REFERENCES

BECKER, R. A., CHAMBERS, J. M. and WILKS, A. R. (1988). *The New S Language*. Wadsworth/Brooks Cole, Pacific Grove, CA.

CHAMBERS, J. M. and HASTIE, T. J. (1991). *Statistical Models in S*. Wadsworth/Brooks Cole, Pacific Grove, CA.

GU, C. and WAHBA, G. (1991). Smoothing spline ANOVA with componentwise Bayesian "confidence intervals." Technical Report 881, Dept. Statistics, Univ. Wisconsin–Madison.

AT&T BELL LABORATORIES
MURRAY HILL, NEW JERSEY 07974

REJOINDER

CHARLES J. STONE

University of California, Berkeley

I wish to thank Buja and Hastie for their interesting and stimulating remarks. In particular, Buja's improvement over my Lemmas 3.3 and 3.4 is very elegant and may be useful in other contexts. His joint work with Donnell and Stuetzle on the analysis of additive dependencies in data sounds intriguing, and I look forward to reading about it soon.

Hastie gives a brief but excellent description of the formula language in S and the ease with which it can be used in the context of linear and generalized linear models to specify main effects as polynomial splines and selected interactions in terms of the corresponding tensor products. He points out that stepwise model selection procedures are also available in S for determining which main effects and interactions to include; that is, in the notation of the present paper, for adaptively choosing $\mathcal{S}$. As he also notes, however, these facilities are not convenient for selecting the number and placement of knots. The high-level stepwise model selection facilities that are currently available in S are compatible with the spirit of the theory developed in the present paper, but not with that of methodologies such as MARS that are adaptive at the level of the individual basis functions, that is, that adaptively select the individual knots and tensor product basis functions.

Recently, in Kooperberg, Stone and Truong (1993b), the theory developed in the present paper has been modified to handle hazard regression, which can be nonproportional and which includes a smooth model for the baseline hazard function. The corresponding MARS-like adaptive methodology is described in Kooperberg, Stone and Truong (1993a). Kooperberg has written a program in C that implements this methodology and an interface based on S. The combined software is available from statlib by sending an email with the body `send hare from S` to statlib@stat.cmu.edu. Concurrently, Kooperberg and Stone (1993) described similar methodology and software for hazard estimation without covariates. Kooperberg, Truong and I are now working on the theory and methodology for logspline spectral density estimation, while Bose, Kooperberg

and I are modifying the theory and methodology to handle polychotomous regression and its application to multiple classification.

Thus, distinct but closely related theories have been or are being developed for regression, logistic and Poisson regression, polychotomous regression, hazard regression and the estimation of hazard, density, conditional density and spectral density functions. It would be worthwhile to synthesize this theoretical work. Similarly, it would be worthwhile to synthesize the corresponding adaptive methodologies and their interfaces based on S.

Buja asks how I think about the problem of confounding, both theoretically and in terms of guidance for practitioners. Theoretically, I have thought of confounding in the context of the present paper as being handled by Lemmas 3.1 and 3.6 under Condition 1 (that the density function f be bounded away from zero and infinity). From Buja's comments, however, I suspect that he and his coauthors have developed a deeper and more interesting understanding of this issue. Practically, I believe that confounding is considerably ameliorated by adaptive model selection at the level of the individual basis functions. In order to lend credence to this belief and to illustrate the use of some of the software and modeling tools in S described by Hastie, I will discuss the analysis of a simple fractional factorial experiment.

The experimental data are taken from John [(1971), page 182]. According to John's description, there were four major operating variables in the treatment of lube oil at a refinery, each of which was quantitative; these were taken as the factors in the experiment and referred to as A, B, C and D. The objective of the experimental program was to find a set of operating conditions that would optimize a measure of quality in the lube oil, which was used as the response variable. Three levels of each factor were used. John assigned the levels 0 for "low," 1 for "medium" and 2 for "high," and he denoted the levels of A, B, C and D by x_1, x_2, x_3 and x_4, respectively. A complete factorial experiment would involve $3^4 = 81$ factor–level combinations. The actual experiment involved the one-third fraction of the complete factorial experiment containing the $3^{4-1} = 27$ factor–level combinations such that $x_1 + x_2 + x_3 + x_4 \equiv 0 \bmod 3$. Thus if we ignore any one of the four factors, we can view the experiment as a complete factorial experiment with one run at each combination of a level of the three remaining factors. The experimental data are shown in Table 1.

