I suppose the message here is that no single adaptive regression technique can perform uniformly best on all examples, which echoes the point made by Professor Friedman in Section 2.

REFERENCES

[1] CHOI, H. S., HAYNOR, D. R. and KIM, Y. (1989). Multivariate tissue classification of MRI images for 3-D volume reconstruction—a statistical approach. *SPIE Medical Imaging III: Image Processing* **1092** 183–193.

[2] O'SULLIVAN, F. (1990). Mixture models for multi-channel image data. Technical report, Dept. Statist., Univ. Washington.

[3] SCHUMAKER, L. L. (1976). Fitting surfaces to scattered data. In *Approximation Theory II* (G. G. Lorentz, C. K. Chui and L. L. Schumaker, eds.) 203–268. Academic, New York.

[4] SEBER, G. A. F. (1976). *Linear Regression Analysis*. Wiley, New York.

[5] SMITH, M. O., ADAMS, J. B. and JOHNSON, P. (1986). Spectral mixture modelling, a new analysis of rock and soil types at the viking lander 1 site. *J. Geophys. Res.* **91** 8098–8112.

DEPARTMENT OF BIOSTATISTICS AND STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

ART OWEN

Stanford University

I like MARS. It looks like a good tool for pulling out the most useful parts of large interaction spaces. Most of my comments are directed at accounting issues: How many degrees of freedom are used in knot selection? How can the cost be lowered? At the end, there are some comments on how one might apply MARS to models for which fast updating is not available.

My main interest in MARS stems from work in computer experiments. In these applications, smooth functions of fairly high complexity are evaluated over high dimensional domains with no sampling error. I plan to use MARS on such functions evaluated over Latin hypercube designs [McKay, Conover and Beckman (1979)]. Some theory for linear modeling of nonrandom responses over such designs is given in Owen (1990).

When there is no noise, one expects that a larger number of knots might be warranted. It then becomes worthwhile to lower the price of a knot somehow.

Degrees of freedom in broken line regression. Consider the broken line regression model

$$Y_i = b_0 + b_1 t_i + \beta(t_i - \theta)_+ + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $t_1 \le t_2 \le \cdots \le t_n$ are nonrandom with $\sum t_i = 0$ and $\sum t_i^2 = n\sigma_t^2$, ε_i are independent $N(0, 1)$ and b_0, b_1, β and θ are parameters. Taking $\beta = 0$ in (1) yields a one-segment model. Taking $\beta \neq 0$ and $t_1 < \theta < t_n$ yields a two-segment model. This model has been studied by Feder (1967), Hinkley (1969),

Davies (1977, 1987) and Knowles and Siegmund (1989). All these authors point out that nonstandard asymptotics apply when $\beta = 0$. In particular, estimation of β and θ uses up more than 2 degrees of freedom in that case. Hinkley (1969) reports that approximately 3 degrees of freedom are used, based on a simulation in which the potential knots were uniformly spaced. When $\beta \neq 0$, the standard asymptotics, that is 2 degrees of freedom, are relevant [Feder (1967), Part 2].

This suggests that when the evidence that $\beta \neq 0$ is extremely strong, that fewer degrees of freedom should be charged than when the evidence is borderline. Of course, for small $\beta \neq 0$ and finite n, the nonstandard asymptote might be the more accurate one.

Following Davies (1987), we consider testing whether the line breaks at θ via

$$Z(\theta) = V^{-1/2}(\theta) \sum_{i=1}^{n} \hat{\varepsilon}_i (t_i - \theta)_+, \tag{2}$$

where

$$\hat{\varepsilon}_i = Y_i - \bar{Y} - t_i \overline{tY} / \sigma_t^2,$$

$$V(\theta) = R_{02} - \frac{1}{n} R_{01}^2 - \frac{1}{n\sigma_t^2} R_{11}^2$$

and

$$R_{jk} = R_{jk}(\theta) = \sum_{i=1}^{n} t_i^j (t_i - \theta)_+^k. \tag{3}$$

The null distribution of $Z(\theta)$ is $N(0, 1)$ for $t_2 < \theta < t_{n-1}$. To test whether the regression line breaks at all, we consider the supremum of $Z(\theta)$ or of $|Z(\theta)|$ over an interval. The null covariance of $Z(\theta)$ and $Z(\phi)$ is

$$\rho(\theta, \phi) = V^{-1/2}(\theta) V^{-1/2}(\phi) \times \left[R_{01,01}(\theta, \phi) - \frac{1}{n} R_{01}(\theta) R_{01}(\phi) - \frac{1}{n\sigma_t^2} R_{11}(\theta) R_{11}(\phi) \right], \tag{4}$$

where

$$R_{01,01}(\theta, \phi) = \sum_{i=1}^{n} (t_i - \theta)_+ (t_i - \phi)_+. \tag{5}$$

