A Nonparametric Approach to Measuring and Testing Curvature

Jason ABREVAYA
Department of Economics, Purdue University, W. Lafayette, IN 47907-2056 (*abrevaya@purdue.edu*)

Wei JIANG
Graduate School of Business, Finance and Economics Division, Columbia University, New York, NY 10027 (*wj2006@columbia.edu*)

This article considers the problem of testing curvature (e.g., linearity, concavity, convexity) in a multivariate nonparametric regression model. A measure of curvature, called the *simplex statistic*, that does not require bandwidth choice and is easy to compute, is introduced. A global test of curvature based on the simplex statistic is also introduced. Localized versions of the test, which require smoothing parameters, are shown to be consistent against more general alternatives than the global test. In the univariate case, the local test of concavity (convexity) is consistent against all nonconcave (nonconvex) alternatives. The simplex statistic can also be used in the context of a partially linear regression model. Applications to examining the curvature of the experience-earnings profile and testing the "style timing" of mutual funds are considered.

KEY WORDS: Concavity; Convexity; Nonparametric test; U-statistics.

1. INTRODUCTION

Concavity and convexity are important concepts in microeconomic theory. The concavity or convexity of a relationship (function) may be predicted by theory; for example, theory tells us that cost functions are concave in input price and that profit functions are convex in output price. The concavity or convexity of a function can have important implications for the way in which economic agents will behave; for example, concavity (convexity) of an individual's utility function corresponds to risk aversion (loving). By definition, concavity or convexity of a function is directly associated with whether there are decreasing returns or increasing returns to a given factor. For instance, to determine whether the returns to education (e.g., on wages) are increasing or decreasing in the level of education, we would want to know if the relationship between wages and education is convex or concave.

Tests of convexity or concavity have appeared in a wide range of empirical studies. For example, there have been studies on testing the curvature between investment and Tobin's q (e.g., Barnett and Sakellaris 1998), between firm value and product price (Borenstein and Farrell 1999), and between mutual fund performance and subsequent money inflow (Chevalier and Ellison 1997). The convex or concave relationship in these examples have important implications for corporate managers' investment strategies and money managers' risk-taking behavior.

Despite the important role of concavity and convexity in economic theory and applications, there has been surprisingly little work in the econometrics literature on statistical testing of concavity and convexity. Although a large literature on nonparametric specification testing has emerged, only a few studies have considered tests for curvature. Moreover, the existing nonparametric methods for curvature are applicable only for the univariate case. There appear to be no nonparametric methods for testing concavity and convexity of a regression function having more than one covariate.

In empirical work, the general approach for testing curvature is to allow nonlinearity in the regression function and perform appropriate hypothesis tests after estimation. Natural candidates for parametric methods that allow nonlinearity are polynomials or linear splines. Although easy to implement, these methods are subject to model misspecification when researchers do not have clear prior information about the shape of the underlying functions. Because a parametric test is essentially a joint test of curvature and parametric functional form, the conclusion of such a test is unclear due to possible misspecification of functional form.

Nonparametric specification of the regression function circumvents this difficulty by avoiding auxiliary assumptions on the functional form and focusing directly on the curvature of the underlying regression function. The related nonparametric curvature tests broadly fall into three categories. The first category includes a wide range of general nonparametric specification tests (see references in Ellison and Ellison 2000). These tests can detect nonlinearity but cannot be used directly to reject a null hypothesis of concavity or convexity. The second category includes tests based on nonparametric estimation of functions subject to convex/concave restriction (e.g., Hildreth 1954; Mammen 1991; Wang 1992; Yatchew 1992). (Relevant surveys of shape-restricted nonparametric estimation include Diack and Thomas-Agnan 1998 and Matzkin 1994.) Studies along this line have focused on the univariate case, and the asymptotic distribution of the estimates when the dimension of covariates is bigger than one is not yet known. The third, and most promising, category includes tests based on nonparametric estimation of second derivatives that have been developed in the statistics literature. Diack and Thomas-Agnan (1998) and Diack (2001) constructed tests of concavity and convexity using cubic-spline estimation, where the test statistics are based on the estimated second derivative evaluated at the spline knots. Dümbgen and Spokoiny (2001) provided a consistent nonparametric test of concavity and convexity that does not require

Journal of Business & Economic Statistics
January 2005, Vol. 23, No. 1
DOI 10.1198/073500104000000316

the choice of any smoothing parameters. Their "multiscale" sup-norm method essentially combines test statistics using all possible bandwidth choices. Unfortunately, the tests of Diack and Thomas-Agnan (1998), Diack (2001), and Dümbgen and Spokoiny (2001) are applicable only to univariate nonparametric models with homoscedastic error disturbances. The test of Dümbgen and Spokoiny (2001) also requires normality of the error disturbances.

This article makes several contributions to the existing literature. First, we introduce new (global) descriptive measures of curvature, called *simplex statistics*. These statistics (and their asymptotic standard errors) do not require bandwidth choice and are easy to compute. Second, because tests based on the global simplex statistics are not consistent against general alternatives, we propose a test based on a discretized process of localized versions of the simplex statistics. In the univariate case, the local test is consistent against general alternatives like those given by Diack and Thomas-Agnan (1998), Diack (2001), and Dümbgen and Spokoiny (2001), but avoids the assumptions of homoscedasticity and normality made for those tests. The proposed test also avoids estimation of the error variance, which is required by Diack and Thomas-Agnan (1998), Diack (2001), and Dümbgen and Spokoiny (2001). The test based on the simplex statistic for the multivariate case appears to be the first proposal for testing curvature in a multivariate nonparametric regression setting.

To fix ideas, we consider the following nonparametric regression model:

$$y_i = f(\mathbf{x}_i) + \epsilon_i, \qquad E[\epsilon_i|\mathbf{x}_i] = 0, \qquad i = 1, \dots, n, \qquad (1)$$

where y_i is a scalar, $\mathbf{x}_i$ is a $1 \times m$ vector $(x_{1i}, \dots, x_{mi})$, and ϵ_i is an error disturbance. The salient features of our nonparametric approach are that it leaves the functional form of f in (1) unspecified to avoid model misspecification, and it avoids smoothed nonparametric estimation of f entirely and thus should be easy to implement. The curvature measure that we propose is based on the defining property of convex/concave functions. Recall that the convexity of a univariate function f is defined as follows: $f(x)$ is convex in x if and only if for any $\lambda \in [0, 1]$ and $x_1, x_2 \in \mathbb{R}$, it is true that $\lambda f(x_1) + (1 - \lambda)f(x_2) \geq f(\lambda x_1 + (1 - \lambda)x_2)$. For concavity, the inequality is reversed. Graphically, a convex (concave) function f lies below (above) the line segment connecting $[x_1, f(x_1)]$ and $[x_2, f(x_2)]$ on the range between the two values. An analogous graphical relationship holds in higher dimensions. To allow visualization in three dimensions, consider the bivariate case ($m = 2$). For three distinct points $\mathbf{x}_i$, $\mathbf{x}_j$, and $\mathbf{x}_k$ in $\mathbb{R}^2$, a convex (concave) function lies above (below) the plane going through the points on the range defined by the triangle formed by the three points.

To use this idea in the context of actual data, suppose that there are four distinct data points $\mathbf{x}_i$, $\mathbf{x}_j$, $\mathbf{x}_k$, and $\mathbf{x}_l$ such that $\mathbf{x}_l$ is in the interior of the closed triangle with vertices $\mathbf{x}_i$, $\mathbf{x}_j$, and $\mathbf{x}_k$, denoted by $\mathbf{x}_l \in \Delta(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$. Then there exists a unique solution (a_1, a_2, a_3) such that

$$\mathbf{x}_l = a_1\mathbf{x}_i + a_2\mathbf{x}_j + a_3\mathbf{x}_k, \qquad (2)$$
$$a_1 > 0, a_2 > 0, a_3 > 0, a_1 + a_2 + a_3 = 1.$$

If all quadruples $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l)$ that satisfy (2) are collected from the observed data, then the relationship

$$a_1 f(\mathbf{x}_i) + a_2 f(\mathbf{x}_j) + a_3 f(\mathbf{x}_k) \geq f(\mathbf{x}_l)$$

always holds if f is convex (inequality reversed for concave f). From (1), each y_i differs from $f(\mathbf{x}_i)$ by an error disturbance, meaning that the relationship

$$a_1 y_i + a_2 y_j + a_3 y_k \geq y_l$$

should usually (i.e., more often than not) hold if the function f in (1) is convex (inequality reversed for concave f). If f is linear, then there should be no tendency for y_l to be either greater or smaller than $a_1 y_i + a_2 y_j + a_3 y_k$. This graphical intuition is formalized and extended to the general multivariate case in Section 2.

