The Annals of Statistics
1995, Vol. 23, No. 6, 1896–1920

TESTING FOR ADDITIVITY IN NONPARAMETRIC REGRESSION

BY R. L. EUBANK,[1] J. D. HART, D. G. SIMPSON[2] AND L. A. STEFANSKI

Texas A & M University, Texas A & M University, University of Illinois and North Carolina State University

Additive models are one means of assuaging the curse of dimensionality when nonparametric smoothing methods are used to estimate multivariable regression functions. It is important to have methods for testing the fit of such models, especially in high dimensions where visual assessment of fit becomes difficult. New tests of additivity are proposed in this paper that derive from Fourier series estimators with data-driven smoothing parameters. Other tests related to the classical Tukey test for additivity are also considered. While the new tests are consistent against essentially any "smooth" alternative to additivity, the Tukey-type tests are found to be inconsistent in certain situations. Asymptotic power of both varieties of tests is studied under local alternatives that tend toward additivity at a parametric rate, and small-sample power comparisons are carried out by means of a simulation study.

1. Introduction. Nonparametric smoothers have become an increasingly popular means of estimating regression functions, especially in problems where the predictor is one-dimensional. In settings with higher-dimensional predictors, smoothing methods that impose no structure on the regression function become less attractive, owing to the sparseness of data in the predictor space. This is the so-called curse of dimensionality. In recent years a number of methods have been proposed that seek to circumvent this problem while retaining a nonparametric flavor. These methods include those based on additive models of the form

$$Y = \sum_{i=1}^{k} f_i(t_i) + \varepsilon, \tag{1.1}$$

where Y is the response variable, $t_1, \dots, t_k$ are the predictor variables and ε is an unobserved error term. The functions $f_1, \dots, f_k$ are unknown and assumed merely to be "smooth."

Received November 1993; revised June 1995.
[1]Research supported in part by NSF Grant DMS-93-00918.
[2]Research supported by NSF Grant DMS-92-07730.
AMS 1991 *subject classifications*. 62E20, 62E25, 62G07, 62G10, 62G20.
Key words and phrases. Additive models, dimension reduction, kernel estimators, Tukey test, Fourier series, order-selection test, local linear smoothers.

When model (1.1) holds, much can be gained in terms of estimation efficiency over the structureless model

$$Y = f(t_1, \dots, t_k) + \varepsilon. \tag{1.2}$$

It is well known that the efficiency of linear smoothers under model (1.2) tends to deteriorate rapidly with increasing dimension k. By contrast, under appropriate conditions, the regression function in (1.1) can be estimated so that the estimation error tends to 0 at the same rate as in the case of a single predictor, that is, $k = 1$ [Stone (1985)]. Hence, nonparametric methods can effectively defeat the curse of dimensionality when the structure in (1.1) is justified.

The problem of testing for additivity has been addressed by Hastie and Tibshirani (1990), Barry (1993) and Spiegelman and Wang (1994). Since one will often use an additive model only as an approximation, diagnostic tests give the data analyst information concerning the extent of departures from additivity. The utility of formal tests increases with higher dimensions since departures from additivity become more difficult to assess by graphical means when the number of independent variables is large.

The purpose of this paper is to analyze a Tukey-type test of additivity proposed by Hastie and Tibshirani (1990) and to introduce new tests of a more omnibus nature. The basic ideas behind our testing methodology are easiest to demonstrate for models with two predictors. Thus, we will focus on $k = 2$ in the sequel and then sketch extensions to higher dimensions.

With two predictors a Tukey-type test for additivity arises from considering an alternative to (1.1) in which the departure from additivity has the form $\gamma f_1 f_2$ for some unknown constant γ. Since additivity in this case is equivalent to $\gamma = 0$, the Tukey (1949) testing paradigm proceeds by estimating γ and then using the estimate to test if γ differs from 0.

In the next section we investigate the large-sample power properties of two tests of Tukey type. One of the two is exactly the test proposed by Tukey (1949) for the classical analysis of variance setting, while the other uses nonparametric smoothers to estimate f_1 and f_2. The former test differs from the latter in that it neglects to take advantage of smoothness of f_1 and f_2. We show that both of these tests can detect certain alternatives that approach additivity at parametric rates. However, Tukey tests are inconsistent against some reasonable departures from additivity. In fact, they cannot detect any alternative that is orthogonal to the product $f_1 f_2$. We thus propose a second set of tests that are consistent against essentially any departure from additivity. These tests utilize Fourier series ideas and are analogous to the "order selection" test proposed by Eubank and Hart (1992). Besides being consistent, they can also detect alternatives converging to additivity at parametric rates.

The rest of the paper proceeds as follows. Distribution theory for the Tukey-type and order-selection tests is considered in Sections 2 and 3, respectively. A simulation study addressing power of the tests is summarized in Section 4, and proofs of theoretical results are given in Section 5.

2. Tukey-type tests. Consider the case where a response variable Y is observed over a grid of design points corresponding to two independent variables t_1 and t_2. We assume that $n = n_1 n_2$ data vectors (t_{1r}, t_{2k}, y_{rk}), $r = 1, \ldots, n_1$, $k = 1, \ldots, n_2$, are observed with

$$y_{rk} = \mu + f_1(t_{1r}) + f_2(t_{2k}) + f_{12}(t_{1r}, t_{2k}) + \varepsilon_{rk}, \quad r = 1, \ldots, n_1,\ k = 1, \ldots, n_2, \tag{2.1}$$

where μ is an unknown constant, t_{1r}, $r = 1, \ldots, n_1$, and t_{2k}, $k = 1, \ldots, n_2$, are design points and the ε_{rk}'s are iid random variables with $E\varepsilon_{11} = 0$ and $\mathrm{Var}(\varepsilon_{11}) = \sigma^2 < \infty$. The functions f_1, f_2 and f_{12} are unknown but satisfy some identifiability and smoothness conditions to be discussed subsequently. Borrowing terminology from analysis of variance, f_1 and f_2 will be referred to as *main effect* functions and f_{12} will be called the *interaction* function.

The goal here is to test the hypothesis that f_{12} is identical to 0 in (2.1). Hastie and Tibshirani [(1990), page 264] have suggested adapting the Tukey (1949) paradigm from classical analysis of variance for this purpose. The basic idea is to fit the following regression function to the data:

$$\mu + f_1(t_1) + f_2(t_2) + \gamma f_1(t_1) f_2(t_2). \tag{2.2}$$

In this context, testing additivity is tantamount to testing $\gamma = 0$. Thus, the problem becomes one of estimating the parameter γ in (2.2) and then using the estimator to construct a test statistic.

Given estimators $\hat{f}_1$ and $\hat{f}_2$ of the main effect functions that satisfy $\sum_{j=1}^{n_i} \hat{f}_i(t_{ij}) = 0$, $i = 1, 2$, a least squares estimator of γ is

$$\hat{\gamma} = \frac{\sum_{r=1}^{n_1} \sum_{k=1}^{n_2} \hat{f}_1(t_{1r}) \hat{f}_2(t_{2k}) \big(y_{rk} - \hat{f}_1(t_{1r}) - \hat{f}_2(t_{2k})\big)}{\sum_{r=1}^{n_1} \hat{f}_1^2(t_{1r}) \sum_{k=1}^{n_2} \hat{f}_2^2(t_{2k})}. \tag{2.3}$$

In particular, if we set

$$\bar{y}_{r\cdot} = \frac{1}{n_2} \sum_{j=1}^{n_2} y_{rj}, \qquad r = 1, \ldots, n_1,$$

$$\bar{y}_{\cdot k} = \frac{1}{n_1} \sum_{j=1}^{n_1} y_{jk}, \qquad k = 1, \ldots, n_2,$$

$$\bar{y}_{\cdot\cdot} = \frac{1}{n} \sum_{r=1}^{n_1} \sum_{k=1}^{n_2} y_{rk}$$

and take

$$\hat{f}_1(t_{1r}) = \bar{y}_{r\cdot} - \bar{y}_{\cdot\cdot}, \qquad r = 1, \ldots, n_1, \tag{2.4a}$$

$$\hat{f}_2(t_{2k}) = \bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot}, \qquad k = 1, \ldots, n_2, \tag{2.4b}$$

then (2.3) becomes

$$\hat{\gamma}_T = \frac{\sum_{r=1}^{n_1}\sum_{k=1}^{n_2}(\bar{y}_{r\cdot}-\bar{y}_{\cdot\cdot})(\bar{y}_{\cdot k}-\bar{y}_{\cdot\cdot})(y_{rk}-\bar{y}_{r\cdot}-\bar{y}_{\cdot k}+\bar{y}_{\cdot\cdot})}{\sum_{r=1}^{n_1}(\bar{y}_{r\cdot}-\bar{y}_{\cdot\cdot})^2\sum_{k=1}^{n_2}(\bar{y}_{\cdot k}-\bar{y}_{\cdot\cdot})^2}. \tag{2.5}$$

This is exactly the Tukey (1949) estimator of γ in (2.2). Its use can be justified in this setting by viewing model (2.1) as a two-way layout with possible interaction and one observation per cell.

The Tukey choice (2.4) for main effect estimators is inefficient in model (2.1) since it does not exploit the smoothness of the main effect functions. Thus, we shall also consider how $\hat{\gamma}$ in (2.3) behaves when nonparametric smoothers are used to estimate f_1 and f_2.

2.1. *Large-sample properties.* To assess asymptotic power properties of tests deriving from (2.3), we study the case where data follow the model

$$y_{rk} = \mu + f_1(t_{1r}) + f_2(t_{2k}) + \frac{1}{\sqrt{n}}g(t_{1r}, t_{2k}) + \varepsilon_{rk}, \quad r = 1,\dots,n_1,\ k = 1,\dots,n_2. \tag{2.6a}$$

The design points in (2.6a) are generated by positive, continuous densities h_1 and h_2 via the relationships

$$\int_0^{t_{ij}} h_i(u)\,du = \frac{(2j-1)}{2n_i}, \qquad j = 1,\dots,n_i,\ i = 1,2. \tag{2.6b}$$

We also assume that $f_1, f_2 \in C^1[0,1]$, $g \in C^1([0,1]\times[0,1]) = \{p: \partial^2 p(u,v)/\partial u\,\partial v$ is continuous in u and $v\}$,

$$\int_0^1 f_i(u)h_i(u)\,du = 1, \qquad i = 1,2, \tag{2.6c}$$

and

$$\int_0^1 g(u,\cdot)h_1(u)\,du \equiv \int_0^1 g(\cdot,u)h_2(u)\,du \equiv 0. \tag{2.6d}$$

Conditions (2.6c)–(2.6d) represent one set of identifiability restrictions which insure that the functions in (2.6a) are uniquely defined.