Bounds for

$$P\left(\sup_{\theta_0 \leq \theta \leq \theta_1} Z(\theta) > c \right)$$

may be derived from the expected number of upcrossings of c by Z. For

example,

$$(6)\qquad P\Big(\sup_{\theta_0\le\theta\le\theta_1} Z(\theta) > c\Big) \le \Phi(-c) + \frac{1}{2\pi}\exp\Big(-\frac{1}{2}c^2\Big)\int_{\theta_0}^{\theta_1}\rho_{11}^{1/2}(\theta)\,d\theta,$$

where Φ is the standard normal distribution function and

$$(7)\qquad \rho_{11}(\theta) = \frac{\partial^2}{\partial\theta\,\partial\phi}\rho(\theta,\phi)\Big|_{\phi=\theta}.$$

Davies (1977) gives (6) as his (3.7), except that he uses

$$(7')\qquad \rho_{11}(\theta) = -\frac{\partial^2}{\partial\phi^2}\rho(\phi,\theta)\Big|_{\phi=\theta}.$$

Definitions (7) and (7′) are equivalent when both derivatives exist (since the process has constant variance). For broken line regression neither exists but the problem is easily handled: Replace $(t_i - \theta)_+$ by $q_\varepsilon(t_i - \theta)$, where

$$q_\varepsilon(x) = \begin{cases} 0, & x \le -\varepsilon, \\ \dfrac{\varepsilon}{4} + \dfrac{x}{2} + \dfrac{x^2}{4\varepsilon}, & |x| \le \varepsilon, \\ x, & x \ge \varepsilon \end{cases}$$

for small positive ε in (1), (2), (3), (5). The function θ_ε is Friedman's piecewise cubic approximation to x_+ which in this case is piecewise quadratic. Now $\partial^2\rho(\theta,\phi)/\partial\theta\,\partial\phi$ exists and is continuous so (6) follows from (3.3) of Leadbetter (1972). The right side of (6) is continuous in $\varepsilon > 0$ when this substitution is made. If we take the limit as $\varepsilon \downarrow 0$, we get

$$\rho_{11}(\theta) = -\frac{1}{4}\left(\frac{V'(\theta)}{V(\theta)}\right)^2 + V^{-1}(\theta)\left[R_{00} - \frac{1}{n}R_{00}^2 - \frac{1}{n\sigma_t^2}R_{10}^2\right],$$

where

$$V' = \frac{d}{d\theta}V(\theta) = 2R_{01} + \frac{2}{n}R_{00}R_{01} + \frac{2}{n\sigma_t^2}R_{10}R_{11},$$

with

$$(8)\qquad R_{j0} = R_{j0}(\theta) = \sum_{t_i > \theta} t_i^j.$$

One could also take the sum in (8) over $t_i \ge \theta$; it makes no difference in (6) because the integrand is only affected at finitely many places.

Since both positive and negative β are of interest, we write

$$
\begin{aligned}
P\Big(\sup_{\theta_0 \le \theta \le \theta_1} |Z| > c\Big) &= P\Big(\sup_{\theta_0 \le \theta \le \theta_1} Z > c\Big) \\
&\quad \times \left[2 - P\Big(\inf_{\theta_0 \le \theta \le \theta_1} Z < -c \,\Big|\, \sup_{\theta_0 \le \theta \le \theta_1} Z > c\Big)\right] \\
&\le 2P\Big(\sup_{\theta_0 \le \theta \le \theta_1} Z > c\Big) \qquad (9) \\
&= P\big(\chi^2_{(1)} > c^2\big) + \frac{1}{\pi} \exp\Big(-\frac{1}{2}c^2\Big) \int_{\theta_0}^{\theta_1} \rho_{11}^{1/2}(\theta)\, d\theta \\
&= P\big(\chi^2_{(1)} > c^2\big) + \frac{1}{\pi} \int_{\theta_0}^{\theta_1} \rho_{11}^{1/2}(\theta)\, d\theta\, P\big(\chi^2_{(2)} > c^2\big).
\end{aligned}
$$