The remainder of the article is organized as follows. Section 2 introduces the simplex statistic for measuring the curvature of f from the model in (1), and discusses asymptotic results and consistency of the related hypothesis tests. Section 3 considers tests based on localized versions of the simplex statistics. These tests are consistent against more general alternatives. Section 4 reports results from Monte Carlo simulations using the proposed tests, including comparisons with existing tests in the literature. Section 5 discusses how the simplex statistic can be used in the context of a partially linear regression model. Section 6 applies the methodology in two empirical applications, one examining the curvature of the experience-earnings profile and the other measuring the "style timing" performance of asset-allocation mutual funds.

2. THE SIMPLEX STATISTICS

2.1 Definitions of Convexity/Concavity and the Test Statistics

We first define convexity (concavity) of a function as follows.

Definition 1. A function $f: A \to \mathbb{R}$, defined on the convex set $A \subseteq \mathbb{R}^m$, is *convex* (*concave*) if

$$a_1 f(\mathbf{x}_1) + a_2 f(\mathbf{x}_2) + \cdots + a_m f(\mathbf{x}_m) \geq (\leq) f(a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_m\mathbf{x}_m) \qquad (3)$$

for any collections of vectors $\mathbf{x}_1, \dots, \mathbf{x}_m \in A$ and nonnegative numbers $a_1, \dots, a_m$ such that $a_1 + a_2 + \cdots + a_m = 1$. If the inequality in (3) is strict, then f is strictly convex (concave) on A.

This definition is used in several economics textbooks (e.g., Mas-Colell, Whinston, and Green 1995; Moore 1999) and is equivalent to other commonly used definitions (see Hardy, Littlewood, and Polya 1988, sec. 3.4). Geometrically, the definition says that the center of gravity of arbitrarily weighted points of the function lies on or above the function if the function is convex. Viewing the weights as probabilities, notice that (3) is a version of Jensen's inequality for convex (concave) functions (Hardy et al. 1988).

Before considering the testable implications of this definition in the context of observed data, we first formalize the sampling scheme.

Assumption 1. An iid sample $\{(\mathbf{x}_i, \epsilon_i)\}_{i=1}^n$ is drawn from the joint distribution of the random variables $(\mathbf{x}, \epsilon)$, where ϵ is symmetrically distributed about 0 (conditional on $\mathbf{x}$). The observed sample is $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$, where y_i is generated according to the model in (1).

This assumption allows for conditional heteroscedasticity of the error disturbances. The conditional symmetry assumption on the error disturbances is necessary for the proposed method of measuring curvature. For instance, if the error disturbances are conditionally asymmetric with the asymmetry varying over $\mathbf{x}$ values, then the data might appear to have a convex or concave relationship even if the regression function is truly linear. Figure 1 depicts such a violation in the univariate x case. The conditional asymmetry in the error disturbance at a single x value (denoted x^*) makes f appear to be concave; in this figure, f is linear with symmetric errors for all $x \neq x^*$ and asymmetric errors at x^* with median$(\epsilon|x^*) < E(\epsilon|x^*)$.

Denote $v_i \equiv (y_i, \mathbf{x}_i)$ to be the data associated with a given observation. For a sample of size n, there are $\binom{n}{m+2}$ ordered $(m+2)$-tuples $\{\mathbf{v}_{t_1}, \ldots, \mathbf{v}_{t_{m+2}}\}$, where $t_1 < \cdots < t_{m+2}$. The implication of (3) can be tested on a given $(m+2)$-tuple if the set $\{t_1, \ldots, t_{m+2}\}$ can be partitioned into $\{[t_{m+2}]\}$ and $\{[t_1], \ldots, [t_{m+1}]\}$ such that the following system has a unique solution:

$$\mathbf{x}_{[t_{m+2}]} = a_1\mathbf{x}_{[t_1]} + a_2\mathbf{x}_{[t_2]} + \cdots + a_{m+1}\mathbf{x}_{[t_{m+1}]}, \qquad (4)$$
$$a_1 > 0, a_2 > 0, \ldots, a_{m+1} > 0, a_1 + a_2 + \cdots + a_{m+1} = 1.$$

The set of all interior linear combinations of $\mathbf{x}_{[t_1]}, \ldots, \mathbf{x}_{[t_{m+1}]}$ is called an m-dimensional *simplex*, denoted by $\Delta(\mathbf{x}_{[t_1]}, \ldots, \mathbf{x}_{[t_{m+1}]})$, and $\mathbf{x}_{[t_{m+2}]}$ is in the interior of $\Delta(\mathbf{x}_{[t_1]}, \ldots, \mathbf{x}_{[t_{m+1}]})$. Based on (3), we call $\{\mathbf{v}_{t_1}, \ldots, \mathbf{v}_{t_{m+2}}\}$ a *convex* $(m+2)$-tuple if $a_1 y_{[t_1]} + \cdots + a_{m+1} y_{[t_{m+1}]} > y_{[t_{m+2}]}$, a *concave* one if $a_1 y_{[t_1]} + \cdots + a_{m+1} y_{[t_{m+1}]} < y_{[t_{m+2}]}$, and a *linear* one if $a_1 y_{[t_1]} + \cdots + a_{m+1} y_{[t_{m+1}]} = y_{[t_{m+2}]}$. If ϵ is continuously distributed, then the occurrence of a linear $(m+2)$-tuple is a zero-probability event.

To ensure that the number of "qualified" $(m+2)$-tuples grows large as n grows large, we make the following assumption on the distribution of the covariates.

Assumption 2. For $(m+2)$ iid draws $\mathbf{x}_1, \ldots, \mathbf{x}_{m+2}$ of the random variable $\mathbf{x}$, the probability that $\mathbf{x}_{m+2}$ is within the interior of the simplex $\Delta(\mathbf{x}_1, \ldots, \mathbf{x}_{m+1})$ is greater than 0.

For the univariate case ($m = 1$), Assumption 2 merely says that there are at least three disjoint regions with positive probability; even three distinct point masses would be sufficient. For $m \geq 2$, Assumption 2 rules out linear dependence among covariates (in which case no simplex would have an interior). The assumption also rules out situations where simplices have interiors but no points lie within these interiors; examples for $m = 2$ would include covariate distributions with points lying on a circle or a square.

In the univariate case ($m = 1$), any observation-triple having (three) distinct x values will be used to test the implication of (3). Figure 2 shows a concave triple in a scatterplot of data. Note that the intermediate point lies above the segment that would connect the outer points. Equivalently, the slope connecting the leftmost and intermediate points is larger than the slope connecting the rightmost and intermediate points. The figure highlights only one observation-triple, but there are many observation-triples within the scatterplot, some of them concave triples and some of them convex triples.

In the multivariate case ($m \geq 2$), checking whether or not a point $\mathbf{x} \in \mathbb{R}^m$ is in the interior of the simplex $\Delta(\mathbf{x}_1, \ldots, \mathbf{x}_{m+1})$ amounts to solving the following system of linear equations:

$$\mathbf{x} = a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \cdots + a_{m+1}\mathbf{x}_{m+1};$$
$$a_1 + a_2 + \cdots + a_{m+1} = 1. \qquad (5)$$

If the system is nondegenerate, then $m+1$ equations with $m+1$ unknowns will produce a unique solution. The point $\mathbf{x}$ is in the interior of the simplex $\Delta(\mathbf{x}_1, \ldots, \mathbf{x}_{m+1})$ if and only if $a_1, \ldots, a_{m+1}$ are all positive. It is straightforward to show that if $\mathbf{x}$ is inside $\Delta(\mathbf{x}_1, \ldots, \mathbf{x}_{m+1})$, then $\mathbf{x}_j$ is not inside $\Delta(\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}, \mathbf{x}_{j+1}, \ldots, \mathbf{x}_{m+1})$ for any $j \in \{1, 2, \ldots, m+1\}$. The statistic for testing curvature will not consider any $(m+2)$-tuple for which there is no point lying in the interior of the simplex formed by the remaining points.