Model (2.6) provides an alternative to additivity in which the mean function converges to an additive model at the rate $n^{-1/2}$. This formulation allows us to obtain more precise information about our tests than would be possible under a fixed alternative. In the latter instance one can typically establish only consistency. In contrast, by studying the behavior of our tests under the local alternative (2.6), we can derive explicit, large-sample power formulae which also imply consistency against certain fixed alternatives. The special case of $g \equiv 0$ in (2.6) gives us the limiting null distribution for a test.

Our first result describes the large-sample behavior of the standard Tukey statistic (2.5). Throughout this section we use the notation

$$\|f_i\|^2 = \int_0^1 f_i^2(u)h_i(u)\,du, \qquad i = 1,2,$$

and

$$\langle g, f_1 f_2 \rangle = \int_0^1 \int_0^1 g(u,v) f_1(u) f_2(v) h_1(u) h_2(v)\, du\, dv.$$

THEOREM 2.1. *In model* (2.6) *assume that* (i) $E\varepsilon_{11}^6 < \infty$, $E\varepsilon_{11}^j = 0$, $j = 1,3,5$, (ii) $\|f_i\| > 0$, $i = 1,2$, *and* (iii) $n_1/n_2 \to \theta \in (0,1)$ *as* $n \to \infty$. *Then*

$$\sqrt{n}\,\hat{\gamma}_T \to_{\mathscr{D}} N\left(\frac{\langle g, f_1 f_2 \rangle}{(\|f_1\|\,\|f_2\|)^2}, \frac{\sigma^2}{(\|f_1\|\,\|f_2\|)^2} \right) \tag{2.7}$$

as $n \to \infty$.

From Theorem 2.1 and the fact that in this case (see Section 5) $n_i^{-1}\sum_{j=1}^{n_i} \hat{f}_i^2(t_{ij}) \to_P \|f_i\|^2$, $i = 1,2$, we see that

$$\begin{aligned} T_{1n} &= \frac{\sum_{r=1}^{n_1}\sum_{k=1}^{n_2}(\bar{y}_{r\cdot} - \bar{y}_{\cdot\cdot})(\bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot})(y_{rk} - \bar{y}_{r\cdot} - \bar{y}_{\cdot k} + \bar{y}_{\cdot\cdot})}{\left[\sigma^2 \sum_{r=1}^{n_1}(\bar{y}_{r\cdot} - \bar{y}_{\cdot\cdot})^2 \sum_{k=1}^{n_2}(\bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot})^2\right]^{1/2}} \\ &= \frac{\hat{\gamma}_T}{\sigma}\left[\sum_{r=1}^{n_1}(\bar{y}_{r\cdot} - \bar{y}_{\cdot\cdot})^2 \sum_{k=1}^{n_2}(\bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot})^2\right]^{1/2} \end{aligned} \tag{2.8}$$

has a limiting $N(\langle g, f_1 f_2 \rangle/(\sigma\|f_1\|\,\|f_2\|), 1)$ distribution. In particular, T_{1n} will be asymptotically standard normal under an additive model. This means we can reject additivity at level α if $|T_{1n}| > Z_{\alpha/2}$, for Z_p the $100(1-p)$th percentile of the standard normal distribution.

The asymptotic power of T_{1n} is

$$\begin{aligned} \lim_{n\to\infty} P\big(|T_{1n}| > Z_{\alpha/2}\big) &= \Phi\left(-Z_{\alpha/2} - \frac{\langle g, f_1 f_2 \rangle}{\sigma\|f_1\|\,\|f_2\|}\right) \\ &\quad + \Phi\left(-Z_{\alpha/2} + \frac{\langle g, f_1 f_2 \rangle}{\sigma\|f_1\|\,\|f_2\|}\right), \end{aligned} \tag{2.9}$$

where Φ is the standard normal distribution function. The power is monotone increasing in $|\langle g, f_1 f_2 \rangle|/\|f_1\|\,\|f_2\|$ and monotone decreasing in σ. Thus, power is maximized for fixed σ when $g \propto f_1 f_2$. The worst case occurs when $g \perp f_1 f_2$ in the sense that $\langle g, f_1 f_2 \rangle = 0$. In that event the asymptotic power of T_{1n} is equal to the level α.

Theorem 2.1 also has the implication that a test based on T_{1n} will be consistent against a fixed alternative as in (2.1) with $f_{12} \in C^1([0,1] \times [0,1])$ and $\langle f_{12}, f_1 f_2 \rangle \neq 0$. If, however, $\langle f_{12}, f_1 f_2 \rangle = 0$, then T_{1n} will have a standard normal limiting distribution and the test is inconsistent.

It is also of interest to study what transpires when estimators other than the treatment means (2.4) are used in (2.3). As previously noted, (2.4) is a poor choice for main effect estimators under model (2.6) since the convergence rates are only of probability order $n_i^{-1/2}$, $i = 1,2$. We will give examples of estimators shortly which use the smoothness of the functions being estimated to attain rates $O_p(n^{-2/5}) = O_p(n_i^{-4/5})$, $i = 1, 2$. In addition, one will generally want to construct additivity tests using the nonparametric smoothers employed in fitting the data instead of the raw treatment means. Thus, our next result describes the properties of $\hat{\gamma}$ for the case where f_1 and f_2 are estimated using a generic set of smoothers.

THEOREM 2.2. *Assume that* (i) $E\varepsilon_{11} = 0$, (ii) $\operatorname{Var} \varepsilon_{11} = \sigma^2 < \infty$, (iii) $\|f_i\| > 0$, $i = 1, 2$, *and* (iv) $n_1/n_2 \to \theta \in (0, 1)$ *as* $n \to \infty$. *Let* $\hat{f}_1$ *and* $\hat{f}_2$ *in* (2.3) *be estimators of* f_1 *and* f_2 *that satisfy the following*:

(a) $\sum_{j=1}^{n_i} \hat{f}_i(t_{ij}) = o_p(1)$, $i = 1, 2$;
(b) f_i, $i = 1, 2$, *is continuously differentiable with*

$$\sup_u \left| \hat{f}_i^{(j)}(u) - f_i^{(j)}(u) \right| = o_p(1), \qquad j = 0, 1,\ i = 1, 2.$$

Then

$$\sqrt{n}\, \hat{\gamma} \to_{\mathscr{D}} N\left(\frac{\langle g, f_1 f_2 \rangle}{(\|f_1\|\,\|f_2\|)^2}, \frac{\sigma^2}{(\|f_1\|\,\|f_2\|)^2} \right). \tag{2.10}$$

Theorem 2.2 implies that tests for additivity can also be based on

$$T_{2n} = \frac{\sum_{r=1}^{n_1} \sum_{k=1}^{n_2} \hat{f}_1(t_{1r}) \hat{f}_2(t_{2k}) \big(y_{rk} - \hat{f}_1(t_{1r}) - \hat{f}_2(t_{2k}) \big)}{\big(\sigma^2 \sum_{r=1}^{n_1} \hat{f}_1^2(t_{1r}) \sum_{k=1}^{n_2} \hat{f}_2^2(t_{2k}) \big)^{1/2}}. \tag{2.11}$$

This statistic will have the same large-sample properties as T_{1n} in (2.8). Thus, rejection of additivity at level α will be indicated if $|T_{2n}| > Z_{\alpha/2}$, and the asymptotic power for this test is the same as in (2.9).

To explore some of the practical implications of Theorem 2.2, assume now that the estimators of f_1 and f_2 are linear smoothers. It suffices to deal with estimation of either one of the main effects since the other can be analyzed similarly. Thus, consider estimation of f_1 using

$$\hat{f}_1(\cdot) = \sum_{i=1}^{n_1} w_i(\cdot)(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}), \tag{2.12}$$

where $\{w_i(\cdot)\}_{i=1}^{n_1}$ is a set of weights that depend on the design and satisfy the following:

(2.13a) $w_i(u)$ is a continuously differentiable function of u for each i and each n_1;

$$\sup_u \sum_{i=1}^{n_1} w_i(u)^2 = O\left(\frac{1}{n_1 b} \right); \tag{2.13b}$$

$$\sup_u \sum_{i=1}^{n_1} w_i'(u)^2 = O\left(\frac{1}{n_1 b^3} \right); \tag{2.13c}$$

$$\sum_{i=1}^{n_1} w_i(u) = 1 \quad \text{for all } u \in [0,1]; \tag{2.13d}$$

$$\sup_u \left| \sum_{i=1}^{n_1} w_i'(u) \right| = o(1), \tag{2.13e}$$

$$\begin{aligned} &\sup_u \left| \sum_{i=1}^{n_1} w_i(u) f_1(t_{1i}) - f_1(u) \right| \\ &\qquad = o(1) = \sup_u \left| \sum_{i=1}^{n_1} w_i'(u) f_1(t_{1i}) - f_1'(u) \right|, \end{aligned} \tag{2.13f}$$

$$\sum_{i=1}^{n_1} |w_i(u_1) - w_i(u_2)| \le \frac{C_1}{b}|u_1 - u_2| \quad \text{for all } u_1, u_2 \tag{2.13g}$$

and

$$\sum_{i=1}^{n_1} |w_i'(u_1) - w_i'(u_2)| \le \frac{C_2}{b^{1+\delta}}|u_1 - u_2|^{\delta} \quad \text{for all } u_1, u_2. \tag{2.13h}$$

In relations (2.13), b represents a nonnegative bandwidth-type parameter with $b \to 0$ as $n = n_1 n_2 \to \infty$, C_1 and C_2 are finite, nonnegative constants and $\delta \in (0,1]$ is a Lipschitz constant.

It is possible to establish the following proposition.