John states that "the two-factor interactions are hopelessly jumbled." This can easily be seen in S by looking at the output of

```
alias(aov(y ~ (A + B + C + D)^2)).
```

Indeed, for the given design, the model involving all the main effects and two-factor interactions is nonidentifiable, which is an extreme form of confounding. On the other hand, if we exclude all two-factor interactions involving one of the factors or all those not involving that factor, the remaining components are orthogonal. In particular,

```
summary(aov(y ~ (A+B+C)^2+D))
```

TABLE 1
Lube oil data

0000	4.2	0012	5.9	0021	8.2
0102	13.1	0111	16.4	0120	30.7
0201	9.5	0210	22.2	0222	31.0
1002	7.7	1011	16.5	1020	14.3
1101	11.0	1110	29.0	1122	55.0
1200	8.5	1212	37.4	1221	66.3
2001	11.4	2010	21.1	2022	57.9
2100	13.5	2112	51.6	2121	76.5
2202	31.0	2211	74.5	2220	85.1

TABLE 2
ANOVA table

	DF	SS	MS	*F*	*P*-value
A	2	4496.3	2248.1	51.301	0.00017
B	2	2768.7	1384.3	31.590	0.00065
C	2	5519.8	2759.9	62.979	0.00009
D	2	283.4	141.7	3.233	0.11149
A : *B*	4	310.8	77.7	1.773	0.25267
A : *C*	4	1232.9	308.2	7.034	0.01886
B : *C*	4	669.7	167.4	3.820	0.07069
Residuals	6	262.9	43.8		

yields the ANOVA table shown in Table 2. Alternatively,

$$\texttt{summary.aov(lm(y} \sim \texttt{(poly(x1,2) + poly(x2,2) + poly(x3,2))\^{}2 + poly(x4,2)))}$$

yields the equivalent ANOVA table shown in Table 3.
Moreover,

$$\texttt{summary(lm(y} \sim \texttt{(poly(x1,2) + poly(x2,2) + poly(x3,2))\^{}2 + poly(x4,2)))}$$

yields the coefficient table shown in Table 4. In this table, $\texttt{poly}_1(\texttt{x}_\texttt{m},2)$ and $\texttt{poly}_2(\texttt{x}_\texttt{m},2)$ are the orthonormal polynomials given by

$$\texttt{poly}_1(\texttt{x}_\texttt{m}, 2) = (\texttt{x}_\texttt{m} - 1)/\sqrt{18} \quad \text{and} \quad \texttt{poly}_2(\texttt{x}_\texttt{m}, 2) = [(\texttt{x}_\texttt{m} - 1)^2 - \tfrac{2}{3}]/\sqrt{6}.$$

Examination of the six linear×quadratic and the three quadratic×quadratic interaction terms in Table 4 suggests that these nine terms could be dropped from the model. After similar analyses involving other combinations of the four factors, we are led to the conclusion that in analyzing the lube oil data, we should restrict attention to models that exclude the 12 linear×quadratic and the 6 quadratic×quadratic interaction terms involving x_1, x_2, x_3 and x_4. In other words, we should restrict attention to quadratic polynomial models in these variables, as would ordinarily be done in the context of response surface

TABLE 3
Alternative ANOVA table

	DF	SS	MS	*F*	*P*-value
poly(x_1, 2)	2	4496.3	2248.1	51.301	0.00017
poly(x_2, 2)	2	2768.7	1384.3	31.590	0.00065
poly(x_3, 2)	2	5519.8	2759.9	62.979	0.00009
poly(x_4, 2)	2	283.4	141.7	3.233	0.11149
poly(x_1, 2) : poly(x_2, 2)	4	310.8	77.7	1.773	0.25267
poly(x_1, 2) : poly(x_3, 2)	4	1232.9	308.2	7.034	0.01886
poly(x_2, 2) : poly(x_3, 2)	4	669.7	167.4	3.820	0.07069
Residuals	6	262.9	43.8		