Let $\mathcal{S}$ denote the set of all $(m+2)$-tuples of $1 \times m$ vectors such that one vector lies in the interior of the simplex spanned

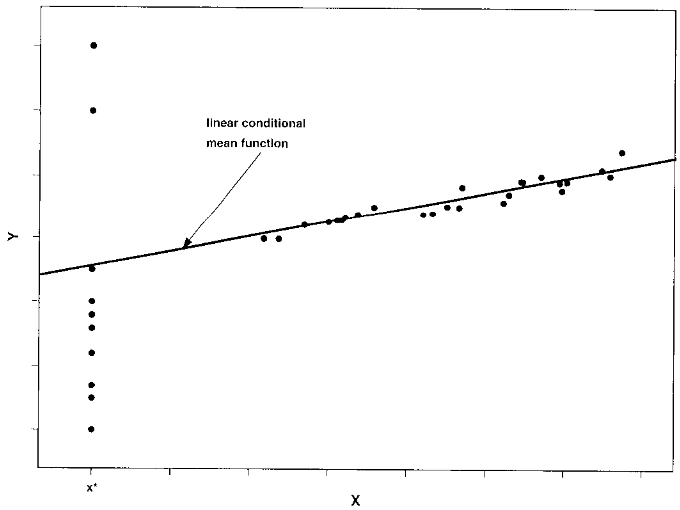

Figure 1. A Violation of the Error-Disturbance Assumptions.

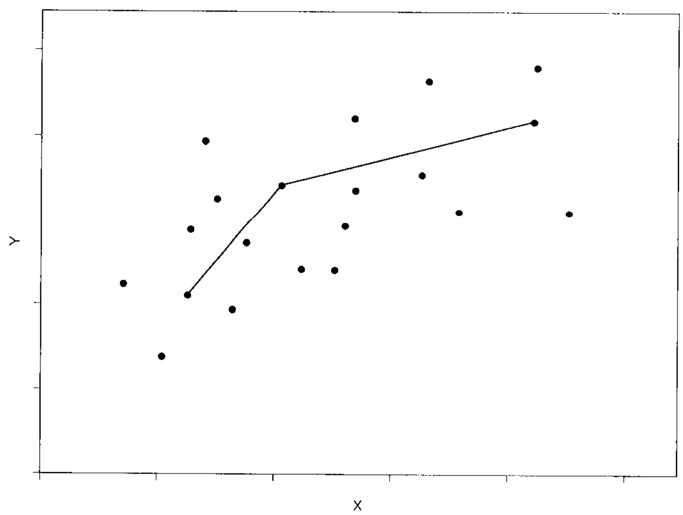

Figure 2. A "Concave" Observation-Triple.

by the remaining vectors. Let N be the number of $(m+2)$-tuples from the observed sample belonging to $\mathcal{S}$; that is,

$$N \equiv \sum_{1 \le t_1 < \cdots < t_{m+2} \le n} \mathbb{1}\big(\{\mathbf{x}_{t_1}, \ldots, \mathbf{x}_{t_{m+2}}\} \in \mathcal{S}\big), \qquad (6)$$

where $\mathbb{1}(\cdot)$ is the indicator function (equal to 1 if the argument is true and 0 otherwise). The more linearly correlated the independent variables $\mathbf{x}$ are among one another, the lower N will be given n and m. The largest possible value for N is $\binom{n}{m+2}$. Note that Assumption 2 guarantees that $N \to \infty$ as $n \to \infty$.

The simplex statistic, denoted by U_n, is defined as

$$U_n \equiv \binom{n}{m+2}^{-1} \sum_{1 \le t_1 < \cdots < t_{m+2} \le n} \mathbb{1}\big(\{\mathbf{x}_{t_1}, \ldots, \mathbf{x}_{t_{m+2}}\} \in \mathcal{S}\big) \times \operatorname{sign}\big(a_1 y_{[t_1]} + \cdots + a_{m+1} y_{[t_{m+1}]} - y_{[t_{m+2}]}\big), \qquad (7)$$

where $\operatorname{sign}(q) = 1$ if $q > 0$, -1 if $q < 0$, and 0 if $q = 0$ and $a_1, \ldots, a_{m+1}$ is the unique solution to (4). (Although $a_1, \ldots, a_{m+1}$ obviously depend on the vectors, we have suppressed the dependence on the $\mathbf{x}$'s to simplify notation.) It is straightforward to show that

$$U_n = \binom{n}{m+2}^{-1} \big[\#\text{ of convex } (m+2)\text{-tuples} - \#\text{ of concave } (m+2)\text{-tuples}\big].$$

Alternatively, rescaling U_n by $N^{-1}\binom{n}{m+2}$ yields the statistic

$$U_n' = N^{-1} \sum_{1 \le t_1 < \cdots < t_{m+2} \le n} \mathbb{1}\big(\{\mathbf{x}_{t_1}, \ldots, \mathbf{x}_{t_{m+2}}\} \in \mathcal{S}\big) \times \operatorname{sign}\big(a_1 y_{[t_1]} + \cdots + a_{m+1} y_{[t_{m+1}]} - y_{[t_{m+2}]}\big), \qquad (8)$$

which has a ready interpretation of

$$U_n' = \text{percentage of convex } (m+2)\text{-tuples} - \text{percentage of concave } (m+2)\text{-tuples}.$$

The term "percentage" refers to the percentage among the $(m+2)$-tuples within $\mathcal{S}$. Because U_n and U_n' differ by only a factor of $N^{-1}\binom{n}{m+2}$, the asymptotic properties for U_n automatically apply to U_n' up to scale. [In the univariate case with no ties in the x values, $U_n = U_n'$, because $N = \binom{n}{3}$.] Practically, one can use either (7) or (8) for the purpose of testing curvature; (8) has a more intuitive interpretation in terms of the probability of convex $(m+2)$-tuples in excess of that of concave ones. For discussion of asymptotic properties, however, we focus on (7), because it has a standard U-statistic form. We call U_n nad U_n' *simplex statistics* in the remainder of the article. Although we focus on hypothesis testing based on these statistics, researchers can use the U_n' statistic as an overall descriptive statistic for the curvature displayed in their data.

2.2 Asymptotic Results

In this section we discuss the asymptotic properties of the simplex statistic defined in (7). All proofs are provided in the Appendix. First, we apply the law of large numbers for U-statistics (e.g., Serfling 1980, thm. 5.4A). For notational ease, define

$$w(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}) \equiv a_1 y_{[1]} + a_2 y_{[2]} + \cdots + a_{m+1} y_{[m+1]} - y_{[m+2]}. \qquad (9)$$

The probability limit of the simplex statistic is given by

Theorem 1. If Assumption 1 holds, then

$$U_n \xrightarrow{p} \theta \equiv E\big[\mathbb{1}(\{\mathbf{x}_1, \ldots, \mathbf{x}_{m+2}\} \in \mathcal{S}) \times \operatorname{sign}(w(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}))\big], \qquad (10)$$

where $\mathbb{1}(\cdot)$ is the indicator function and $\{a_1, \ldots, a_{m+1}\}$ uniquely solves (4) if $\{\mathbf{x}_1, \ldots, \mathbf{x}_{m+2}\} \in \mathcal{S}$ and can be assigned arbitrary values otherwise.

Deriving the asymptotic normality of U_n requires some additional notation. The *kernel function* of U_n is given by

$$h(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}) \equiv \mathbb{1}(\{\mathbf{x}_1, \ldots, \mathbf{x}_{m+2}\} \in \mathcal{S}) \cdot \operatorname{sign}(w(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2})). \qquad (11)$$

Note that the kernel function is symmetric in its arguments by the definitions of $\mathcal{S}$ and $w(\cdot)$. Then define

$$\tilde{h}(\mathbf{v}) \equiv E\big(h(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{m+2})|\mathbf{v}\big) \qquad (12)$$

and

$$\varsigma \equiv E\big[\big(\tilde{h}(\mathbf{v}) - \theta\big)^2\big]. \qquad (13)$$

The asymptotic normality of U_n follows from Hoeffding (1948):

Theorem 2. If Assumptions 1 and 2 hold, then

$$\sqrt{n}(U_n - \theta) \xrightarrow{d} \mathrm{N}\big(0, (m+2)^2\varsigma\big).$$

A consistent estimator of ς is

$$\widehat{\varsigma} \equiv n^{-1} \sum_{t_1=1}^{n} \left(\binom{n-1}{m+1}^{-1} \times \sum_{\substack{t_2 < \cdots < t_{m+2} \\ t_2 \ne t_1, \ldots, t_{m+2} \ne t_1}} h\big(\mathbf{v}_{t_1}, \mathbf{v}_{t_2}, \ldots, \mathbf{v}_{t_{m+2}}\big) - U_n\right)^2. \qquad (14)$$

The scalar multiplying ς in the asymptotic variance expression is the square of the order of the U-statistic [i.e., $(m+2)^2$ for the $(m+2)$-order U-statistic U_n].

2.3 Hypothesis Testing

In the absence of the error disturbance, the equivalence between convexity/concavity of the f function and convexity/concavity of the $(m+2)$-tuples is given by Definition 1. If f is convex (concave), then all $(m+2)$-tuples are convex (concave). Under conditional symmetry of the error disturbance (assumed in Assumption 1), the following theorem relates the sign of θ [defined in (10)] directly to the curvature of the function f.