PROPOSITION 2.1. *Assume that the* ε_{ij} *in* (2.6) *are iid with zero mean and finite fourth moment, and let conditions* (2.13a)–(2.13h) *hold. Then, for any* $\theta > 0$,

$$\sup_u |\hat{f}_1(u) - f_1(u)| = O_p\left(\frac{1}{n^{\theta} b} + \frac{1}{n^{1-\theta/2} b} \right) + o_p(1) \tag{2.14a}$$

and

$$\sup_u |\hat{f}_1'(u) - f_1'(u)| = O_p\left(\frac{1}{n^{\delta\theta} b^{1+\delta}} + \frac{1}{n^{1-\theta/2} b^3} \right) + o_p(1). \tag{2.14b}$$

The proof of this result follows from conditions (2.13a)–(2.13h) and a partitioning argument where the interval $[0,1]$ is divided into $n_1^{2\theta}$ subintervals, as in the proof of Theorem 11.2 in Müller (1988). Instead of Müller's exponential inequality we have used Theorem 2 of Whittle (1960) with fourth-order moments.

Proposition 2.1 allows us to verify condition (b) of Theorem 2.2 by checking conditions on the weight function used for the smoother. For a second-order smoother with smoothing parameter/bandwidth b, a global optimal level of smoothing is obtained when b is of order $n^{-1/5}$. In that case we can take θ to be any value in $(1/5, 4/5)$ and therefore need $\delta > 1/3$ in (2.14) for global consistency of both $\hat{f}_1$ and $\hat{f}_1'$. This restriction on δ can be weakened by requiring further moments for the random errors.

A specific case where (2.14) can be verified is provided by local linear smoothers [see, e.g., Fan (1992)]. One may check that Proposition 2.1 holds for such estimators when they are defined in terms of a kernel that has finite support and two continuous derivatives.

Another apparent implication of Theorem 2.2 is that there is little to be gained in large samples from using smoothers over the raw treatment means when constructing Tukey-type tests. However, our simulation results in Section 4 show that smoothing can be beneficial and that test statistics like (2.11) can perform significantly better than (2.8) in finite samples.

To use T_{1n} or T_{2n} in practice, one will require a consistent estimator of σ under model (2.6). There are a variety of estimators from which to choose. These include parallels of mean squared error estimators from analysis of variance and nonparametric difference estimators such as those in Hall, Kay and Titterington (1990). Most of these estimators are consistent under mild conditions. However, their finite-sample properties are frequently unsatisfactory. We will discuss two consistent estimators we have found to be effective in Section 4.

The fundamental conclusions about Tukey-type tests that follow from Theorems 2.1 and 2.2 are quite intuitive given the form of the test statistics. We would expect tests based on either (2.8) or (2.11) to do the following: (i) perform well if the interaction function is well approximated by a constant multiple of $f_1 f_2$; (ii) perform poorly when the interaction is not aligned with a product of the main effects; and (iii) perform worse with noisy data.

To extend the Tukey-type tests to $k > 2$ dimensions, one proceeds along the lines of the discussion surrounding (2.2) and models the interaction part of the mean function as a linear combination of $\binom{k}{2}$ pairwise main effect products, $\binom{k}{3}$ three-way main effect products and so on. For example, with $k = 3$ predictors t_1, t_2 and t_3, the interaction is treated as having the form

$$\sum_{1 \le i < j \le 3} \gamma_{ij} f_i(t_i) f_j(t_j) + \gamma_{123} f_1(t_1) f_2(t_2) f_3(t_3),$$

for f_1, f_2 and f_3 the main effect functions. Assuming a model like (2.3) with responses y_{rkl} on an $n_1 \times n_2 \times n_3$ design grid, the γ_{ij} can be estimated analogously to (2.3) and γ_{123} can be estimated by

$$\hat{\gamma}_{123} = \frac{\sum_{r=1}^{n_1} \sum_{k=1}^{n_2} \sum_{l=1}^{n_3} \hat{f}_1(t_{1r}) \hat{f}_2(t_{2k}) \hat{f}_3(t_{3l}) \big(y_{rkl} - \hat{f}_1(t_{1r}) - \hat{f}_2(t_{2k}) - \hat{f}_3(t_{3l}) \big)}{\sum_{r=1}^{n_1} \hat{f}_1^2(t_{1r}) \sum_{k=1}^{n_2} \hat{f}_2^2(t_{2k}) \sum_{l=1}^{n_1} \hat{f}_3^2(t_{3l})}.$$

Theorem 2.2 extends in a straightforward fashion to this case and one finds, for example, that

$$\sqrt{n_1 n_2 n_3}\, \hat{\gamma}_{123} \to_{\mathscr{D}} N\left(\frac{\langle g, f_1 f_2 f_3 \rangle}{(\|f_1\| \|f_2\| \|f_3\|)^2}, \frac{\sigma^2}{(\|f_1\| \|f_2\| \|f_3\|)^2} \right)$$

if conditions (i) and (ii) of Theorem 2.2 hold for all three main effect estimators. One can then combine the tests obtained from the four coefficient estimators to obtain an overall test using some type of level protection methodology.

3. Order-selection tests of additivity. In the previous section we saw that Tukey-type tests had the disadvantage of being inconsistent against departures from additivity that are orthogonal to the product of the main effect functions. We shall therefore derive new tests that do not have this problem and can detect essentially any type of nonadditivity.

Again consider model (2.1) and suppose that we use Fourier series methods to estimate interaction. To this end, define $x_{1i} = (i - 0.5)/n_1$, $i = 1, \dots, n_1$, $x_{2j} = (j - 0.5)/n_2$, $j = 1, \dots, n_2$, and sample Fourier coefficients by

$$\hat{\alpha}_{jk} = \frac{2}{n} \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} (y_{rs} - \bar{y}_{\cdot\cdot}) \cos(\pi j x_{1r}) \cos(\pi k x_{2s}),$$

$$j = 1, \dots, n_1 - 1, \ k = 1, \dots, n_2 - 1.$$

The goal here is to develop tests using methods for selecting the order of a Fourier series estimator constructed from the $\hat{\alpha}_{jk}$.

To motivate the form of the tests, first note that the $\hat{\alpha}_{jk}$ are defined in terms of evenly spaced points regardless of whether or not the t_{rs} form a uniform grid. (This is done to take advantage of the orthogonality properties

$$\sum_{r=1}^{n_i} \cos\left(\frac{\pi j (r - 0.5)}{n_i} \right) \cos\left(\frac{\pi k (r - 0.5)}{n_i} \right) = 0, \qquad j \neq k,$$

for $j, k = 0, 1, \dots, n_i - 1$ and $i = 1, 2$.) We can nonetheless use the $\hat{\alpha}_{jk}$ for detecting interaction; specifically, define the orthogonal series estimator

$$\hat{f}^*_{12}(x_1, x_2) = 2 \sum\sum_{(j,k) \in \Lambda} \hat{\alpha}_{jk} \cos(\pi j x_1) \cos(\pi k x_2), \tag{3.1}$$

$$(x_1, x_2) \in [0, 1] \times [0, 1],$$

for Λ some subset of $\Lambda(n_1, n_2) = \{(j, k): 1 \leq j \leq n_1 - 1, 1 \leq k \leq n_2 - 1\}$. It is not difficult to argue that, by appropriate choice of Λ, $\hat{f}^*_{12}$ consistently estimates the function

$$f^*_{12}(x_1, x_2) = f_{12}\big(H_1^{-1}(x_1), H_2^{-1}(x_2)\big), \qquad (x_1, x_2) \in [0, 1] \times [0, 1],$$

where H_1 and H_2 are the cumulative distribution functions corresponding to the densities h_1 and h_2 in (2.6b). Since h_1 and h_2 are positive, $f^*_{12} \equiv 0$ if and only if $f_{12} \equiv 0$. Hence, there is evidence of an interaction between t_1 and t_2 when $\hat{f}^*_{12}$ differs significantly from 0.

We propose that tests of additivity be based upon a data-driven choice for the set Λ to be used with $\hat{f}^*_{12}$. Since changing Λ in $\hat{f}^*_{12}$ amounts to the usual practice of varying the smoothing parameter(s) of a function estimator, one means of choosing Λ is to optimize an estimated risk criterion. One such criterion is

$$R(\Lambda; C) = n \sum\sum_{(j,k) \in \Lambda} \hat{\alpha}^2_{jk} - C \hat{\sigma}^2 N(\Lambda), \qquad \Lambda \subset \Lambda(n_1, n_2), \tag{3.2}$$

where C is a prespecified positive constant, $N(A)$ denotes the number of elements in the set A and $\hat{\sigma}^2$ is a consistent estimator of σ^2. Since $\Lambda = \varnothing$ corresponds to the null model of additivity, a test can be obtained by rejecting additivity whenever the maximizer $\hat{\Lambda}(C)$ of $R(\Lambda; C)$ over an appropriate collection of subsets of $\Lambda(n_1, n_2)$ is such that $\hat{\Lambda}(C)$ is nonempty. These tests are referred to as *order-selection tests*, since Λ determines the order of the series estimator $\hat{f}_{12}^*$. Maximizing the criterion $R(\Lambda; 2)$ is equivalent to minimizing a nearly unbiased estimator of $E\int(\hat{f}_{12}^* - f_{12}^*)^2$; see Eubank and Hart (1992) for more details.

We will focus attention on two special cases of the test derived from $\hat{\Lambda}(C)$. Specifically, we wish to consider maximizing $R(\Lambda; C)$ over a collection of sets of the form $\{\varnothing, \Lambda_1, \dots, \Lambda_m\}$ with $m = \min(n_1 - 1, n_2 - 1)$. For the Λ_j we will use either

$$\Lambda_j = \{(r, s): 1 \le r \le j, 1 \le s \le j\}, \qquad j = 1, 2, \dots, \tag{3.3}$$

or

$$\Lambda_j = \{(r, s): r \ge 1, s \ge 1, r + s \le j\}, \qquad j = 1, 2, \dots. \tag{3.4}$$

These two schemes correspond to risk criteria for series estimators with only one smoothing parameter. Choosing the Λ_j as in (3.3) produces a series estimator of f_{12}^* of the form

$$\hat{f}_{12}^*(x_1, x_2) = 2 \sum_{j=1}^{\lambda} \sum_{k=1}^{\lambda} \hat{\alpha}_{jk} \cos(\pi j x_1) \cos(\pi k x_2),$$

while the choice (3.4) gives

$$\hat{f}_{12}^*(x_1, x_2) = 2 \sum\sum_{j+k \le \lambda} \hat{\alpha}_{jk} \cos(\pi j x_1) \cos(\pi k x_2),$$

for λ some integer between 0 and m.