For a given set of t_i, ρ_{11} can be computed and the integral in (9) can then be numerically evaluated. Updating formulae are easy to derive for the R_{jk}. Assume that $\Sigma t_i = 0$ and $\Sigma t_i^2 = n$, that is, $\sigma_t^2 = 1$. The integral of $\rho_{11}^{1/2}$ is invariant when a nonsingular location scale transformation is applied to the t_i and to the limits of integration. Suppose we want $\rho_{11}(\theta_i)$ for $t_2 < \theta_1 \le \theta_2 \le \cdots \le \theta_m < t_{n-1}$. Let $s_1 \le s_2 \le \cdots \le s_{n+m}$ be obtained by pooling and sorting the data points t_i and the evaluation points θ_i. Let D_i be 1 if s_i arose as a data point and 0 if s_i arose as an evaluation point. With this set up, $R_{00}(s_1) = n - 1$, $R_{01}(s_1) = -nt_1$, $R_{10}(s_1) = -t_1$ and $R_{11}(s_1) = n$. For $2 \le i \le n + m$, $R_{00}(s_i) = R_{00}(s_{i-1}) - D_i$, $R_{01}(s_i) = R_{01}(s_{i-1}) - (s_i - s_{i-1})R_{00}(s_{i-1})$, $R_{10}(s_i) = R_{10}(s_{i-1}) - D_i s_i$ and $R_{11}(s_i) = R_{11}(s_{i-1}) - (s_i - s_{i-1})R_{10}(s_{i-1})$. Now ρ_{11} can be computed on noting that $R_{02}(\theta) = R_{11}(\theta) - \theta R_{01}(\theta)$ and the values corresponding to evaluation points can be extracted. At an evaluation point θ_i that coincides with a data point t_j, $\rho_{11}(\theta_i)$ will depend on the order in which the two points appear in the list of s_i.

For uniformly distributed t_i, we can find a simple approximation to (9). Let $t_i = (i - 0.5)/n - 0.5$ so $\sigma_t^2 = (n^2 - 1)/(12n^2) \doteq \frac{1}{12}$ and approximate R_{jk} by

$$\tilde{R}_{jk}(\theta) = n \int_{\theta}^{1/2} u^j (u - \theta)^k\, du$$

and similarly for $\tilde{R}_{01,01}$. Use $\tilde{V}$ and $\tilde{\rho}$ to denote the corresponding changes to V and ρ. Some calculus gives

$$\tilde{V}(\theta) = \frac{n}{3}\left(\frac{1}{4} - \theta^2\right)^3$$

and

$$\tilde{\rho}_{11} = \frac{3}{4}\left(\frac{1}{4} - \theta^2\right)^{-2}$$

and so

$$P\Big(\sup_{\theta_0\le\theta\le\theta_1}|Z|>c\Big)$$
$$\le P\big(\chi^2_{(1)}>c^2\big)+\frac{\sqrt{3}}{2\pi}\int_{\theta_0}^{\theta_1}\frac{d\theta}{\left(\frac14-\theta^2\right)^{1/2}}P\big(\chi^2_{(2)}>c^2\big)$$
$$=P\big(\chi^2_{(1)}>c^2\big)+\frac{\sqrt{3}}{2\pi}\left(\log\frac{\frac12+\theta_1}{\frac12-\theta_1}-\log\frac{\frac12+\theta_0}{\frac12-\theta_0}\right)P\big(\chi^2_{(2)}>c^2\big).$$

Therefore for $0<\varepsilon<\frac12$,

$$(10)\quad P\Big(\sup_{|\theta|\le(1/2)-\varepsilon}|Z|>c\Big)\le P\big(\chi^2_{(1)}>c^2\big)+\frac{\sqrt{3}}{\pi}\log\left(\frac{1-\varepsilon}{\varepsilon}\right)P\big(\chi^2_{(2)}>c^2\big).$$

Knowles and Siegmund [(1989), Section 4] obtain (10) for uniformly spaced t_i and large n using the Hotelling–Naiman volume of tubes approach. For large n, our c corresponds to $n^{1/2}w$ in their notation. One substitutes their equation (10) in their (6) and letting $n\to\infty$, one gets the one-tailed (i.e., sup Z) version of (10).

Figure 1 compares the bound in (10) to tail probabilities from chi-squared distributions on degrees of freedom ranging from 1 through 4 by steps of 0.5.