Theorem 3. If Assumptions 1 and 2 hold, then:

a. $f(\cdot)$ is linear $\Rightarrow \theta = 0$.
b. $f(\cdot)$ is convex $\Rightarrow \theta \geq 0$.
c. $f(\cdot)$ is concave $\Rightarrow \theta \leq 0$.

The quantity θ is strictly positive (negative) for a convex (concave) f that is strictly convex (concave). Conditional symmetry of the error distributions guarantees that the linear combination $a_1\epsilon_1 + \cdots + a_{m+1}\epsilon_{m+1} - \epsilon_{m+2}$ is also symmetric, conditional on $(\mathbf{x}_1, \ldots, \mathbf{x}_{m+2})$.

The simplex statistic U_n in (7) is nothing but the sample analog of θ. Therefore, Theorem 3 provides natural hypothesis testing of curvature based on U_n. Under the null hypothesis that f is linear, $\theta = 0$ (under suitable assumptions), and the asymptotic distribution of U_n (under the null as well as under the alternative) is given by Theorem 2. Based on Theorem 3, a test of $H_0: \theta = 0$ is consistent against any strictly convex or strictly concave alternative, a test of $H_0: \theta \leq 0$ is consistent against a strictly convex alternative, and a test of $H_0: \theta \geq 0$ is consistent against a strictly concave alternative. Let z_α be the α-percentile value of the standard normal distribution. Table 1 summarizes the hypothesis tests based on the simplex statistic.

The null hypothesis of linearity can also be tested using other nonparametric specification tests; we provide comparisons of the simplex statistic test with one such nonparametric specification test in the Monte Carlo section. One advantage of the simplex statistic is that it does not require the choice of window width to which the results of nonparametric specification tests can be sensitive. Further, nonparametric specification tests cannot in general directly tell whether the nonlinearity takes the form of convexity or concavity. The drawback of the simplex statistic is that a test of $H_0: \theta = 0$ is not consistent against all nonlinear alternatives, because $\theta = 0$ does not imply linearity. A function can have convex and concave regions that offset each other and give a statistic close to 0. Similarly, $\theta \geq 0$ ($\theta \leq 0$) does not imply that the function is convex (concave) either. To develop a test that is consistent against local convexity/concavity, we consider a localized version of the simplex statistic in the next section.

3. TESTS BASED ON LOCALIZED SIMPLEX STATISTICS

As mentioned in the previous section, a test of linearity based on the simplex statistic is not necessarily consistent against all nonlinear alternatives. In this respect, it is similar to the symmetry test of Davis and Quade (1978) and Randles, Fligner, Policello, and Wolfe (1980) and the rank correlation test of Kendall (1938) (based on Kendall's "tau"). A test of constancy of the regression function f (in the univariate case) based on the global Kendall's tau statistic is not consistent against all nonconstant alternatives (because f may be increasing for some x's and decreasing for other x's). Ghosal, Sen, and van der Vaart (2000) recently developed tests based on *localized* Kendall's tau statistics, which, unlike the test based on the global Kendall's tau, are consistent against all general alternatives. In particular, the tests of constancy are consistent against all nonconstant alternatives, and the tests of monotonicity are consistent against all nonmonotonic alternatives.

In this section we adopt the approach of Ghosal et al. (2000) by defining localized versions of the simplex statistics and proposing test statistics based on the "process" of localized simplex statistics. The main departure from Ghosal et al. (2000) is to consider a discretized, rather than a continuous, process. There are several reasons for using a discretized process in this context. First, whereas the process considered by Ghosal et al. (2000) is approximated by a stationary Gaussian process (due to the underlying tau statistic and the assumption of homoscedasticity), the continuous process of localized simplex statistics cannot in general be approximated by a stationary Gaussian process *even* in the case of homoscedasticity. As a result, the classical extremal theory based on stationary Gaussian processes cannot be applied, making derivation of a test statistic's asymptotic distribution extremely difficult. Second, the test statistic based on the discretized version does not suffer from two empirical drawbacks of the continuous version: (1) the continuous version is undersized (i.e., does not reject often enough under the null), as seen in table 1 of Ghosal et al. (2000) and independent simulations by the authors, and (2) the size of the continuous version seems to be sensitive to the bandwidth choice.

The covariate-distribution assumption (Assumption 2) is strengthened somewhat to derive consistency results for the localized tests. In particular, Assumption 2 is replaced by the following.

Assumption 3. The vector $\mathbf{x}$ has a continuous joint distribution with positive density bounded below by $\eta > 0$ over the support $\mathcal{X} \equiv [c_1, d_1] \times \cdots \times [c_m, d_m]$ for some $c_1, d_1, \ldots, c_m, d_m \in \mathbb{R}$.

Table 1. Summary of Hypothesis Tests Using the Simplex Statistic

To test	Null hypothesis	Reject if (level $= \alpha$)	Consistent against
Linearity	$\theta = 0$	$\left\lvert \frac{\sqrt{n}U_n}{(m+2)\sqrt{\xi}} \right\rvert > z_{\alpha/2}$	Convex or concave alternatives
Concavity	$\theta \leq 0$	$\frac{\sqrt{n}U_n}{(m+2)\sqrt{\xi}} > z_\alpha$	Convex alternatives
Convexity	$\theta \geq 0$	$\frac{\sqrt{n}U_n}{(m+2)\sqrt{\xi}} < -z_\alpha$	Concave alternatives

The continuity of $\mathbf{x}$'s distribution is required so that asymptotically there will be sufficient data locally to construct a test with power against general alternatives. For any $j = 1, \ldots, m$, denote the window width in the jth dimension by h_j. Corresponding to this window width, let g_j denote the number of nonoverlapping windows that fit in the jth dimension, given by $g_j = \lfloor \frac{d_j - c_j}{2h_j} \rfloor$. Without loss of generality, we assume that $h_1 = \cdots = h_m = h$ to simplify notation. (One can linearly transform the data so that the components of $\mathbf{x}$ have equal variance, in which case applying the uniform window width is reasonable. One just needs to transform the data back to the original metric when interpreting the results.) The simplex statistic is evaluated (locally) at $G \equiv \prod_{j=1}^{m} g_j$ locations. The evaluation locations are denoted by $\mathbf{x}_1^*, \ldots, \mathbf{x}_G^*$, where any convention for the ordering can be adopted. Finally, we denote the subpopulation for which $\mathbf{x}$ values fall in a local window as

$$V_h(\mathbf{x}^*) \equiv \{(y, \mathbf{x}) : x_1^* - h < x_1 < x_1^* + h, \ldots, x_m^* - h < x_m < x_m^* + h\}, \quad (15)$$

and let $p_h(\mathbf{x}^*)$ be the number of observations in the set $V_h(\mathbf{x}^*)$.

Define the *localized simplex statistic* at a given $\mathbf{x}^*$ as

$$U_{n,h}(\mathbf{x}^*) \equiv \binom{p_h(\mathbf{x}^*)}{m+2}^{-1} \sum_{t_1 < \cdots < t_{m+2}} \left(\prod_{k=1}^{m+2} K_h(\mathbf{x}_{t_k} - \mathbf{x}^*) \right) \times \mathbb{1}\left(\{\mathbf{x}_{t_1}, \ldots, \mathbf{x}_{t_{m+2}}\} \in \mathcal{S}\right) \times \operatorname{sign}\left(w(\mathbf{v}_{t_1}, \ldots, \mathbf{v}_{t_{m+2}})\right), \quad (16)$$

where $w(\cdot)$ is defined in (9) and $K_h(\mathbf{v}) \equiv K(\mathbf{v}/h)$, with the kernel function K satisfying the following assumption.

Assumption 4. The kernel function $K : \mathbb{R}^m \to \mathbb{R}$ satisfies the following conditions: (a) $K(\cdot)$ is a nonnegative, symmetric, and continuous function with support $[-1, 1]^m$, and (b) $K(\cdot)$ is twice continuously differentiable on $(-1, 1)^m$.

When K is the uniform kernel function, $U_{n,h}(\mathbf{x}^*)$ is just the simplex statistic applied to the subset of data whose $\mathbf{x}_i$ values fall into the local window of $\mathbf{x}^*$.

Note that the kernel function associated with the localized simplex statistic is

$$k(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}; \mathbf{x}, h) \equiv \left(\prod_{k=1}^{m+2} K_h(\mathbf{x}_k - \mathbf{x}) \right) \times \mathbb{1}(\{\mathbf{x}_1, \ldots, \mathbf{x}_{m+2}\} \in \mathcal{S}) \cdot \operatorname{sign}(w(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2})). \quad (17)$$

The following theorem gives the asymptotic distribution of the localized simplex statistic for a *fixed* window width.