We shall also consider another statistic that makes use of a data-driven smoothing parameter. This statistic may be regarded as an F-type ratio with random degrees of freedom. Define $R(0) = 0$ and $R(\lambda) = R(\Lambda_\lambda; 2)$, $\lambda = 1, \dots, \min(n_1, n_2) - 2$, where Λ_λ is defined as in (3.3). Let $\hat{\lambda}$ be the maximizer of $R(\lambda)$ over $\lambda = 0, 1, \dots, \min(n_1, n_2) - 2$, and let $S(m_1, m_2) = \sum_{j=1}^{m_1} \sum_{k=1}^{m_2} \hat{\alpha}_{jk}^2$ for each $m_1, m_2 \ge 1$. The proposed test statistic is $\mathscr{F}$, where $\mathscr{F} = 0$ when $\hat{\lambda} = 0$ and

$$\mathscr{F} = \frac{S(\hat{\lambda}, \hat{\lambda}) / \hat{\lambda}^2}{\{S(n_1 - 1, n_2 - 1) - S(\hat{\lambda}, \hat{\lambda})\} / ((n_1 - 1)(n_2 - 1) - \hat{\lambda}^2)} \tag{3.5}$$

otherwise. Note that $\mathscr{F}$ has the form of a classic F-ratio in which the degrees of freedom are not fixed a priori.

3.1. *Distributional properties under additivity.* We first consider the order-selection tests based on $\hat{\Lambda}(C)$. For the purpose of large-sample developments it will be easier to study an equivalent form of the test. Let $\mathscr{L}$ be a collection of subsets of $\Lambda(n_1, n_2)$, and note that the test which rejects additivity if and only if

$$\arg \max_{\Lambda \in \mathscr{L} \cup \{\varnothing\}} R(\Lambda; C) \neq \varnothing$$

is equivalent to the test that rejects additivity if and only if $V_n \geq C$, where

$$V_n = \max_{\Lambda \in \mathscr{L}} \frac{1}{N(\Lambda)} \sum\sum_{(j,k) \in \Lambda} \left(\frac{n \hat{\alpha}_{jk}^2}{\hat{\sigma}^2} \right).$$

This form shows more explicitly how the test is sensitive to nonzero Fourier coefficients and also makes computation of P-values relatively straightforward.

For our purposes the role of the constant C is to control the level of the test. Under our moment assumptions, for any fixed, positive integers m_1 and m_2, the collection of random variables $\{n \hat{\alpha}_{jk}^2 / \hat{\sigma}^2\colon\ 1 \leq j \leq m_1,\ 1 \leq k \leq m_2\}$ converges in distribution to the collection $\{Z_{jk}^2\colon\ 1 \leq j \leq m_1,\ 1 \leq k \leq m_2\}$ as $n \to \infty$, where the Z_{jk}'s are iid standard normal random variables. Since each $n \hat{\alpha}_{jk}^2 / \hat{\sigma}^2$ is asymptotically distributed as a chi-squared random variable with one degree of freedom, the only reasonable choices for C are those greater than $1 = E(Z_{jk}^2)$.

We now state a theorem that indicates the large-sample distribution of V_n and hence the limiting level of the order-selection test.

THEOREM 3.1. *Let model* (2.1) *hold with* $f_{12} \equiv 0$, *and assume that the* ε_{ij}*'s are iid with mean* 0 *and finite fourth moment. Define*

$$V_n = \max_{1 \leq i \leq m_n} \frac{1}{N(\Lambda_i)} \sum\sum_{(j,k) \in \Lambda_i} \left(\frac{n \hat{\alpha}_{jk}^2}{\hat{\sigma}^2} \right),$$

where $m_n \to \infty$ *as* $n \to \infty$ *and* $\{\Lambda_1, \Lambda_2, \dots\}$ *is an increasing sequence of subsets of* $\{(i,j)\colon\ i \geq 1,\ j \geq 1\}$ *such that the following holds*:

$$N(\Lambda_i) \geq Bi(\log(i+1))^a \tag{3.6}$$
for all $i \geq 1$ *and some constants* $B > 0$ *and* $a > 1$.

Then the statistic V_n *converges in distribution to the random variable* V *as* $n_1, n_2 \to \infty$, *where*

$$V = \max_{i \geq 1} \frac{1}{N(\Lambda_i)} \sum\sum_{(j,k) \in \Lambda_i} Z_{jk}^2$$

and the Z_{jk}*'s are iid standard normal random variables.*

Under the conditions of Theorem 3.1, $\sum\sum_{(j,k) \in \Lambda_i} Z_{jk}^2 / N(\Lambda_i)$ tends to 1 almost surely as $i \to \infty$, implying that V has support $(1, \infty)$ and thus further confirming that reasonable choices of C are larger than 1.

In the special cases where the $\{\Lambda_i\}$ are defined as in (3.3) and (3.4), respectively, we obtain test statistics

$$V_{1n} = \max_{1 \le \lambda \le m} \frac{1}{\lambda^2} \sum_{i=1}^{\lambda} \sum_{j=1}^{\lambda} \left(\frac{n \hat{\alpha}_{jk}^2}{\hat{\sigma}^2} \right)$$

and

$$V_{2n} = \max_{2 \le \lambda \le m} 2\lambda^{-1}(\lambda - 1)^{-1} \sum_{j=1}^{\lambda-1} \sum_{k=1}^{\lambda-j} \left(\frac{n \hat{\alpha}_{jk}^2}{\hat{\sigma}^2} \right),$$

for $m = \min(n_1 - 1, n_2 - 1)$. Each of (3.3) and (3.4) satisfies (3.6), and hence, under the other conditions of Theorem 3.1, V_{1n} and V_{2n} converge in distribution to

$$V_1 = \max_{\lambda \ge 1} \lambda^{-2} \sum_{j=1}^{\lambda} \sum_{k=1}^{\lambda} Z_{jk}^2$$

and

$$V_2 = \max_{\lambda \ge 2} 2\lambda^{-1}(\lambda - 1)^{-1} \sum_{j=1}^{\lambda-1} \sum_{k=1}^{\lambda-j} Z_{jk}^2,$$

where the Z_{jk}'s are iid $N(0,1)$. We can therefore reject additivity if V_{in} exceeds $C_{i\alpha}$, the $100(1 - \alpha)$th percentage point of V_i, $i = 1, 2$.

Ultimately we will resort to simulation to approximate the cumulative distribution functions F_1 and F_2 of V_1 and V_2. We are unaware of any analytic expressions for these distributions. However, there is a relationship between them and a known distribution which gives some insight into their properties.

Let $Z_1, Z_2, \ldots$ be a sequence of iid $N(0,1)$ random variables. Using results of Spitzer (1956), one can argue that the random variable S,

$$S = \max_{r \ge 1} \frac{1}{r} \sum_{j=1}^{r} Z_j^2,$$

has the following distribution function:

$$F_S(s) = \begin{cases} 0, & s \le 1, \\ \exp\left(-\sum_{j=1}^{\infty} \frac{P\left(\chi_j^2 > js\right)}{j} \right), & s > 1, \end{cases}$$

where χ_j^2 has the χ^2 distribution with j degrees of freedom. The distribution functions F_1 and F_2 can be shown to satisfy

(3.7) $$F_S(t) \le F_i(t), \quad i = 1, 2, \text{ for all } t.$$

Percentiles of S are thus larger than the corresponding ones of either V_1 or V_2. Consequently, one obtains a conservative large-sample test by comparing V_{1n} or V_{2n} with percentiles of S, a few of which are given in Eubank and Hart [(1992), Table 2.1].

It is also interesting to note that $F_1(C)$ and $F_2(C)$ are the limiting probabilities that risk criteria of the form (3.2) choose the correct model when the null hypothesis of additivity is true. In the case $C = 2$, which corresponds to estimation of the risk $E\int(\hat{f}^*_{12} - f^*_{12})^2$, relation (3.7) and Eubank and Hart [(1992), Table 2.1] imply that $F_i(2) \geq 0.71$, $i = 1, 2$. Simulation studies indicate that $F_1(2) \approx 0.78$.

Using arguments like those in Eubank and Hart (1992), we can also obtain the large-sample distribution of $\mathscr{F}$ in (3.5).

THEOREM 3.2. *Let the conditions of Theorem* 3.1 *hold, and define* Z_{jk}, $j, k = 1, 2, \dots,$ *as in that theorem. Set* $\tilde{R}(\lambda) = \sum_{j=1}^{\lambda}\sum_{k=1}^{\lambda}(Z_{jk}^2 - 2)$ *for each* $\lambda \geq 1$, *and define the random variable* $\mathscr{F}^*$ *to be* 0 *if* $\tilde{R}(\lambda) < 0$ *for each* $\lambda \geq 1$ *and to be* $(1/\hat{\lambda}^2)\sum_{j=1}^{\hat{\lambda}}\sum_{k=1}^{\hat{\lambda}}Z_{jk}^2$ *otherwise, where* $\hat{\lambda}$ *is the maximizer of* $\tilde{R}(\lambda)$ *with respect to* λ. *Then* $\mathscr{F}$ *converges in distribution to* $\mathscr{F}^*$ *as* $n_1, n_2 \to \infty$.

3.2. *Power of order-selection tests*. We consider both fixed and local alternatives to the additive model. For f_{12} as in (2.1) define the Fourier coefficients α_{jk} by

$$\alpha_{jk} = 2\int_0^1\int_0^1 f_{12}\big(H_1^{-1}(u), H_2^{-1}(v)\big)\cos(\pi ju)\cos(\pi kv)\, du\, dv,$$
$$j = 1, 2, \dots, k = 1, 2, \dots .$$

Concerning fixed alternatives, we then have the following result, whose proof follows along the lines of that for Theorem 4.1 in Eubank and Hart (1992).

THEOREM 3.3. *Suppose model* (2.1) *holds with* f_{12} *continuous over* $[0, 1] \times [0, 1]$, H_1^{-1} *and* H_2^{-1} *continuous on* $[0, 1]$, *and* $\alpha_{jk} \neq 0$ *for some* $j \geq 1$ *and* $k \geq 1$. *Then the power* (*at any positive nominal level*) *of the test corresponding to either* V_{1n} *or* V_{2n} *tends to* 1 *as* n_1 *and* n_2 *tend to* ∞.