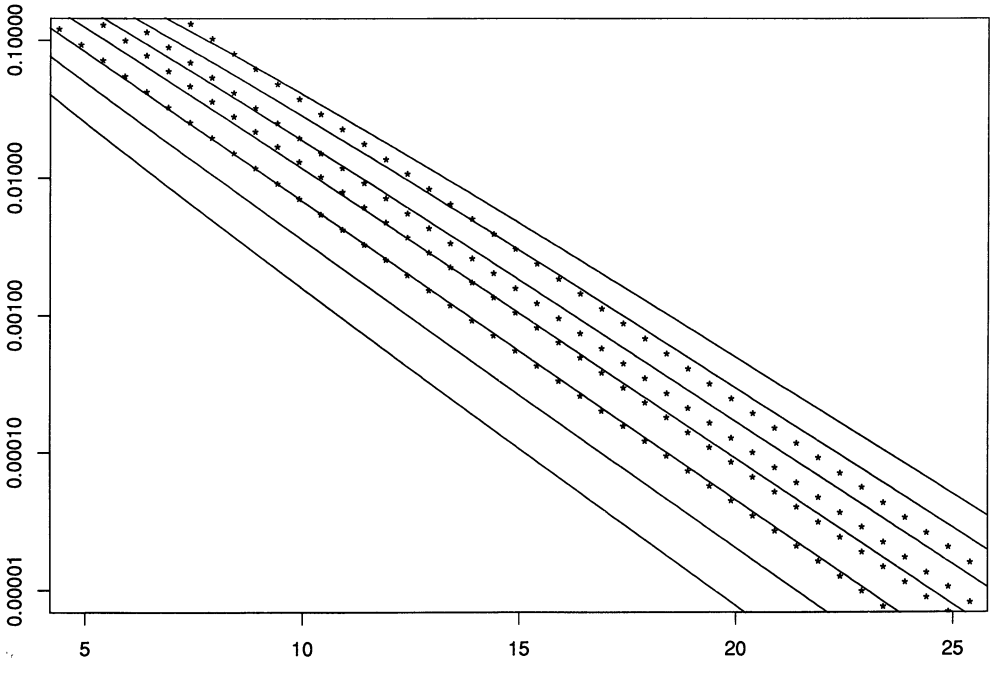

FIG. 1. *Tail probabilities: The solid lines give (right) tail probabilities from chi-squared distributions on degrees of freedom* 1 *through* 4 *in steps of* $\frac12$. *The asterisks plot tail probabilities from formula* (10) *using* ε = 0.2, 0.05, 0.01, 0.0001.

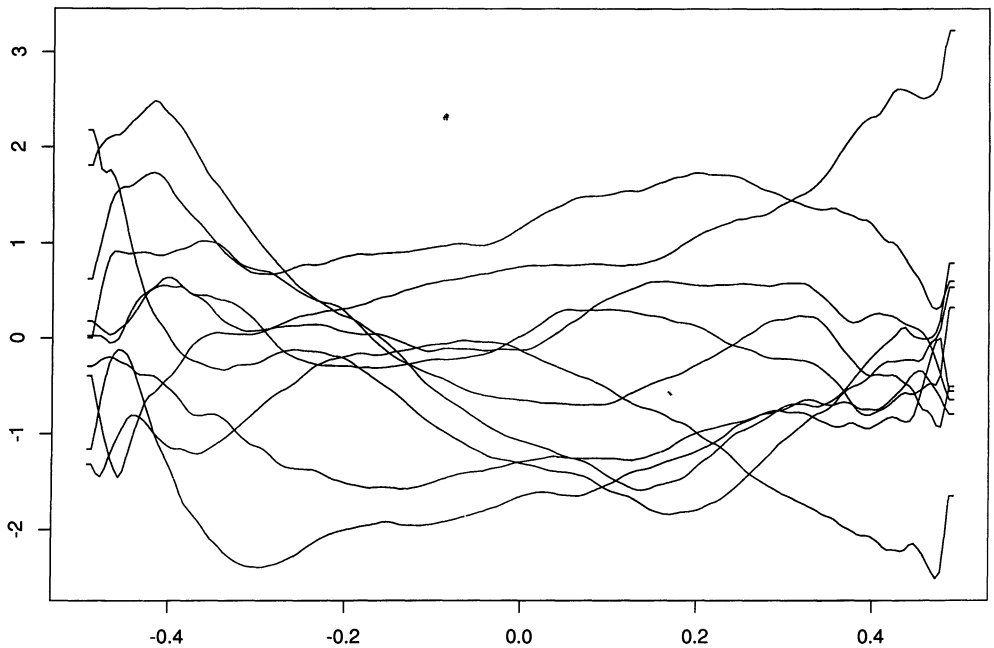

FIG. 2. *Realizations of* $Z(\theta)$: *Shown are* 10 *realizations of the Gaussian process with mean* 0, *constant unit variance and correlation* $\tilde{\rho}$, *a continuous approximation to* ρ *of formula* (4).