Theorem 4. If Assumptions 1, 3, and 4 hold and $h > 0$ is fixed, then

$$\sqrt{p_h(\mathbf{x}^*)}\left(U_{n,h}(\mathbf{x}^*) - \theta(\mathbf{x}^*, h)\right) \xrightarrow{d} \mathrm{N}\left(0, (m+2)^2 \varsigma(\mathbf{x}^*, h)\right), \quad (18)$$

where

$$\theta(\mathbf{x}, h) \equiv E\left[k(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}; \mathbf{x}, h) \mid \mathbf{v}_1, \ldots, \mathbf{v}_{m+2} \in V_h(\mathbf{x})\right], \quad (19)$$

$$\tilde{k}(\mathbf{v}; \mathbf{x}, h) \equiv E\left[k(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{m+2}; \mathbf{x}, h) \mid \mathbf{v}_2, \ldots, \mathbf{v}_{m+2} \in V_h(\mathbf{x})\right], \quad (20)$$

and

$$\varsigma(\mathbf{x}, h) = E\left[\left(\tilde{k}(\mathbf{v}; \mathbf{x}, h) - \theta(\mathbf{x}, h)\right)^2 \middle| \mathbf{v} \in V_h(\mathbf{x})\right]. \quad (21)$$

A consistent estimator of $\varsigma(\mathbf{x}^*, h)$ is formed analogously to (14),

$$\widehat{\varsigma}(\mathbf{x}^*, h) \equiv p_h(\mathbf{x}^*)^{-1} \sum_{t_1=1}^{p_h(\mathbf{x}^*)} \left(\binom{p_h(\mathbf{x}^*) - 1}{m+1}^{-1} \times \sum_{\substack{t_2 < \cdots < t_{m+2} \\ t_2 \neq t_1, \ldots, t_{m+2} \neq t_1}} k(\mathbf{v}_{t_1}, \ldots, \mathbf{v}_{t_{m+2}}; \mathbf{x}^*, h) - U_{n,h}(\mathbf{x}^*) \right)^2. \quad (22)$$

We have considered the case of fixed h first to emphasize that computation of the localized statistic and its associated asymptotic variance (in Thm. 4) is very similar to the global statistic (in Thm. 2). The main implementational difference is that the localized statistics need to be evaluated at multiple locations.

The test statistics that detect local departure from linearity, concavity, or convexity are based on the extreme values of standardized discretized statistics. First, we normalize $U_{n,h}(\mathbf{x}^*)$ by its standard deviation,

$$\widetilde{U}_{n,h}(\mathbf{x}^*) \equiv \frac{\sqrt{p_h(\mathbf{x}^*)}\, U_{n,h}(\mathbf{x}^*)}{(m+2)\sqrt{\widehat{\varsigma}(\mathbf{x}^*, h)}}. \quad (23)$$

When f is linear, $\theta(\mathbf{x}, h) = 0$ for all $\mathbf{x}$ and h. In this case we obtain the following result.

Theorem 5. If Assumptions 1, 3, and 4 hold, $nh^m \to \infty$, and f is linear, then the standardized statistic $\widetilde{U}_{n,h}(\mathbf{x}^*) \xrightarrow{d} \mathrm{N}(0, 1)$.

The condition $nh^m \to \infty$ (along with the positive-density assumption on the covariates) implies that $p_h(\mathbf{x}^*) \to \infty$, so that asymptotic normality follows as in Theorem 4. There are G such standardized discretized statistics. (Note that Thm. 5 allows for $h \to 0$.) Due to the nonoverlapping windows over which they are evaluated and the bounded-support assumption on K (Assumption 4), these statistics will be asymptotically independent of each other (under suitable assumptions on the bandwidth rate).

The two relevant extreme-value statistics are defined as

$$M_n \equiv \max_{j=1,\ldots,G} \{\widetilde{U}_{n,h}(\mathbf{x}_j^*)\} \quad (24)$$

and

$$m_n = \min_{j=1,\ldots,G} \{\widetilde{U}_{n,h}(\mathbf{x}_j^*)\}. \quad (25)$$

Intuitively, a large positive value for M_n should provide evidence against concavity, whereas a large negative value for m_n should provide evidence against convexity. In either case, the use of extreme values provides the ability to detect local violations of concavity or convexity. Suppose that one is interested in testing the null hypothesis of convexity,

$$H_0 : f(\mathbf{x}) \text{ is convex on } \mathcal{X} \equiv [c_1, d_1] \times \cdots \times [c_m, d_m]. \quad (26)$$

The null hypothesis will be rejected (at a specified level α) when the test statistic m_n is below some critical value (which depends on α). This test will be consistent against alternatives for which f is concave on a subset of $\mathcal{X}$.

Theorem 6 provides the appropriate critical values for hypothesis testing, based on the asymptotic distributions of M_n and m_n when f is linear.

Theorem 6. If Assumptions 1, 3, and 4 hold, $h \to 0$, $nh^m / \log h^{-1} \to \infty$, and f is linear, then

$$a_n(M_n - b_n) \xrightarrow{d} W_1, \quad (27)$$

$$a_n(-m_n - b_n) \xrightarrow{d} W_1, \quad (28)$$

and

$$a_n\big(\max\{|M_n|, |m_n|\} - b_n\big) \xrightarrow{d} W_2, \quad (29)$$

where

$$a_n = \sqrt{2 \log G}, \qquad b_n = \sqrt{2 \log G} - \frac{\log \log G + \log 4\pi}{2\sqrt{2 \log G}}, \quad (30)$$

where W_1 is a (standard) type I extreme value distribution with $\Pr(W_1 < w) = \exp(-\exp(-w))$, and W_2 is a type I extreme distribution with $\Pr(W_2 < w) = \exp(-2\exp(-w))$.

The proof of Theorem 6 is based on classical extreme value theory (see, e.g., Leadbetter, Lindgren, and Rootzen 1983). From Theorem 6, we can construct the critical regions that have asymptotic significance level α [for any $\alpha \in (0, 1)$] as

$$M_n \ge b_n - \frac{\log \log(1-\alpha)^{-1}}{a_n} = \sqrt{2 \log G} - \frac{\log \log G + \log 4\pi + 2 \log \log(1-\alpha)^{-1}}{2\sqrt{2 \log G}}, \quad (31)$$

$$m_n \le -b_n + \frac{\log \log(1-\alpha)^{-1}}{a_n} = -\sqrt{2 \log G} + \frac{\log \log G + \log 4\pi + 2 \log \log(1-\alpha)^{-1}}{2\sqrt{2 \log G}}, \quad (32)$$

and

$$\max\{|M_n|, |m_n|\} \ge b_n - \frac{\log(\log(1-\alpha)^{-1}/2)}{a_n} = \sqrt{2 \log G} - \frac{\log \log G + \log \pi + 2 \log \log(1-\alpha)^{-1}}{2\sqrt{2 \log G}}, \quad (33)$$

where a_n and b_n are as defined in (30). Note that $G \sim h^{-m}$, so that rate conditions on G could have been used in Theorem 6 instead.

Finally, we establish the consistency of the localized test against general alternatives in Theorem 7.

Theorem 7. Suppose that Assumptions 1, 3, and 4 hold, $h \to 0$, and $nh^{4+m} / \log h^{-1} \to \infty$. Further, suppose that f is twice continuously differentiable on support $\mathcal{X}$, and denote $H(f; \mathbf{x})$ to be the $m \times m$ Hessian matrix of f evaluated at $\mathbf{x} \in \mathcal{X}$. Then the following results hold:

a. Consistency of concavity test. Suppose that for some $\delta > 0$, the eigenvalues of $H(f; \mathbf{x})$ are greater than δ for all $\mathbf{x} \in [L_1, U_1] \times \cdots \times [L_m, U_m]$, where $c_1 \le L_1 < U_1 \le d_1, \ldots, c_m \le L_m < U_m \le d_m$. Then the test based on M_n and the critical value in (31) is consistent at any level $\alpha \in (0, 1)$.

b. Consistency of convexity test. Suppose that for some $\delta > 0$, the eigenvalues of $H(f; \mathbf{x})$ are less than $-\delta$ for all $\mathbf{x} \in [L_1, U_1] \times \cdots \times [L_m, U_m]$, where $c_1 \le L_1 < U_1 \le d_1, \ldots, c_m \le L_m < U_m \le d_m$. Then the test based on m_n and the critical value in (32) is consistent at any level $\alpha \in (0, 1)$.

c. Consistency of linearity test. Suppose that for some $\delta > 0$, the eigenvalues of $H(f; \mathbf{x})$ are greater than δ for all $\mathbf{x} \in [L_1, U_1] \times \cdots \times [L_m, U_m]$ or the eigenvalues of $H(f; \mathbf{x})$ are less than $-\delta$ for all $\mathbf{x} \in [L_1, U_1] \times \cdots \times [L_m, U_m]$, where $c_1 \le L_1 < U_1 \le d_1, \ldots, c_m \le L_m < U_m \le d_m$. Then the test based on $\max\{|M_n|, |m_n|\}$ and the critical value in (33) is consistent at any level $\alpha \in (0, 1)$.