Now consider local alternatives to additivity and define Fourier coefficients for g in (2.6) by

$$\eta_{jk} = 2\int_0^1\int_0^1 g\big(H_1^{-1}(u), H_2^{-1}(v)\big)\cos(\pi ju)\cos(\pi kv)\, du\, dv,$$
$$j = 1, 2, \dots, \quad k = 1, 2, \dots .$$

Arguing as for Theorem 4.2 of Eubank and Hart (1992), we can establish the following result.

THEOREM 3.4. *Suppose model* (2.6) *holds with* g *uniformly continuous over* $[0, 1] \times [0, 1]$, h_1 *and* h_2 *bounded on* $[0, 1]$ *and the* ε_{rs}*'s iid with finite fourth moments. Then the statistic* V_{1n} *converges in distribution* (*as* $n_1, n_2 \to \infty$) *to the random variable*

$$V_1^* = \max_{\lambda \geq 1} \frac{1}{\lambda^2} \sum_{j=1}^{\lambda} \sum_{k=1}^{\lambda} \left(Z_{jk} + \frac{\eta_{jk}}{\sigma}\right)^2,$$

where the Z_{jk}*'s are iid standard normal random variables.*

It is clear that if at least one η_{jk} is nonzero, then the random variable V_1^* is stochastically larger than V_1. This implies that V_{1n} can indeed detect alternatives converging to the null hypothesis at rate $1/\sqrt{n}$. A similar result can also be established for V_{2n}.

3.3. *Extension to higher dimensions.* Order-selection-type tests can also be developed for cases of more than two predictors. For example, if there are $k = 3$ independent variables t_1, t_2 and t_3, the interaction part of the mean function can be expressed as a sum of three bivariate Fourier series expansions and a trivariate Fourier series

$$2^{3/2} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \alpha_{jkl} \cos(\pi j t_1)\cos(\pi k t_2)\cos(\pi l t_3).$$

One can estimate each of the four terms in this expansion using Fourier series estimators as in (3.1) with orders selected using statistics analogous to (3.3). This will lead to parallels of the test statistic V_n in Theorem 3.1. The three bivariate cases follow as before. For the trivariate case, one uses the test statistic

$$V_n = \max_{\Lambda \in \mathscr{L}} \frac{1}{N(\Lambda)} \sum\sum\sum_{(j,k,l)\in\Lambda} \left(\frac{n\hat{\alpha}_{jkl}^2}{\hat{\sigma}^2} \right),$$

with $\mathscr{L}$ a collection of subsets of

$$\Lambda(n_1, n_2, n_3) = \{(j,k,l) \colon 1 \le j \le n_1 - 1, 1 \le k \le n_2 - 1, l \le 1 \le n_3 - 1\}$$

and

$$\hat{\alpha}_{jkl} = \frac{2^{3/2}}{n} \sum_{r=1}^{n_1} \sum_{s=1}^{n_2} \sum_{u=1}^{n_3} (y_{rsu} - \bar{y}_{\dots})\cos(\pi j x_{1r})\cos(\pi k x_{2s})\cos(\pi k x_{3u}).$$

The same arguments used to produce Theorem 3.1 show that under the same conditions V_n converges in distribution to the random variable

$$V = \max_{i \ge 1} \frac{1}{N(\Lambda_i)} \sum\sum\sum_{(j,k,l)\in\Lambda_i} Z_{jkl}^2,$$

where the Z_{jkl}'s are iid standard normal random variables and $\{\Lambda_1, \Lambda_2, \dots\}$ is an increasing sequence of subsets of $\{(i,j,k)\colon i \ge 1, j \ge 1, k \ge 1\}$.

4. Simulation study. In this section we present the results of a Monte Carlo power study. The experiment was designed to be comparable to the one in Barry (1993). Accordingly, a uniform design grid was used with $t_{1r} = (r - 0.5)/n_1$, $r = 1, \dots, n_1$, $t_{2k} = (k - 0.5)/n_2$, $k = 1, \dots, n_2$, and the errors were chosen to be Gaussian. We also used the same regression functions, sample sizes and values of σ as in Barry's paper.

Our experience in preliminary simulations revealed that the type of variance estimator used to rescale the tests can have a nontrivial effect on the power of both Tukey and order-selection tests. Out of several estimators that were initially considered, we found the best performances were obtained from a bivariate difference-type estimator

$$\hat{\sigma}_d^2 = \frac{1}{16(n_1-2)(n_2-2)} \sum_{i=2}^{n_1-1} \sum_{j=2}^{n_2-1} \sum_{r=-1}^{1} \sum_{s=-1}^{1} \delta_{ij}^2(r,s),$$

where $\delta_{ij}(r,s) = y_{ij} - y_{i+r,j+s}$, and from

$$\hat{\sigma}_F^2 = \frac{S(n_1-1, n_2-1) - S(\hat{\lambda}, \hat{\lambda})}{(n_1-1)(n_2-1) - \hat{\lambda}^2},$$

with $\hat{\lambda}$ and $S(\cdot,\cdot)$ defined as in (3.5). These are the two estimators that were used in our simulation experiment. They can both be shown to be consistent for σ^2 under model (2.1) and minimal smoothness conditions on the regression function, whether it is of additive form or not.

Define $R = \sigma^2 T_{1n}^2$ [see (2.8)] and $R_0 = \sum_{r=1}^{n_1}\sum_{k=1}^{n_2}(y_{rk} - \bar{y}_{r\cdot} - \bar{y}_{\cdot k} + \bar{y}_{\cdot\cdot})^2$. The following six test statistics were considered in our simulation:

1. *Tukey*: $(n - n_1 - n_2)R/(R_0 - R)$;
2. *Tukey*$_F$: T_{2n} with $\sigma^2 = \hat{\sigma}_F^2$ and cross-validated cubic smoothing splines for $\hat{f}_1$ and $\hat{f}_2$;
3. *Tukey*$_d$: T_{2n} with $\sigma^2 = \hat{\sigma}_d^2$ and cross-validated cubic smoothing splines for $\hat{f}_1$ and $\hat{f}_2$;
4. V_F: V_{1n} with $\hat{\sigma}^2 = \hat{\sigma}_F^2$,
5. V_d: V_{1n} with $\hat{\sigma}^2 = \hat{\sigma}_d^2$,
6. $\mathscr{F}$ [see (3.5)].

We also included V_{2n} in our study. However, its empirical power properties were essentially the same as those for V_{1n} in the cases we considered. Thus, we will report only on the results for V_{1n}.

Critical values for the *Tukey* statistic were obtained from an F distribution, since with Gaussian errors and an additive model $(n - n_1 - n_2)R/(R_0 - R)$ has the F-distribution with degrees of freedom 1 and $n - n_1 - n_2$. The critical values used for *Tukey*$_F$ and *Tukey*$_d$ were those suggested by the asymptotic theory in Section 2.1. This turned out to produce tests whose empirical levels (in 10,000 replications) did not exceed the nominal level of 0.05 by more than 0.024. For the other five tests, simulation (based on 10,000 replications) was used to obtain approximate critical values for the different sample size configurations.

Empirical power results are given in Tables 1 and 2. For each combination of function, sample size configuration and value of σ, the power results are based on 1000 independent replications. Simulations were done for $(n_1, n_2) =$ (5, 5), (5, 20) and (20, 20), but, to save space, we only show the results for (5, 5) and (5, 20) since all three configurations followed similar patterns. All tests were done at level of significance 0.05.

From Tables 1 and 2 we see that no one test performed uniformly better than any other test. However, with the exception of function (g) the “smooth”

TABLE 1
Percentage of rejections in 1000, 0.05-*level tests when* $(n_1, n_2) = (5, 5)$

σ	***Tukey***	***Tukey***$_F$	***Tukey***$_d$	V_F	V_d	$\mathscr{F}$
			(a) $f(x_1, x_2) = x_1 x_2$			
0.1	93.5	96.9	89.3	93.9	86.3	94.1
0.5	7.2	14.1	12.7	12.2	12.5	11.1
1.0	4.4	7.2	6.7	4.9	5.9	6.6
			(b) $f(x_1, x_2) = \exp(5(x_1 + x_2))/(1 + \exp(5(x_1 + x_2))) - 1$			
0.1	58.1	81.3	76.0	65.9	68.1	59.3
0.5	4.9	10.2	9.9	7.2	7.9	8.2
1.0	5.5	6.8	6.2	5.4	4.8	6.1
			(c) $f(x_1, x_2) = 0.5(1 + \sin(2\pi(x_1 + x_2)))$			
0.1	5.4	46.4	35.0	99.7	96.5	99.7
0.5	5.6	20.3	15.6	14.5	9.9	19.8
1.0	6.4	10.2	9.7	4.4	3.2	9.9
			(d) $f(x_1, x_2) = 64(x_1 x_2)^3(1 - x_1 x_2)^3$			
0.1	79.6	99.6	82.7	99.3	26.7	100
0.5	6.9	14.4	11.3	9.1	5.8	8.9
1.0	5.5	9.3	7.3	5.4	4.8	6.0
			(e) $f(x_1, x_2)$ *is the product of sawtooths*			
0.1	12.2	12.5	3.3	4.4	1.3	4.8
0.5	5.3	7.8	7.2	4.2	4.5	4.9
1.0	5.2	7.1	6.4	3.4	4.0	4.4
			(f) $f(x_1, x_2)$ *equals* 1 *if* $x_1 > 0.5$, $x_2 > 0.5$ *and* 0 *otherwise*			
0.1	100	100	100	100	100	100
0.5	18.0	34.8	29.6	31.0	33.0	34.0
1.0	5.3	12.4	12.2	11.0	13.4	10.8
			(g) $f(x_1, x_2) = (x_1 + x_2)/2 +$ (1 *outlier*)			
0.1	99.5	17.5	33.8	17.4	21.7	14.4
0.5	38.8	18.6	21.4	8.0	12.7	6.1
1.0	8.8	10.0	10.3	4.8	6.8	5.5

Tukey tests ($Tukey_F$ and $Tukey_d$) were superior to the ordinary Tukey test, that is, *Tukey*. The case of function (c) is particularly noteworthy. Here $\|f_1\| = \|f_2\| = 0$, and we cannot appeal to Theorems 2.1 and 2.2 to obtain the large-sample distribution of the Tukey test. However, it is clear from Tables 1 and 2 that *Tukey* has power approximately equal to its level (0.05), while $Tukey_F$ and $Tukey_d$ have substantially larger power. As Barry (1993) notes, main effects are absent from function (c), and hence we expect the power of a Tukey test to be low. This is all the more reason to be encouraged by the performance of $Tukey_F$ and $Tukey_d$.

