that the local scoring procedure is a Fisher scoring step for maximizing (1) (see Hastie and Tibshirani, 1986a, for details). Now a typical smoother matrix is close to symmetric, has eigenvectors that are close to polynomials, and has eigenvalues that tend to decrease with increasing order of the eigenvector. Hence, the penalty term in (1) puts greater penalty on the higher order polynomial components of each $\mathbf{s}_j$. There is also a close tie here to smoothing splines. If we start with a penalty of the form $\sum_1^p \lambda_j \mathbf{s}_j^t K_j \mathbf{s}_j$, where $K_j$ is an appropriate quadratic penalty matrix, we derive a local scoring procedure that uses cubic spline smoothers. Hence, there is close relation of local scoring to the work of O'Sullivan, Yandell, and Raynor (1986), Green (1985), and Green and Yandell (1985). These authors consider a