We briefly note that the assumption of twice differentiability is not needed for consistency of the proposed tests. This assumption, which likely is innocuous in most applications, allows the alternatives to be stated in terms of the Hessian matrices rather than $\theta(\mathbf{x}, h)$ quantities. The former is a more familiar way to view concavity/convexity.

The window width h needs to shrink to 0 at a rather slow rate, proportional to $n^{-\gamma}$ for $\gamma < \frac{1}{4+m}$, to ensure consistency of the test. This rate reflects the difficulty of detecting local curvature, because any function appears approximately linear as the local window shrinks. Apart from the $\log h^{-1}$, this rate is analogous to those in kernel estimation of second-order derivatives. For example, when $m = 1$, the rate of convergence for kernel estimation of second-order derivatives is $n^{-1/5}$. As such, a test to detect local convexity/concavity based on kernel estimation would likely require a similar window-width convergence rate.

The linearity test based on (33) is consistent against the alternative that f is strictly convex or concave on a positive-measure subset of the support. The concavity test based on (31) [or convexity test based on (32)] is consistent against the alternative that f is strictly convex (or concave) on a positive-measure subset of the support. For the univariate case, the test of linearity is consistent against any alternative with a local nonlinearity, and the test of convexity (concavity) is consistent against any alternative with a local nonconvexity (nonconcavity). Abrevaya and Jiang (2003) also considered the consistency of the univariate test against a sequence of local alternatives. The local power of the test is found to be similar to that of Diack (2001). In the multivariate case, however, the tests are not necessarily consistent if the violation of linearity (or convexity/concavity) involves regions that are nonlinear but neither convex nor concave; for example, the function $f(x_1, x_2) = x_1^2 + x_2^2 - 5x_1x_2$ is neither convex or concave on any $[c_1, d_1] \times [c_2, d_2] \subseteq \mathbb{R}^2$.

Given that there are alternative ways of defining concavity and convexity (i.e., other than Definition 1), there may be alternative ways of testing concavity and convexity as well. For instance, a function f is convex at a point $\mathbf{x}$ if the minimum eigenvalue of the Hessian matrix evaluated at $\mathbf{x}$ is nonnegative. It seems likely that one could construct a test based on nonparametric estimates of the Hessian matrix at various locations (similar to the grid approach used for the simplex statistic). A cubic-spline approach to this problem would represent a multivariate extension of the work of Diack (2001) and Diack and Thomas-Agnan (1998). Though our experience with the cubic-spline approach in the univariate setting has not been promising (see Sec. 4), the extension to the multivariate case is worthy of future research. The benefit of analyzing the Hessian matrices directly is that the researcher can learn something about the magnitudes of the second derivatives of f. In contrast, the simplex approach is easier to implement, because it does not require estimation of f or its derivatives, but the simplex statistics are not informative about the magnitudes of the second derivatives of f.

4. MONTE CARLO SIMULATIONS

In this section we examine the performance of the global and local versions of the simplex statistic in univariate and bivariate $\mathbf{x}$ designs. We conducted the simulations on small- to moderate-sized samples (ranging from 25 to 1,000 observations). For the univariate case, we compare our approach with the testing alternatives proposed by Diack (2001) and Dümbgen and Spokoiny (2001). For the multivariate case, we compare our approach with the nonparametric specification test of Ellison and Ellison (2000), because there is no existing nonparametric alternative for directly testing concavity or convexity.

4.1 One-Covariate Case

We consider several different Monte Carlo designs for the univariate case, with x drawn from a uniform distribution on $[0, 100]$ in each:

$$\text{(D1)} \quad y = 100(x/100)^{\alpha} + \epsilon, \tag{34}$$

$$\text{(D2)} \quad y = 400(x/100 - .5)^3 + \epsilon, \tag{35}$$

and

$$\text{(D3)} \quad y = \begin{cases} 10c(x/100 - .5)^3 \\ \quad - c\exp[-100(x/100 - .25)^2] \\ \quad + \epsilon, \quad \text{if } x < 50 \\ .1c(x/100 - .5) \\ \quad - c\exp[-100(x/100 - .25)^2] \\ \quad + \epsilon, \quad \text{if } x > 50. \end{cases} \tag{36}$$

The value of α in design (D1) determines the global curvature of the regression function, with $\alpha = 1$ linear, $\alpha > 1$ convex, and $\alpha < 1$ concave. Figure 3 shows the functions for (D1) corresponding to $\alpha = 1$, $\alpha = 1.2$, and $\alpha = 1.4$, which are the values used in the simulations that follow. Designs (D2) and (D3) have both concave and convex regions. Design (D2) is a cubic function, concave for $x < 50$ and convex for $x > 50$. Design (D3), taken from Ghosal et al. (2000), is mostly concave but has a sharp convex "dip" around $x = 25$. [The constant c is chosen so that the function $f(\cdot)$ ranges from 0 to 100 as in designs (D1) and (D2).]

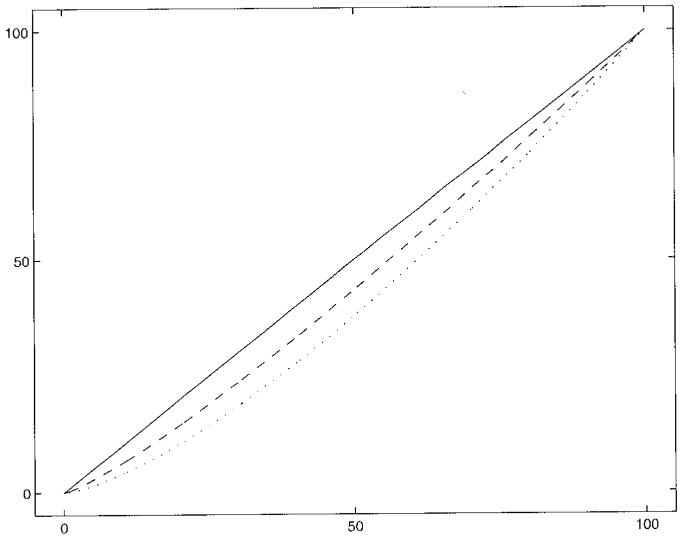

Figure 3. Shape of Functions for Monte Carlo Design (— $\alpha = 1$; - - $\alpha = 1.2$; $\cdots$ $\alpha = 1.4$).

Table 2 examines the size of the global linearity test by testing $H_0: \theta = 0$ for design (D1) with $\alpha = 1$. The sample size ($n = 25, 50, 100, 200$), the error distribution (normal or uniform), and the error standard deviation (10, 20, or proportional to $\sqrt{x}$) are varied, yielding 20 designs. For each design, we performed 1,000 simulations. The table reports the mean of U_n, the mean of the estimated asymptotic standard errors, and the empirical standard deviation of the U_n values. The final two columns report the rejection rates for the 10% and 5% levels. The nonparametric nature of the test is evident, because it is robust to nonnormality and conditional heteroscedasticity of the disturbance term. Overall, the rejection rates based on the asymptotic critical values are very encouraging. Others (e.g., Ellison and Ellison 2000) have pointed out that nonparametric specification tests tend to converge to their asymptotic distributions relatively slowly; as a result, the size of these tests tends to be quite erratic, even with hundreds of observations. In contrast, the asymptotic distribution for the simplex statistic is quite accurate even for very small sample sizes.