With the exception of function (g) (which includes an outlier), use of the variance estimator $\hat{\sigma}_F^2$ tended to yield tests with at least slightly larger power than using $\hat{\sigma}_d^2$. This is consistent with the fact that $\hat{\sigma}_F^2$ has smaller asymptotic variance than $\hat{\sigma}_d^2$.

TABLE 2
Percentage of rejections in 1000, 0.05-*level tests when* $(n_1, n_2) = (5, 20)$

σ	*Tukey*	$Tukey_F$	$Tukey_d$	V_F	V_d	$\mathscr{F}$
			(a) $f(x_1, x_2) = x_1 x_2$			
0.1	100	100	100	100	100	100
0.5	11.5	30.8	30.1	35.4	34.3	35.2
1.0	4.8	8.9	8.3	11.6	11.9	12.6
			(b) $f(x_1, x_2) = \exp(5(x_1 + x_2))/(1 + \exp(5(x_1 + x_2))) - 1$			
0.1	99.9	100	100	100	100	99.8
0.5	6.0	15.0	14.4	21.7	20.6	21.7
1.0	5.3	9.0	7.9	8.7	8.1	8.5
			(c) $f(x_1, x_2) = 0.5(1 + \sin(2\pi(x_1 + x_2)))$			
0.1	6.9	70.2	55.9	100	100	100
0.5	5.5	38.9	35.8	98.4	97.5	97.2
1.0	4.8	16.8	15.7	30.1	24.6	33.4
			(d) $f(x_1, x_2) = 64(x_1 x_2)^3(1 - x_1 x_2)^3$			
0.1	100	100	100	100	100	100
0.5	27.2	48.3	45.4	61.9	50.7	59.7
1.0	7.3	13.8	13.5	15.1	12.8	11.3
			(e) $f(x_1, x_2)$ *is the product of sawtooths*			
0.1	84.2	87.2	77.6	18.7	8.0	19.0
0.5	5.7	7.4	6.6	7.0	6.6	6.2
1.0	4.5	6.0	6.1	6.1	5.3	4.5
			(f) $f(x_1, x_2)$ *equals* 1 *if* $x_1 > 0.5$, $x_2 > 0.5$ *and* 0 *otherwise*			
0.1	100	100	100	100	100	100
0.5	80.3	94.5	92.4	96.6	94.7	93.9
1.0	13.9	34.4	32.7	46.7	45.7	44.5
			(g) $f(x_1, x_2) = (x_1 + x_2)/2 +$ (1 *outlier*)			
0.1	62.6	8.8	17.0	4.6	23.4	0.4
0.5	16.4	5.9	6.9	5.3	5.7	4.2
1.0	7.7	6.7	6.2	3.9	4.2	4.7

Among the new tests proposed in this paper, the one most analogous to Barry's (1993) W-test is that based on $\mathscr{F}$. Comparing the results for $\mathscr{F}$ with those for W in Barry (1993), it can be seen that the two tests are roughly comparable except in the case of functions (d) and (e). For function (d) the $\mathscr{F}$-test is clearly superior to the W-test, while for function (e) the W-test has at least a small advantage over $\mathscr{F}$. The performance of V_F was very similar to that of $\mathscr{F}$.

In summary, we have seen evidence that the Tukey tests employing smooth estimates of main effects are preferable to the ordinary Tukey test. We have also seen that the order-selection tests based on V_F and $\mathscr{F}$ are comparable to the W-test of Barry (1993). We agree with Barry that the W- and hence V_F- and $\mathscr{F}$-tests tend to be more reliable than Tukey tests since W, V_F and $\mathscr{F}$ have reasonable power in cases where the Tukey test has extremely low power, that is, cases where the product of the main effect functions is orthogonal to the interaction function. Inasmuch as V_F and $\mathscr{F}$ are simpler to compute, practitioners may sometimes prefer these tests to Barry's W-test.

5. Proofs. In this section we present proofs for results in Sections 2 and 3. We begin by establishing two lemmas. For this purpose we initially assume that our $n = n_1 n_2$ responses follow the model

$$y_{rk} = a + f_{1r} + f_{2k} + \frac{1}{\sqrt{n}} g_{rk} + \varepsilon_{rk}, \qquad r = 1, \dots, n_1,\ k = 2, \dots, n_2. \tag{5.1a}$$

Here the ε_{rk} are iid random variables with $E\varepsilon_{11} = 0$ and $E\varepsilon_{11}^2 = \sigma^2 < \infty$. The parameters a, $\{f_{1r}\}$, $\{f_{2k}\}$ and $\{g_{rk}\}$ are allowed to depend on n, although we will not explicitly indicate this in our notation. They are assumed to satisfy

$$\sum_r f_{1r} = \sum_k f_{2k} = \sum_r g_{rk} = \sum_k g_{rk} = 0, \tag{5.1b}$$

$$\frac{1}{n_i} \sum_{j=1}^{n_i} f_{ij}^2 \to C_i^2 > 0 \quad \text{as } n_i \to \infty \text{ for } i = 1, 2, \tag{5.1c}$$

and

$$\frac{1}{n} \sum_{r=1}^{n_1} \sum_{k=1}^{n_2} f_{1r} f_{2k} g_{rk} \to C_{12} \quad \text{as } n \to \infty, \tag{5.1d}$$

for some finite constants C_1, C_2 and C_{12}. We also require that there exists a nonnegative, finite constant C such that

$$\max_j |f_{ij}| \le C \qquad \forall\ n_i,\ i = 1, 2, \tag{5.1e}$$

and

$$\max_{j,l} |g_{jl}| \le C \qquad \forall\ (n_1, n_2). \tag{5.1f}$$

We will now establish some results about Tukey-type tests under model (5.1). Subsequent developments will then focus on showing that model (2.6) falls into this framework. Our first lemma describes the asymptotic distribution theory for the classical Tukey test.

LEMMA 5.1. *Let*

$$T_n = \sqrt{n}\, \frac{\sum_{r=1}^{n_1} \sum_{k=1}^{n_2} (\bar{y}_{r\cdot} - \bar{y}_{\cdot\cdot})(\bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot})(y_{rk} - \bar{y}_{r\cdot} - \bar{y}_{\cdot k} + \bar{y}_{\cdot\cdot})}{\sum_{r=1}^{n_1} (\bar{y}_{r\cdot} - \bar{y}_{\cdot\cdot})^2 \sum_{k=1}^{n_2} (\bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot})^2}.$$

Assume that model (5.1) *holds*, $E\varepsilon_{11}^j = 0$, $j = 1, 3, 5$, $E\varepsilon_{11}^6 < \infty$ *and that* $n_1/n_2 \to \theta \in (0,1)$ *as* $n \to \infty$. *Then*

$$T_n \to_{\mathscr{D}} N\left(\frac{C_{12}}{(C_1 C_2)^2}, \frac{\sigma^2}{(C_1 C_2)^2}\right)$$

as $n \to \infty$.

PROOF. First define

$$\bar{\varepsilon}_{r\cdot} = \frac{1}{n_2}\sum_k \varepsilon_{rk}, \qquad \bar{\varepsilon}_{\cdot k} = \frac{1}{n_1}\sum_r \varepsilon_{rk}, \qquad \bar{\varepsilon}_{\cdot\cdot} = \frac{1}{n}\sum_r \sum_k \varepsilon_{rk},$$

$$\hat{f}_{1r} = (\bar{y}_{r\cdot} - \bar{y}_{\cdot\cdot}) = f_{1r} + \bar{\varepsilon}_{r\cdot} - \bar{\varepsilon}_{\cdot\cdot} \quad \text{and} \quad \hat{f}_{2k} = (\bar{y}_{\cdot k} - \bar{y}_{\cdot\cdot}) = f_{2k} + \bar{\varepsilon}_{\cdot k} - \bar{\varepsilon}_{\cdot\cdot}.$$

With this notation

$$T_n = \frac{\sqrt{n}\sum_r \sum_k \hat{f}_{1r}\hat{f}_{2k} y_{rk}}{\left(\sum_r \hat{f}_{1r}^2 \sum_k \hat{f}_{2k}^2\right)},$$

since $\sum_{j=1}^{n_i} \hat{f}_{ij} = 0$, $i = 1, 2$.

Note, for example, that

$$n_1^{-1}\sum \hat{f}_{1r}^2 = n_1^{-1}\sum f_{1r}^2 + 2n_1^{-1}\sum f_{1r}(\bar{\varepsilon}_{r\cdot} - \bar{\varepsilon}_{\cdot\cdot}) + n_1^{-1}\sum \bar{\varepsilon}_{\cdot r}^2 - \bar{\varepsilon}_{\cdot\cdot}^2.$$

Now $\bar{\varepsilon}_{\cdot\cdot} = O_p(n^{-1/2})$ and, from Markov's inequality, $n_1^{-1}\sum_r \bar{\varepsilon}_{r\cdot}^2 = O_p(n_1^{-1}) = O_p(n^{-1/2})$. Thus, we conclude that

$$n_i^{-1}\sum_{j=1}^{n_i} \hat{f}_{ij}^2 = C_i^2 + o_p(1), \qquad i = 1, 2,$$

and it suffices to study the limiting properties of $n^{-1/2}\sum_r \sum_k \hat{f}_{1r}\hat{f}_{2k} y_{rk}/(C_1 C_2)^2$.