Chi-squared tail probabilities are plotted by lines, probabilities based on (10) are plotted as asterisks. For Figure 1, $\varepsilon = 0.2$, 0.05, 0.01 and 0.0001 are used.

The points for $\varepsilon = 0.01$ fall close to the curve for $\chi^2_{(3)}$. So searching the central 98% of the range for θ uses roughly 3 degrees of freedom under the null hypothesis. Hinkley (1969) notes from simulations that roughly 3 degrees of freedom are used. The relative error in these probabilities is less than 0.1 when the value of the bound (10) is between 0.05 and 0.005. While the tail probability in (10) is not in the chi-squared family, it would appear that if one were to approximate it in the chi-squared family for purposes of model selection, that 3 degrees of freedom would be a reasonable choice. Using 3 degrees of freedom might even be a little conservative since it would be common to search over less than the central 98% of the range.

The choice $\varepsilon = 0.2$ corresponds to searching the central 60% of the range and uses approximately 2 degrees of freedom under the null hypothesis. It also uses (asymptotically) 2 degrees of freedom under the alternative $\beta \neq 0$, $|\theta| < 0.3$. So it might be reasonable to make all splits in the central 60% of whatever range is being searched and charge 2 degrees of freedom.

Figure 2 shows 10 realizations of the Gaussian process $Z(\theta)$ on $(-\frac{1}{2}, \frac{1}{2})$ with covariance given by $\tilde{\rho}$. The mean is 0 and the variance is 1 over the whole range. The process turns more rapidly (correlations are smaller) for large $|\theta|$, so more upcrossings occur near the edges.

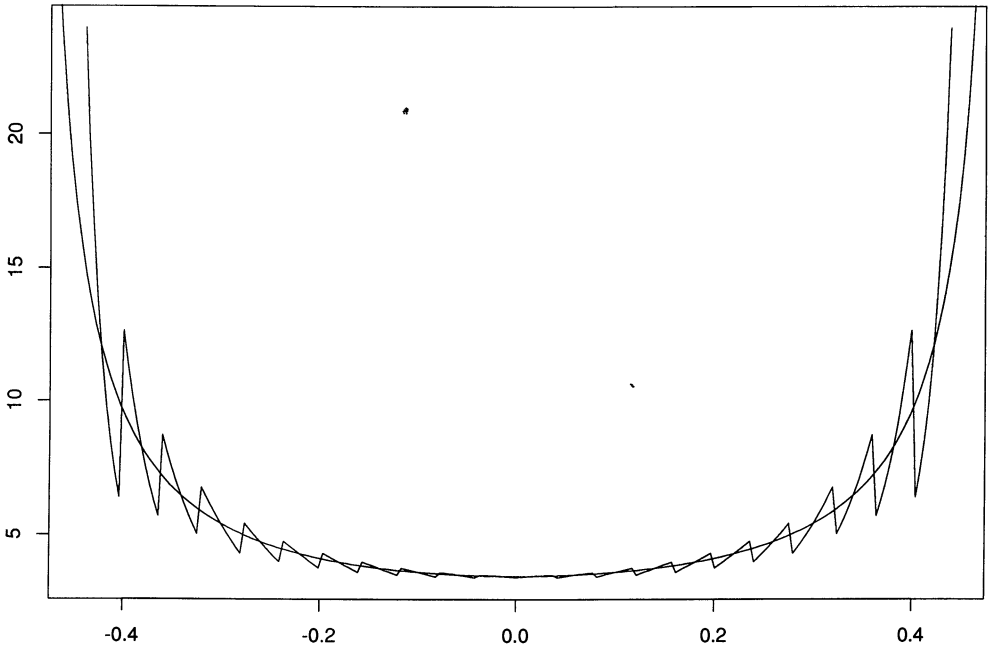

FIG. 3. *Comparison of* $\tilde{\rho}_{11}^{1/2}$ *and* $\rho_{11}^{1/2}$: *The jagged curve is* $\rho_{11}^{1/2}$ *from formula* (4) *with* $n = 25$. *The smooth curve is the continuous approximation* $\tilde{\rho}_{11}^{1/2}$.