Next, we examine the size and power of the local version of the simplex test. To have some meaningful benchmarks, we compare the performance with the methods of Diack (2001) and Dümbgen and Spokoiny (2001). Table 3 summarizes the results from simulations using designs (D1) (with $\alpha = 1, 1.2, 1.4$), (D2), and (D3). The tests are abbreviated in the table, with "S" for simplex test, "DS" for Dümbgen and Spokoiny (2001), and "D" for Diack (2001). Because the DS and D tests are only applicable to homoscedastic error disturbances, we restrict attention to designs with homoscedastic ϵ. For all simulations, ϵ is drawn from a normal distribution, with standard deviation 10 in designs (D1) and (D2) and standard deviation 20 in design (D3). Rejection rates are based on 2,000 simulations for each design and method. The rejection rates correspond to the

Table 2. Size of the Univariate Simplex Test Under the Null of Linearity

n	Error distribution	SD	Mean bias	Average asymptotic standard error	Empirical SD	10% rejection	5% rejection
25	Normal	10	.0026	.0550	.0543	10.2	5.8
	Normal	20	−.0014	.0548	.0540	9.8	6.1
	Uniform	10	.0008	.0559	.0541	10.7	5.4
	Uniform	20	−.0004	.0559	.0550	10.4	5.8
	Normal	$5\sqrt{x}$	.0002	.0535	.0512	9.3	5.1
50	Normal	10	−.0006	.0344	.0348	11.1	5.7
	Normal	20	−.0026	.0346	.0347	11.0	6.9
	Uniform	10	−.0010	.0353	.0350	9.5	5.3
	Uniform	20	−.0006	.0355	.0338	9.9	4.9
	Normal	$5\sqrt{x}$	.0032	.0329	.0319	10.3	4.8
100	Normal	10	−.0016	.0233	.0227	10.2	4.7
	Normal	20	−.0006	.0232	.0232	10.4	5.9
	Uniform	10	.0010	.0238	.0238	11.0	4.9
	Uniform	20	.0018	.0238	.0235	9.5	5.5
	Normal	$5\sqrt{x}$	−.0006	.0223	.0215	9.4	4.3
200	Normal	10	.0004	.0161	.0154	9.2	4.4
	Normal	20	0	.0161	.0160	10.4	4.8
	Uniform	10	0	.0165	.0164	9.7	4.5
	Uniform	20	.0004	.0165	.0164	10.2	5.1
	Normal	$5\sqrt{x}$	−.0010	.0154	.0148	9.4	4.5

one-sided test of concavity (i.e., reject when there is a local region of convexity). Although the actual test of Dümbgen and Spokoiny (2001) is a multiscale test (using multiple bandwidths simultaneously), in Table 3 we consider a version of the DS test that uses only a single bandwidth. The purpose is to compare the underlying filtering method of Dümbgen and Spokoiny (2001) with the spline method of Diack (2001) and the simplex statistic. Abrevaya and Jiang (2003) reported results from simulations using multiscale versions of the DS test and the simplex test.

Both the DS and D methods require estimation of the variance of ϵ. For the D test, the B-spline method of Diack (2001) was followed explicitly. The estimated error variance used was $\hat{\sigma}^2 = n^{-1}\sum_i \hat{\epsilon}_i^2$, where $\{\epsilon_1, \dots, \epsilon_n\}$ are the estimated residuals from the spline estimation. The method outlined in section 4.1 of Dümbgen and Spokoiny (2001) was used to implement the DS test. The estimated error variance used was $\hat{\sigma}^2 = (6(n-2))^{-1}\sum_{i=2}^{n-2}(2\tilde{y}_i - \tilde{y}_{i-1} - \tilde{y}_{i+1})^2$, where $\tilde{y}$ denotes the residual from the linear projection of y on x.

For the simplex test, the number of bins chosen was between $G = 4$ and $G = 6$ for the sample sizes considered ($n = 100, 200, 400, 1{,}000$). The choices in the table correspond to between roughly 25 observations per bin (for $n = 100$ and $G = 4$) and 167 observations per bin (for $n = 1{,}000$ and $G = 6$). The results of Table 2 suggest that such sample sizes are sufficient for the application of the local simplex statistics. Across sample sizes, the DS test was evaluated using the same bin choices as the simplex statistic. The simulations indicated that the size of both the simplex test and the DS test were very insensitive to the choice of bandwidth. For the simulations reported, the size is very accurate for both the simplex and DS test. For the Diack (2001) test, the size was more sensitive to bandwidth choice. The number of interior spline knots (e.g., equal to 8 for $n = 100$) was chosen such that the size of the D test was most

Table 3. Size and Power of Local Tests for Curvature

n	Test	No. of bins or knots	(D1), linear		(D1), $\alpha = 1.2$		(D1), $\alpha = 1.4$		(D2) cubic		(D3) convex dip	
			10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
100	S	4	9.8	4.7	10.3	5.4	12.1	7.7	13.1	6.6	79.6	69.2
	DS	4	10.4	5.6	10.2	5.2	13.1	6.6	32.2	20.8	82.0	73.3
	D	8	9.3	4.8	9.6	5.2	9.1	4.4	9.0	4.8	78.4	69.1
200	S	4	10.2	5.1	12.4	7.3	13.9	8.2	28.5	18.5	98.7	97.2
	S	5	9.9	5.1	11.2	6.0	12.0	7.1	15.3	8.0	71.0	59.8
	DS	4	10.5	5.1	11.1	5.2	13.4	7.3	40.7	29.0	97.1	94.8
	DS	5	10.1	5.2	9.8	4.9	12.7	7.1	34.1	23.8	75.4	63.8
	D	8	9.4	4.6	9.6	4.8	9.6	5.0	9.4	4.8	98.1	96.8
400	S	4	10.7	5.3	13.9	8.0	16.7	9.4	53.0	42.4	100	100
	S	5	9.9	4.7	12.1	7.3	14.2	8.7	25.4	16.5	100	100
	DS	4	10.1	5.3	12.3	6.8	14.5	7.8	57.0	45.2	100	100
	DS	5	10.2	10.2	11.6	5.6	13.0	7.3	41.3	29.0	100	100
	D	9	9.7	5.2	9.3	4.9	10.0	5.2	8.8	4.4	100	100
1,000	S	5	9.7	4.8	13.8	8.1	16.6	10.2	53.6	38.8	100	100
	S	6	10.5	5.4	12.3	6.8	16.3	10.5	28.8	22.1	100	100
	DS	5	9.8	4.7	10.5	5.8	17.5	9.6	60.0	47.3	100	100
	DS	6	10.2	5.0	11.7	5.7	13.9	7.0	45.6	34.0	100	100
	D	12	9.3	4.9	8.8	4.3	9.4	4.4	9.5	4.4	100	100

accurate. In cases where a range of choices gave accurate size, a smaller number of knots was chosen, because this choice gave greater power for the nonlinear designs.

The results in Table 3 indicate that the simplex statistic and the underlying filtering method of the DS test have comparable size and power characteristics. Across the designs, the rejection rates of the S and DS tests are similar, with the DS test showing greater power in the cubic design and the S test showing slightly greater power in the nonlinear (D1) designs. The power of the D test is comparable to that of the other two tests for design (D3) but is much lower than that of the other designs. The problem is that the number of knots is relatively high and the nonlinear functions in (D1) and (D2) are relatively smooth, making it difficult for the D test to discern nonlinearity. The power of the D test for these designs increases if the number of knots is reduced, but there is significant overrejection in the linear case with the reduced number of knots.

4.2 Two-Covariate Case

In this section we report results from simulations using U_n' with bivariate **x**. We use the following Monte Carlo design:

$$y = 50(x_1/100)^{\alpha} + 50(x_2/100)^{\alpha} + \epsilon, \qquad (37)$$

where $\alpha = 1$ yields linearity, $\alpha < 1$ yields concavity, and $\alpha > 1$ yields convexity. Here x_1 and x_2 were chosen to be independently and uniformly distributed on the interval [0, 100]. Table 4 examines the size of the linearity test (i.e., testing $H_0: \theta = 0$ when $\alpha = 1$) for $n = 50, 100, 200, 400$. Designs with uniform and normal error distributions are considered, with the error standard deviation set to 10, 20, randomly heteroscedastic ($20u$, where u is a standard uniform [0, 1] random variable), or proportional to $\sqrt{x_1 + x_2}$.

For each of the 24 designs in Table 4, we performed 1,000 simulations. For the standard errors of the simplex statistic, we used both the asymptotic method and a "quadruples" bootstrap method. For each simulation, we constructed 400 bootstrap U_n' statistics in the following way: (a) draw one observation (y_i, x_{1i}, x_{2i}) and then find all possible quadruples that include this observation; (b) repeat (a) from the original sample n times; (c) calculate the U_n' statistic from these $\binom{n}{4}$ quadruples; and (d) rescale the bootstrap statistic by $\binom{4}{1} = 4$ (because in each quadruple we fixed one observation). Rejections were based on the standard deviation of statistics over the bootstrap samples. Arcones and Giné (1992; thm. 2.4) proved that the bootstrap statistic constructed in this way has the same limit distribution as the original U_n' under weak moment conditions. Note that this bootstrap sampling is valid under iid errors as well as conditionally heteroscedastic errors. The quadruple bootstrap is much less costly in computation compared with data resampling or error replacement, in addition to being robust to error distributions. We also note that this type of bootstrap could be used for the univariate case, but our experience suggests that the asymptotic approximation is quite accurate even for very small sample sizes in the univariate case.