Let

$$\begin{aligned} E_n &= \frac{1}{\sqrt{n}}\sum_r \sum_k \left(\hat{f}_{1r}\hat{f}_{2k} - f_{1r}f_{2k}\right) y_{rk} \\ &= \frac{1}{\sqrt{n}}\sum_r \sum_k \left(\hat{f}_{1r}\hat{f}_{2k} - f_{1r}f_{2k}\right)\varepsilon_{rk} + \frac{1}{n}\sum_r \sum_k \left(\hat{f}_{1r}\hat{f}_{2k} - f_{1r}f_{2k}\right) g_{rk} \\ &= E_{1n} + E_{2n}. \end{aligned}$$

Writing

$$E_{2n} = \frac{1}{n}\sum_r \sum_k \left(\hat{f}_{1r} - f_{1r}\right)\hat{f}_{2k} g_{rk} + \frac{1}{n}\sum_r \sum_k \left(\hat{f}_{2k} - f_{2k}\right) f_{1r} g_{rk},$$

we see, for example, that

$$\left|\frac{1}{n}\sum_r \sum_k \left(\hat{f}_{1r} - f_{1r}\right)\hat{f}_{2k} g_{rk}\right| \le \max_{r,k}|g_{rk}|\left\{\frac{1}{n_1}\sum_r \left(\hat{f}_{1r} - f_{1r}\right)^2 \left(C_2 + o_p(1)\right)\right\}^{1/2}.$$

Since

$$\frac{1}{n_1}\sum_r\left(\hat{f}_{1r}-f_{1r}\right)^2=\frac{1}{n_1}\sum_r\left(\bar{\varepsilon}_{r\cdot}-\bar{\varepsilon}_{\cdot\cdot}\right)^2=O_p\left(n_1^{-1}\right),$$

we then conclude that $E_{2n}=O_p(n_1^{-1/2}+n_2^{-1/2})$.

Now let us deal with

$$E_{1n}=\frac{1}{\sqrt{n}}\sum_r\sum_k\left(\hat{f}_{1r}-f_{1r}\right)\hat{f}_{2k}\varepsilon_{rk}+\frac{1}{\sqrt{n}}\sum_r\sum_k\left(\hat{f}_{2k}-f_{2k}\right)f_{1r}\varepsilon_{rk}.$$

The proof becomes quite tedious at this point, requiring computation of the means and variances of the two terms in this sum. We therefore point out only the essential steps for the analysis of

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_r\sum_k\left(\hat{f}_{1r}-f_{1r}\right)\hat{f}_{2k}\varepsilon_{rk}&=\frac{1}{\sqrt{n}}\sum_r\sum_k\left(\bar{\varepsilon}_{r\cdot}-\bar{\varepsilon}_{\cdot\cdot}\right)f_{2k}\varepsilon_{rk}\\&\quad+\frac{1}{\sqrt{n}}\sum_r\sum_k\left(\bar{\varepsilon}_{r\cdot}-\bar{\varepsilon}_{\cdot\cdot}\right)\left(\bar{\varepsilon}_{\cdot k}-\bar{\varepsilon}_{\cdot\cdot}\right)\varepsilon_{rk}\\&=A_{1n}+A_{2n}.\end{aligned}$$

We can write

$$A_{1n}=\frac{1}{\sqrt{n}}\sum_r\sum_k f_{2k}\bar{\varepsilon}_{r\cdot}\varepsilon_{rk}-\frac{1}{\sqrt{n}}\bar{\varepsilon}_{\cdot\cdot}\sum_r\sum_k f_{2k}\varepsilon_{rk}.$$

The last term is $O_p(n^{-1/2})$ since $\bar{\varepsilon}_{\cdot\cdot}=O_p(n^{-1/2})$, $E\sum_r\sum_k f_{2k}\varepsilon_{rk}=0$ and $\operatorname{Var}\sum_r\sum_k f_{2k}\varepsilon_{rk}=\sigma^2 n_1\sum_k f_{2k}^2=O(n)$. For the first term we find that

$$E\sum_r\sum_k f_{2k}\bar{\varepsilon}_{r\cdot}\varepsilon_{rk}=0$$

and

$$\operatorname{Var}\left(\sum_r\sum_k f_{2k}\bar{\varepsilon}_{r\cdot}\varepsilon_{rk}\right)=n_1\frac{4\left(E\varepsilon_{11}^4-\sigma^4\right)+n_2\sigma^2}{n_2}\left(C_2+o(1)\right).$$

Therefore, $A_{1n}=O_p(n_2^{-1/2})$.

Finally,

$$A_{2n}=\frac{1}{\sqrt{n}}\sum_r\sum_k\varepsilon_{rk}\bar{\varepsilon}_{r\cdot}\bar{\varepsilon}_{\cdot k}-\frac{n_2}{\sqrt{n}}\bar{\varepsilon}_{\cdot\cdot}\sum_r\bar{\varepsilon}_{r\cdot}^2-\frac{n_1}{\sqrt{n}}\bar{\varepsilon}_{\cdot\cdot}\sum_k\bar{\varepsilon}_{\cdot k}^2+\sqrt{n}\,\bar{\varepsilon}_{\cdot\cdot}^3.$$

The last three terms in this sum are $O_p(n_1^{-1})$, $O_p(n_2^{-1})$ and $O_p(n^{-1})$, respectively. For the first term one finds that

$$E\sum_r\sum_k\bar{\varepsilon}_{r\cdot}\bar{\varepsilon}_{\cdot k}\varepsilon_{rk}=0$$

and, using the assumption that $E\varepsilon_{11}^6<\infty$,

$$\operatorname{Var}\left(\sum_r\sum_k\bar{\varepsilon}_{r\cdot}\bar{\varepsilon}_{\cdot k}\varepsilon_{rk}\right)=O(1).$$

Consequently, $A_{2n}=O_p(n^{-1/2})$.

To complete the proof, we apply the Lindeberg–Feller theorem to show that $n^{-1/2}\Sigma_r\Sigma_k f_{1r}f_{2k}\varepsilon_{rk} \to_{\mathscr{D}} N(0, \sigma^2 C_1^2 C_2^2)$. Since $n^{-1}\Sigma_r\Sigma_k f_{1r}f_{2k}g_{rk} \to C_{12}$, the lemma follows from an application of Slutsky's thoerem. □

The next lemma gives the limiting properties of a variant of the Tukey test where the estimators and main effects $\{f_{1r}\}$ and $\{f_{2k}\}$ exhibit certain smoothness properties.

LEMMA 5.2. *Assume that model* (5.1) *holds and that* $\{\hat{f}_{1r}\}$ *and* $\{\hat{f}_{2k}\}$ *are estimators of the main effects that satisfy the following*:

(i) $$\sum_j \hat{f}_{ij} = o_p(1), \qquad i = 1,2;$$

(ii) $$\max_j \left| \hat{f}_{i(j+1)} - f_{i(j+1)} - \left(\hat{f}_{ij} - f_{ij}\right)\right| = o_p(n_i^{-1}), \qquad i = 1,2;$$

(iii) $$\max_j \left| \hat{f}_{i(j+1)} - \hat{f}_{ij} \right| = O_p(n_i^{-1}), \qquad i = 1,2;$$

(iv) $$\max_{i=1,2} \left| \hat{f}_{in_i} - f_{in_i} \right| = o_p(1);$$

(v) $$\sum_r \left(\hat{f}_{ir} - f_{ir}\right)^2 = o_p(n_i).$$

Also assume that $n_1/n_2 \to \theta \in (0,1)$,

$$\max_j | f_{i(j+1)} - f_{ij} | = O(n_i^{-1}), \qquad i = 1,2,$$

and that $a = O(\sqrt{n})$ *in* (5.1). *Then*,

$$T_n = \frac{\sqrt{n}\,\Sigma_{r=1}^{n_1}\Sigma_{k=1}^{n_2}\hat{f}_{1r}\hat{f}_{2k}\left(y_{rk} - \hat{f}_{1r} - \hat{f}_{2k}\right)}{\Sigma_{r=1}^{n_1}\hat{f}_{1r}^2\,\Sigma_{k=1}^{n_2}\hat{f}_{2k}^2}$$
$$\to_{\mathscr{D}} N\left(\frac{C_{12}}{(C_1C_2)^2}, \frac{\sigma^2}{(C_1C_2)^2}\right).$$

PROOF. Using condition (v), we can show that $n_i^{-1}\Sigma_j\hat{f}_{ij}^2 = C_i^2 + o_p(1)$, $i = 1,2$, and consequently we need only study

$$\frac{1}{\sqrt{n}}\sum_r\sum_k \hat{f}_{1r}\hat{f}_{2k}\left(y_{rk} - \hat{f}_{1r} - \hat{f}_{2k}\right) = \frac{1}{\sqrt{n}}\sum_r\sum_k \hat{f}_{1r}\hat{f}_{2k}y_{rk} + o_p(1)$$

as a result of (i), (v) and the fact that $\sqrt{n} = O(n_i)$.

Now

$$\frac{1}{\sqrt{n}}\sum_r\sum_k \hat{f}_{1r}\hat{f}_{2k}y_{rk}$$
$$= \frac{1}{\sqrt{n}}\sum_r\sum_k \hat{f}_{1r}\hat{f}_{2k}\varepsilon_{rk} + \frac{1}{n}\sum_r\sum_k \hat{f}_{1r}\hat{f}_{2k}g_{rk} + \frac{1}{\sqrt{n}}a\sum_r\hat{f}_{1r}\sum_k\hat{f}_{2k} + o_p(1).$$

The third term in this expression is $o_p(1)$ from (i) and $a = O(\sqrt{n})$. For the second term we have

$$\begin{aligned}\left|\frac{1}{n}\sum_r\sum_k\left(\hat{f}_{1r}\hat{f}_{2k} - f_{1r}f_{2k}\right)g_{rk}\right| &\le \max|g_{rk}|\frac{1}{\sqrt{n}}\left[\left(\sum_r\left(\hat{f}_{1r} - f_{1r}\right)^2\sum_k\hat{f}_{2k}^2\right)^{1/2}\right.\\ &\qquad\left.+\left(\sum_k\left(\hat{f}_{2k} - f_{2k}\right)^2\sum_r f_{1r}^2\right)^{1/2}\right]\\ &= o_p(1),\end{aligned}$$

and hence $n^{-1}\Sigma_r\Sigma_k\hat{f}_{1r}\hat{f}_{2k}g_{rk} \to_P C_{12}$.