Figure 3 compares $\tilde{\rho}_{11}^{1/2}$ and $\rho_{11}^{1/2}$, where the latter is computed for $n = 25$. The steep parts of the jagged curve are meant to be vertical. Since (10) only uses integrals of $\rho_{11}^{1/2}$, the continuous approximation using $\tilde{\rho}$ should be reasonably accurate, especially if each limit of integration is near the midpoint of two consecutive t_i's.

Now suppose that t_i is the $(i - 0.5)/n$ quantile of the distribution with density

$$f(x) = e^{-x-1}, \qquad x \geq -1. \tag{11}$$

This is the unit exponential distribution shifted left one unit so as to have mean zero. For this distribution, $\tilde{R}_{00} = \tilde{R}_{01} = e^{-1-\theta}$, $\tilde{R}_{10} = (\theta + 1)e^{-1-\theta}$, $\tilde{R}_{02} = 2e^{-1-\theta}$ and $\tilde{R}_{11} = (2 + \theta)e^{-1-\theta}$ for $\theta \geq -1$. The function $\tilde{\rho}$ tends to $\frac{1}{4}$ as $\theta \to \infty$ and $\tilde{\rho}$ is asymptotic to $(\frac{3}{4})(1 + \theta)^{-2}$ as $\theta \to -1$. So searching for a breakpoint in an interval of given length should cost more if that interval is near the left edge of the predictor space than if it is near the right edge. Perhaps this is to be expected because there tend to be more points t_i per unit length at the left end. Figure 4 plots $\tilde{\rho}_{11}^{1/2}(F^{-1}(u))/f(F^{-1}(u))$ versus u, where f is the density in (11) and F is the corresponding distribution function. Asterisks are used for $\tilde{\rho}_{11}^{1/2}$ taken from the exponentially distributed t_i and the smooth curve is for $\tilde{\rho}_{11}^{1/2}$ taken from the uniform distribution. If one decides to search over the range between two sample quantiles, then the null probability

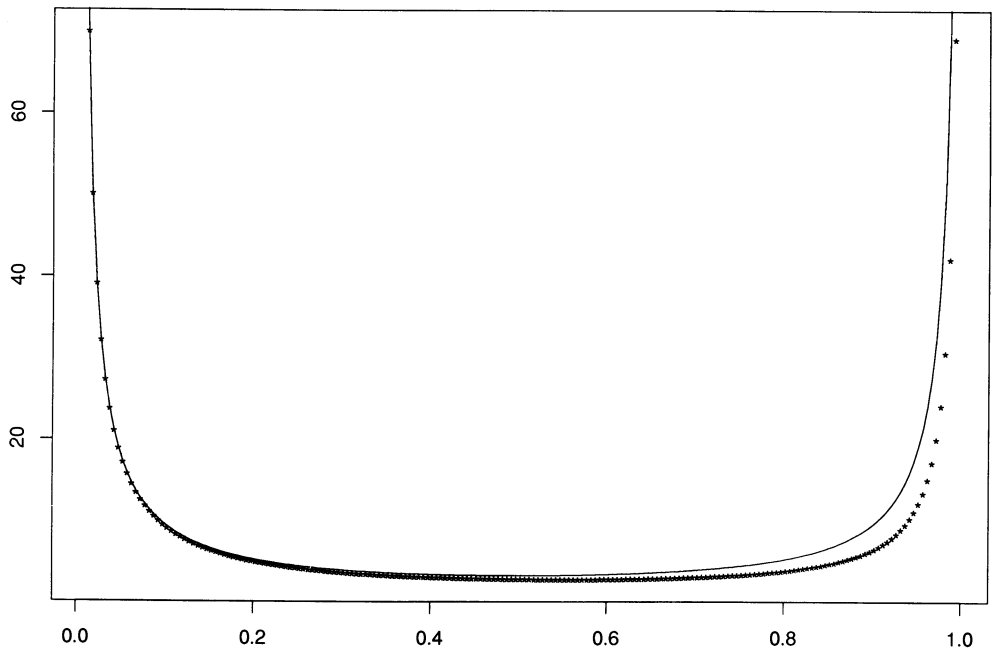

FIG. 4. *Comparison of uniform and exponential designs: The smooth curve is the continuous approximation* $\tilde{\rho}_{11}^{1/2}(\rho)$ *assuming a uniform distribution of design points. The asterisks are* $\tilde{\rho}_{11}^{1/2}(\theta)/f(\theta)$, *where* $\tilde{\rho}$ *is the continuous approximation to* ρ *for design points from the exponential distribution f given by equation* (11). *The x axis is* $F^{-1}(\theta)$, *where F is the appropriate distribution (uniform or exponential).*

of an upcrossing is very close under an exponential design to what it is under a uniform design. Bounds like those in (10) based on the uniform design would be conservative for an exponential design, since the asterisks lie below the curve in Figure 4.