TABLE 4
Coefficient table

	Value	SE	*t*
(Intercept)	29.981	1.274	23.534
$\text{poly}_1(x_1, 2)$	66.327	6.620	10.019
$\text{poly}_2(x_1, 2)$	9.852	6.620	1.488
$\text{poly}_1(x_2, 2)$	51.454	6.620	7.773
$\text{poly}_2(x_2, 2)$	−11.009	6.620	−1.663
$\text{poly}_1(x_3, 2)$	74.270	6.620	11.219
$\text{poly}_2(x_3, 2)$	−1.946	6.620	−0.294
$\text{poly}_1(x_4, 2)$	14.614	6.620	2.208
$\text{poly}_2(x_4, 2)$	−8.355	6.620	−1.262
$\text{poly}_1(x_1, 2)\text{poly}_1(x_2, 2)$	83.700	34.398	2.433
$\text{poly}_2(x_1, 2)\text{poly}_1(x_2, 2)$	−2.425	34.398	−0.070
$\text{poly}_1(x_1, 2)\text{poly}_2(x_2, 2)$	32.216	34.398	0.937
$\text{poly}_2(x_1, 2)\text{poly}_2(x_2, 2)$	18.500	34.398	0.538
$\text{poly}_1(x_1, 2)\text{poly}_1(x_3, 2)$	180.750	34.398	5.255
$\text{poly}_2(x_1, 2)\text{poly}_1(x_3, 2)$	−8.747	34.398	−0.254
$\text{poly}_1(x_1, 2)\text{poly}_2(x_3, 2)$	−23.123	34.398	−0.672
$\text{poly}_2(x_1, 2)\text{poly}_2(x_3, 2)$	−2.650	34.398	−0.077
$\text{poly}_1(x_2, 2)\text{poly}_1(x_3, 2)$	114.450	34.398	3.327
$\text{poly}_2(x_2, 2)\text{poly}_1(x_3, 2)$	−50.836	34.398	−1.478
$\text{poly}_1(x_2, 2)\text{poly}_2(x_3, 2)$	−46.332	34.398	−1.347
$\text{poly}_2(x_2, 2)\text{poly}_2(x_3, 2)$	−15.850	34.398	−0.461

exploration [see Box and Draper (1987)]. In S, we can use

$$\texttt{lm(y} \sim \texttt{poly(x1, x2, x3, x4, 2))}$$

to fit the full quadratic model. After examining summary statistics from this fit, we may be led to drop some of the linear×linear interaction terms or some of the individual quadratic terms. Whichever model we end up with in this manner, we can be sure that confounding will not be a serious practical issue.

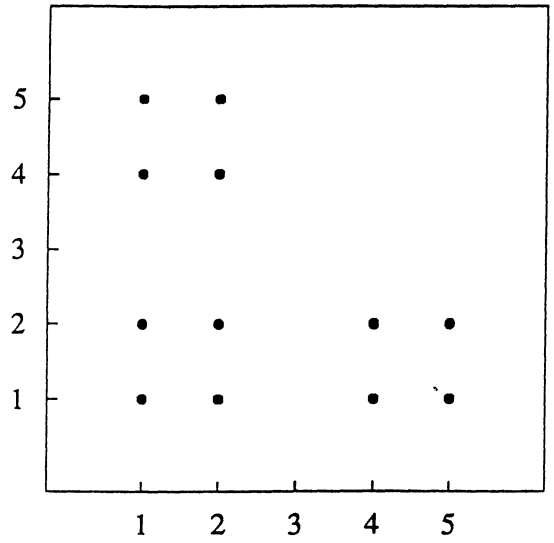

FIG. 1. *An irregular design.*

To lend further support to the belief that confounding can be ameliorated by adaptive model selection at the basis function level, consider two variables each ranging over [0, 6], the space of linear splines in the first variable having a single knot at 3, and the space of linear splines in the second variable also having a single knot at 3. Observe that 1, x_1, $(x_1 - 3)_+$, x_2, $(x_2 - 3)_+$, x_1x_2, $(x_1-3)_+x_2$, $x_1(x_2-3)_+$ and $(x_1-3)_+(x_2-3)_+$ form a basis of the tensor product of these two spaces. Consider the design of size 12 shown in Figure 1. The nine-dimensional tensor product space is clearly nonidentifiable relative to this design since $(x_1 - 3)_+(x_2 - 3)_+$ equals zero at all the design points. If the basis functions are entered one at a time by stepwise addition, however, the last basis function would never enter and the resulting model would necessarily be identifiable and not too highly confounded.

In closing this discussion, I would like to emphasize that, in my own collaborative work, theoretical results for nonadaptive procedures such as those contained in the present paper have served as a useful framework for developing more practical, highly adaptive methodologies.

REFERENCES

BOX, G. E. P. and DRAPER, N. R. (1987). *Empirical Model Building and Response Surfaces*. Wiley, New York.

JOHN, P. W. M. (1971). *Statistical Design and Analysis of Experiments*. Macmillan, New York.

KOOPERBERG, C. and STONE, C. J. (1993). Hazard estimation with flexible tails. Technical Report 388, Dept. Statistics, Univ. California, Berkeley.

KOOPERBERG, C., STONE, C. J. and TRUONG, Y. K. (1993a). Hazard regression. Technical Report 389, Dept. Statistics, Univ. California, Berkeley.

KOOPERBERG, C., STONE, C. . and TRUONG, Y. K. (1993b). The L_2 rate of convergence for hazard regression. Technical Report 390, Dept. Statistics, Univ. California, Berkeley.

DEPARTMENT OF STATISTICS
STATISTICAL LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720