Finally, we must deal with $E_n = n^{-1/2}\Sigma_r\Sigma_k(\hat{f}_{1r}\hat{f}_{2k} - f_{1r}f_{2k})\varepsilon_{rk} = E_{1n} + E_{2n}$, where

$$E_{1n} = n^{-1/2}\sum_r\sum_k\left(\hat{f}_{1r} - f_{1r}\right)\hat{f}_{2k}\varepsilon_{rk}$$

and

$$E_{2n} = n^{-1/2}\sum_r\sum_k\left(\hat{f}_{2k} - f_{2k}\right)f_{1r}\varepsilon_{rk}.$$

Define

$$S_{rk} = \sum_{i=1}^{r}\sum_{j=1}^{k}\varepsilon_{ij} \tag{5.2}$$

with $S_{00} = S_{r0} = S_{0k} = 0$, $r = 1,\dots,n_1$, $k = 1,\dots,n_2$, and set

$$-u_r = \hat{f}_{1(r+1)} - f_{1(r+1)} - \left(\hat{f}_{1r} - f_{1r}\right) \quad\text{and}\quad v_k = \hat{f}_{2k} - \hat{f}_{2(k+1)}.$$

Then, using summation by parts, we obtain

$$\begin{aligned}E_{1n} = {}& \frac{1}{\sqrt{n}}\left(\hat{f}_{1n_1} - f_{1n_1}\right)\left[\hat{f}_{2n_2}S_{n_1n_2} + \sum_{k=1}^{n_2-1}v_kS_{n_1k}\right]\\ &+\frac{1}{\sqrt{n}}\hat{f}_{2n_2}\sum_{r=1}^{n_1-1}u_rS_{rn_2} + \frac{1}{\sqrt{n}}\sum_{r=1}^{n_1-1}\sum_{k=1}^{n_2-1}u_rv_kS_{rk}.\end{aligned} \tag{5.3}$$

By Donsker's theorem, $\max_{r,k}|S_{rk}| = O_p(\sqrt{n})$. Using this along with conditions (ii)–(iv), we see that each term in (5.3) is $o_p(1)$. The same approach can be used to see that $E_{2n} = o_p(1)$.

The above arguments show that T_n has the same limiting distribution as $(n^{-1/2}\Sigma_r\Sigma_k f_{1r}f_{2k}\varepsilon_{rk} + C_{12})/(C_1C_2)^2$. The lemma then follows from the Lindeberg–Feller theorem. □

PROOF OF THEOREMS 2.1 AND 2.2. We will now use Lemmas 5.1 and 5.2 to prove, respectively, Theorems 2.1 and 2.2. The results follow once we show that model (2.6) can be formulated as in (5.1).

Define

$$a = \mu + n^{-3/2} \sum_i \sum_j g(t_{1i}, t_{2j}),$$

$$f_{1r} = f_1(t_{1r}) - n_1^{-1} \sum_j f_1(t_{1j}) + n^{-1/2}\left(n_2^{-1} \sum_j g(t_{1r}, t_{2j}) - n^{-1} \sum_i \sum_j g(t_{1i}, t_{2j})\right),$$

$$f_{2k} = f_2(t_{2k}) - n_2^{-1} \sum_j f_2(t_{2j}) + n^{-1/2}\left(n_1^{-1} \sum_j g(t_{1j}, t_{2k}) - n^{-1} \sum_i \sum_j g(t_{1i}, t_{2j})\right)$$

and

$$g_{rk} = g(t_{1r}, t_{2k}) - n_2^{-1} \sum_j g(t_{1r}, t_{2j}) - n_1^{-1} \sum_j g(t_{1j}, t_{2k}) + n^{-1} \sum_i \sum_j g(t_{1i}, t_{2j}).$$

Using (2.6a)–(2.6d) and the facts that $f_i \in C^1[0,1]$, $i = 1, 2$, and $g \in C^1([0,1] \times [0,1])$, we can show that

$$f_{ij} = f_i(t_{ij}) + O(n_i^{-1}), \qquad i = 1, 2, \tag{5.4a}$$

$$g_{rk} = g(t_{1r}, t_{2j}) + O(n_1^{-1} + n_2^{-1} + n^{-1}) \tag{5.4b}$$

and

$$a = \mu + O(n^{-1/2}). \tag{5.4c}$$

Thus, conditions (5.1b)–(5.1f) hold with $C_i^2 = \|f_i\|^2$ and $C_{12} = \langle g, f_1 f_2 \rangle$.

Theorem 2.1 can now be obtained directly as an application of Lemma 5.1. Upon setting $\hat{f}_{ij} = \hat{f}_i(t_{ij})$ in Lemma 5.2 and using arguments like those for (5.4) along with the continuous differentiability of f_1, f_2 and g_{12}, it is straightforward to see that conditions (i) and (ii) of Theorem 5.2 imply conditions (i)–(iv) of Lemma 5.2. The condition $\max_j |f_{i(j+1)} - f_{ij}| = O(n_i^{-1})$ is satisfied as a result of (5.4a) and the differentiability of f_i, $i = 1, 2$. □

PROOF OF THEOREM 3.1. The proof of Theorem 3.1 is analogous to that of Theorem 3.1 in Eubank and Hart (1992) for univariate regression. Ironically, the proof in this two-dimensional setting is simpler, not requiring use of the Hájek–Rényi–type inequality needed in the one-dimensional case. This is consistent with the findings of Sain, Baggerly and Scott (1994), which imply that cross-validation algorithms for smoothing parameter selection tend to perform better with increasing dimension.

By Slutsky's theorem, it is enough to prove the theorem with $\hat{\sigma}$ in V_n replaced by σ, which, w.l.o.g., we take to be 1. Let $\{k_n\}$ be an increasing and unbounded sequence of positive integers such that $k_n < m_n$ for all n, and write $V_n = \max(U_{1n}, U_{2n})$, where

$$U_{1n} = \max_{1 \le i \le k_n} \frac{1}{N(\Lambda_i)} \sum_{(r,s)\in\Lambda_i} n\hat{\alpha}_{rs}^2 \quad \text{and} \quad U_{2n} = \max_{k_n < i \le m_n} \frac{1}{N(\Lambda_i)} \sum_{(r,s)\in\Lambda_i} n\hat{\alpha}_{rs}^2.$$

We now consider

$$P(V_n \le x) = P(U_{1n} \le x, U_{2n} - 1 \le x - 1)$$

for $x > 1$ and must show that this probability converges to $P(V \le x)$ for V defined in Theorem 3.1.

We will first establish that $P(U_{2n} - 1 \le x - 1) \to 1$ as $n \to \infty$. To do this, it is sufficient to establish that

$$(5.5) \qquad P\left(\bigcup_{i=k_n+1}^{m_n} \left\{ \frac{1}{N(\Lambda_i)} \left| \sum\sum_{(r,s)\in\Lambda_i} \left(n\hat{\alpha}_{rs}^2 - 1\right) \right| > (x-1) \right\} \right) \to 0.$$

By Markov's inequality, the probability in (5.5) is bounded by

$$(5.6) \qquad \frac{1}{(x-1)^2} \sum_{i=k_n+1}^{m_n} E\left| \sum\sum_{(r,s)\in\Lambda_i} \left(n\hat{\alpha}_{rs}^2 - 1\right) \right|^2 \frac{1}{N^2(\Lambda_i)}.$$

It is straightforward to show that the expectation in (5.6) is bounded by $AN(\Lambda_i)$, uniformly in i, for some positive constant A. Hence (5.6) is bounded by

$$\frac{A}{(x-1)^2} \sum_{i=k_n+1}^{m_n} \frac{1}{N(\Lambda_i)},$$

which tends to 0 as $n \to \infty$ because k_n is unbounded and

$$N(\Lambda_i) \ge Bi(\log(i+1))^a \quad \text{for } B > 0 \text{ and } a > 1.$$

To complete the proof, we need to show that U_{1n} converges in distribution to V. Let $\{Z_{jk}: j \ge 1, k \ge 1\}$ be iid $N(0,1)$ random variables, and define U_n exactly as is U_{1n} but with $n\hat{\alpha}_{rs}^2$ replaced by Z_{rs}^2. We must show that $P(U_{1n} \le x) - P(U_n \le x) \to 0$. The proof will then be finished since $P(U_n \le x)$ is monotone decreasing in n and hence has a limit, which by definition we take to be $P(V \le x)$.

Arguing as in Eubank and Hart (1992), a Berry–Esseen–type result of Bhattacharya and Ranga Rao (1976) implies that

$$(5.7) \qquad |P(U_{1n} \le x) - P(U_n \le x)| \le C\left(N(\Lambda_{k_n})\right) N^2(\Lambda_{k_n}) \frac{E(\varepsilon_{11}^4)}{\sqrt{n}},$$

where $C(m)$ is a constant that depends only on m. We are free to let k_n tend to infinity as slowly as we want, and hence we can choose $\{k_n\}$ such that the right-hand side of (5.7) tends to 0 as $n \to \infty$. □

Acknowledgments. This paper arose from discussions that occurred at a meeting in Keystone, Colorado, in June 1992. The meeting was partially funded by National Cancer Institute Grant CA-57030. The authors appreciate the comments of two referees and an Associate Editor.

REFERENCES

BARRY, D. (1993). Testing for additivity of a regression function. *Ann. Statist.* **21** 235–254.

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

EUBANK, R. L. and HART, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria. *Ann. Statist.* **20** 1412–1425.

FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004.

HALL, P., KAY, J. W. and TITTERINGTON, D. M. (1990). Asymptotically optimal difference based estimation of variance in nonparametric regression. *Biometrika* **77** 521–528.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. Chapman and Hall, London.

MÜLLER, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data*. Springer, New York.

SAIN, S. R., BAGGERLY, K. A. and SCOTT, D. W. (1994). Cross-validation of multivariate densities. *J. Amer. Statist. Assoc.* **89** 807–817.

SPIEGELMAN, C. and WANG, C. Y. (1994). Detecting interactions using low dimensional searches in high dimensional data. *Chemometrics and Intelligent Laboratory Systems* **23** 293–299.

SPITZER, F. (1956). A combinatorial lemma and its applications to probability theory. *Trans. Amer. Math. Soc.* **82** 323–339.

STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13** 689–705.

TUKEY, J. (1949). One degree of freedom test for non-additivity. *Biometrics* **5** 232–242.

WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory Probab. Appl.* **5** 302–305.

R. L. EUBANK
J. D. HART
DEPARTMENT OF STATISTICS
TEXAS A & M UNIVERSITY
COLLEGE STATION, TEXAS 77843

D. G. SIMPSON
DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
CHAMPAIGN, ILLINOIS 61820

L. A. STEFANSKI
DEPARTMENT OF STATISTICS
NORTH CAROLINA STATE UNIVERSITY
RALEIGH, NORTH CAROLINA 27695