Equation (1) is a model for the problem of deciding if and where to bend a line. If a variable has not yet entered one might consider

$$Y_i = b_0 + \beta(t_i - \theta)_+ + \varepsilon_i \tag{12a}$$

or

$$Y_i = b_0 + \beta(\theta - t_i)_+ + \varepsilon_i. \tag{12b}$$

Note that (12a) at $\theta = t_1$ and (12b) at $\theta = t_n$ are the same (affine) model and that (12a) at $\theta = t_n$ and (12b) at $\theta = t_1$ are both the constant model.

For the model in (12a), we take

$$Z(\theta) = V^{-1/2}(\theta) \sum_{i=1}^{n} \hat{\varepsilon}_i (t_i - \theta)_+,$$

where $\hat{\varepsilon}_i = Y_i - \bar{Y} = \varepsilon_i - \bar{\varepsilon}$ and $V(\theta) = R_{02} - R_{01}^2/n$. The correlation is

$$\rho(\theta, \phi) = V^{-1/2}(\theta)V^{-1/2}(\phi)\left[R_{01,01}(\theta, \phi) - \frac{1}{n}R_{01}(\theta)R_{01}(\phi)\right],$$

with

$$\rho_{11}(\theta) = V^{-1}(\theta)\left[R_{00} - \frac{1}{n}R_{00}^2\right] - \frac{1}{4}\left(\frac{V'(\theta)}{V(\theta)}\right)^2.$$

For a uniform spacing of t_i, one finds

$$\tilde{\rho}_{11}^{1/2}(\theta) = \frac{12^{1/2}(1-\lambda)^{1/2}}{\lambda(4-3\lambda)},$$

where $\lambda = \lambda(\theta) = \frac{1}{2} - \theta$.

If $\beta = 0$, then

$$P\left(\sup_{-(1/2)\le\theta\le\theta_1} |Z(\theta)| > c\right) \le P\left(\chi_{(1)}^2 > c^2\right) + \frac{1}{\pi}\int_{-1/2}^{\theta_1} \tilde{\rho}_{11}^{1/2}(\theta)\, d\theta\, P\left(\chi_{(2)}^2 > c^2\right).$$

Since the process Z can be "glued onto" another one for testing $\beta \neq 0$ in (12b), the null probability of splitting the constant model is bounded by

$$P\left(\chi_{(1)}^2 > c^2\right) + \frac{2}{\pi}\int_{-1/2}^{\theta_1} \tilde{\rho}_{11}^{1/2}(\theta)\, d\theta\, P\left(\chi_{(n)}^2 > c^2\right). \tag{13}$$

For searching the central 98% of the range, the coefficient of $P(\chi_{(2)}^2 > c^2)$ in (10) is approximately 2.53. If one considers model (12a) over all but the rightmost 1% of the range and (12b) over all but the leftmost 1% of the range, the corresponding coefficient is approximately 2.64. So the degrees of freedom used up in deciding whether to split a constant regression are much the same as those used in splitting a linear regression.

In the backward stepwise part (Algorithm 3), how many degrees of freedom should be charged when a knot caused two regressors to be added to the model and one of them gets dropped? I would guess from the analysis of (12a) and (12b) that the full charge for the knot should be assessed, but from Friedman's talk at Interface '90 it seems that half the charge for the knot is assessed.

An alternative to restricting the search to a central subinterval, such as the central 60%, is to search the whole interval, but apply a penalty that increases as the potential knot location nears the end of the range. Davies [(1977), equation 3.3] quotes an upcrossing bound for $P(\sup_\theta Z(\theta) - c(\theta) > 0)$ for continuously differentiable $c(\theta)$. For the process described by the continuous approximation to the uniform design case, the upcrossing bound is especially simple when $c(\theta) = A + B\log((0.5 + |\theta|)/(0.5 - |\theta|))$, where $A, B > 0$. (This c has a cusp, but the bound should still be applicable.) One finds that

$$P\left(\sup_\theta Z(\theta) - c(\theta) > 0\right) \le 2\left(\Phi(B') - \tfrac{1}{2} + \varphi(B')/B'\right)P\left(\chi_{(1)}^2 > A^2\right), \tag{14}$$

where $B' = 2B/3^{1/2}$ and φ is the standard normal density.

Conclusions. It appears that Hinkley's (1969) heuristic of charging 3 degrees of freedom for adding a line segment is reasonably accurate in a variety of settings.

By restricting the search to a subinterval it may be possible to reduce the cost of breaking a line to 2 degrees of freedom. A smoothly varying preference for central splits based on (14) could also be used to lower the cost of knot selection.

The calculations in the preceding section are most relevant to splitting the constant function B_0. When splitting another basis function B_m along variable T, perhaps

$$Y_i = B_m(X_i)(b_0 + b_1 t_i + \beta(t_i - \theta)_+) + \varepsilon_i, \quad i = 1, \ldots, n,$$

should replace (1), or a similar change should be made to (12a) and (12b) depending on context.

Finally, I would like to address Friedman's comments on updating formulae in Section 5.4. When searching a variable for a split point, it may not be necessary to consider every value. The test statistic $Z(\theta)$ should tend to have very smooth sample paths. In the smoothly approximated uniform design case

$$\min_{|\theta| \le 0.4} \max_{\phi \in \{0, \pm 0.2, \pm 0.35\}} \tilde{\rho}(\theta, \phi) \ge 0.94.$$

That is the central 80% of the range can be effectively scanned by considering only 5 points. Note that the realizations in Figure 2 tend to be quite smooth in the middle. I would expect the true underlying function would also make a smooth contribution to $Z(\theta)$, perhaps smoother than that of the noise, though I also expect pathological cases are possible. So after five evaluations one should have a good idea where the maximum is and whether it is worth including in the model. Then one could spend a few more evaluations on a more local search, or wait until the backward stepwise algorithm has finished to refine the search. When a knot is put at 0.2, the next step would involve looking at five places between -0.5 and 0.2 and at five places between 0.2 and 0.5. If this works it should be possible to extend MARS to robust regressions, generalized linear models and the proportional hazards model. Davies (1977, Section 5) has some suggestions on how to perform various tasks on representative points θ, and on picking those representative points.

Acknowledgments. I would like to thank Iain Johnstone and David Siegmund for helpful discussions. In particular Professor Johnstone showed me some unpublished work of his applying the Hotelling–Naiman volume of tubes methodology to broken line regression and including a plot similar to Figure 1.

REFERENCES

DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247–254.

DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43.

FEDER, P. I. (1967). On the likelihood ratio statistic with applications to broken line regression. Ph.D. dissertation, Dept. Statist., Stanford Univ.

HINKLEY, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56** 495–504.

KNOWLES, M. D. and SIEGMUND, D. O. (1989). On Hotelling's approach to testing for a nonlinear parameter in regression. *Internat. Statist. Rev.* **57** 205–220.

LEADBETTER, M. R. (1972). Point processes generated by level crossings. *Stochastic Point Processes*. Wiley, New York.

MCKAY, M. D., CONOVER, W. J. and BECKMAN, R. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** 239–245.

OWEN, A. B. (1990). A central limit theorem for Latin hypercube sampling. Technical report, Dept. Statist., Stanford Univ.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

LARRY L. SCHUMAKER

Vanderbilt University

Clearly the fitting of functions of more than two variables is an important problem, and it is nice to see that statisticians are willing to tackle it. Mathematicians have tended to concentrate on the bivariate case (perhaps because even there, much remains to be done). The recently published bibliography [5] provides a fairly up-to-date list of what approximation theorists have been doing. Some of this work does deal with the many variable case. In particular, the papers [1], [2], [3], [4] and [11] deal with adaptive fitting of piecewise polynomials, much in the spirit of the paper under discussion. These papers also deal with the problem of giving error bounds.

Approximation theorists have also recently been interested in the problem of approximating multivariate functions by sums of univariate functions. In this connection I would like to cite [6], [7] (see also the bibliography [5]). Other references can be found in the book [6].

Next, a few comments on the paper under discussion. I am a bit puzzled by the assertion in Section 2.4.2 and later in Section 3.2 that lack of smoothness of the approximating functions limits the accuracy of the approximation. Generally it is true that lack of smoothness of the function to be approximated limits accuracy, while for the approximating functions it is the degree of the polynomials used which is critical. Similarly, I do not understand the discussion of end effects in Section 3.7. The classical natural splines perform badly near the boundaries precisely because they smoothly match linear functions there; i.e., they are constrained at the endpoints in the wrong way. The author uses a basis of piecewise linear functions which are smoothed out to be C^1. If one does not need C^1 functions, it seems it would be better to simply use linear splines to begin with. As far as I know, the approximation properties of