Table 4 reports the mean bias (relative to 0) of U_n over the 1,000 simulations, the empirical standard deviation of the U_n' values, the mean of the estimated asymptotic standard errors, and the mean of bootstrap standard errors. Rejection rates at the 10% and 5% levels are reported for both methods of standard error estimation. The rejection rates are very encouraging, even with these relatively small samples. There is slight overrejection at $n = 50$ using the asymptotic critical values, but the bootstrap method rejection rates are very close to 5% and 10%.

Next, we compared the power of the simplex statistic test with a parametric test and another nonparametric test, specifically the test of Ellison and Ellison (2000). The Ellison–Ellison

Table 4. Size of the Multivariate Simplex Test Under the Null of Linearity

n	*Error distribution*	*SD*	*Mean bias*	*Empirical SD*	*Average asymptotic standard error*	*10% rejection*	*5% rejection*	*Average bootstrap standard error*	*10% rejection*	*5% rejection*
50	Normal	10	.0005	.1259	.1194	12.5	7.0	.1297	9.8	5.6
	Normal	20	.0052	.1212	.1197	11.6	6.4	.1288	9.8	5.3
	Uniform	10	−.0031	.1306	.1110	11.1	6.8	.1337	9.3	5.3
	Uniform	20	.0022	.1333	.1237	13.9	8.0	.1339	11.0	6.1
	Normal	$20u$	.0032	.1248	.1181	12.2	6.8	.1256	10.2	5.3
	Normal	$\sqrt{x_1+x_2}$	−.0025	.1289	.1187	13.7	7.4	.1289	10.2	5.7
100	Normal	10	−.0012	.0830	.0823	10.4	5.8	.0859	9.2	4.6
	Normal	20	−.0007	.0841	.0829	11.4	6.0	.0852	10.2	5.4
	Uniform	10	−.0032	.0868	.0862	10.1	6.0	.0888	10.0	4.8
	Uniform	20	.0006	.0860	.0860	11.5	6.2	.0887	10.7	5.4
	Normal	$20u$	−.0009	.0819	.0810	10.9	6.5	.0836	10.2	5.4
	Normal	$\sqrt{x_1+x_2}$	.0019	.0856	.0828	10.6	6.5	.0853	10.1	5.8
200	Normal	10	.0015	.0603	.0579	11.5	6.2	.0598	10.4	5.7
	Normal	20	−.0001	.0598	.0578	10.8	5.7	.0587	10.3	5.4
	Uniform	10	−.0001	.0605	.0600	10.1	5.2	.0609	10.0	4.9
	Uniform	20	−.0019	.0610	.0602	12.0	5.5	.0608	10.6	5.2
	Normal	$20u$	−.0002	.0574	.0560	10.4	5.5	.0569	10.2	5.2
	Normal	$\sqrt{x_1+x_2}$	.0017	.0575	.0574	10.3	4.9	.0583	9.6	4.7
400	Normal	10	.0031	.0410	.0404	11.0	4.5	.0407	11.1	4.9
	Normal	20	.0011	.0421	.0405	11.3	5.8	.0408	10.7	5.7
	Uniform	10	.0021	.0413	.0422	9.1	4.1	.0425	8.7	4.1
	Uniform	20	−.0009	.0418	.0422	10.1	4.6	.0425	10.2	4.6
	Normal	$20u$	−.0009	.0391	.0392	10.2	4.7	.0395	10.4	4.6
	Normal	$\sqrt{x_1+x_2}$	.0011	.0415	.0402	10.5	5.7	.0406	10.4	5.5

test is a general nonparametric specification test, which we chose both for its simplicity and its performance in relation to other nonparametric specification tests. When applied to our design, the Ellison–Ellison test tries to discern linearity versus nonlinearity. The Ellison–Ellison test statistics are based on quadratic forms in the null (linear) model's residuals. Suppose that the null model is $\mathbf{y} = f(\mathbf{x}; \alpha) + \mathbf{u}$. The idea is that the quadratic form $\tilde{\mathbf{u}}\mathbf{W}\tilde{\mathbf{u}} = \sum w_{ij}\tilde{u}_i\tilde{u}_j$, for some kernel-weighting matrix $\mathbf{W}$, can detect a spatial correlation in the residuals that could result from a functional form misspecification. The test statistics are asymptotically normally distributed.

A natural parametric test of linearity would be through the following regression:

$$y_i = b_0 + b_1 x_{1,i} + b_2 x_{2,i} + b_3 x_{1,i}^2 + b_4 x_{2,i}^2 + \epsilon_i. \tag{38}$$

Linearity implies that

$$b_3 = b_4 = 0, \tag{39}$$

and convexity amounts to

$$b_3 > 0, \qquad b_4 > 0. \tag{40}$$

When one does not have prior knowledge that the marginal effect of one covariate is independent of the other, the regression should include an interaction term as well (see Abrevaya and Jiang 2003).

Because the Ellison–Ellison test discerns nonlinearity but does not point out the direction of curvature (convexity or concavity), we compared power based on two-sided tests of linearity versus nonlinearity. That is, we applied a two-sided confidence interval on the simplex statistics and tested the joint restriction of (39) in the parametric test. In the Ellison–Ellison test, one needs to choose the window width to form the proper weighting matrix $\mathbf{W}$. In our simulations, we chose the window widths to be $n^{-1/5}$ times the standard deviation of the respective covariate components. Other choices in this range gave similar results. (In other simulations, we tried setting the smoothing parameters for the Ellison–Ellison test at the values that gave the most accurate size; the results were also very similar.) For the simplex and Ellison–Ellison tests, we based rejection rates on bootstrap standard errors, because these give more accurate size in finite samples compared with the asymptotic standard errors. The results are gathered in Table 5. The power of the simplex statistic dominates the other nonparametric test in all specifications and sample sizes. The power of the simplex statistic test increases with sample size at a faster rate than that of the Ellison–Ellison test. On the other hand, the power of the nonparametric test is almost indistinguishable from that of the parametric test for most cases and is superior to the latter when heteroscedasticity is present and heteroscedasticity-consistent standard errors are used.

5. PARTIALLY LINEAR REGRESSION MODEL

Because testing the joint curvature of many variables often can be practically infeasible, an appealing modeling strategy is to specify a regression model in which some variables are assumed to enter linearly into the model and others are left flexibly specified. With this type of model, known as a *partially linear regression* model (or a *semilinear regression* model), the test of curvature can be focused on the latter group of variables for which the model is left flexibly specified. In particular, consider the partially linear regression model

$$y_i = \mathbf{z}_i'\boldsymbol{\beta} + f(\mathbf{x}_i) + \epsilon_i, \qquad E[\epsilon_i|\mathbf{x}_i, \mathbf{z}_i] = 0, \qquad i = 1, \ldots, n, \tag{41}$$

where $\mathbf{z}_i$ is a $1 \times k$ covariate vector, $\boldsymbol{\beta}$ is a $k \times 1$ parameter vector, and $\mathbf{x}_i$ is an $1 \times m$ covariate vector.

This model has received a great deal of attention in the econometrics and statistics literature, because it allows a researcher to focus on a finite-dimensional parameter of interest ($\boldsymbol{\beta}$) while allowing an infinite-dimensional nuisance parameter (the function f). Powell (1987), Robinson (1988), and Yatchew (1997) each developed $\sqrt{n}$-consistent estimators of $\boldsymbol{\beta}$ without parametric restrictions on the distribution of ϵ or the functional form of f. The function f can be estimated in a second stage via nonparametric regression of the residuals $y_i - \mathbf{z}_i'\hat{\boldsymbol{\beta}}$ on $\mathbf{x}_i$ (where $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$-consistent first-stage estimate). Due to the different rates of convergence for estimation

Table 5. Comparison of Power for Multivariate Linearity Tests

n	Error SD	Simplex 10% rejection	Simplex 5% rejection	Ellison–Ellison 10% rejection	Ellison–Ellison 5% rejection	Parametric 10% rejection	Parametric 5% rejection
50	10	18.1	12.3	10.2	5.4	19.5	11.8
	20	12.8	7.8	9.9	5.3	12.7	7.0
	20u	22.5	13.1	10.7	5.5	17.2	10.5
	$\sqrt{x_1 + x_2}$	19.8	13.4	9.9	5.1	18.8	11.8
100	10	28.9	19.3	10.8	5.5	28.6	18.8
	20	15.2	9.1	10.3	5.2	15.4	8.9
	20u	33.3	21.9	10.2	5.3	24.7	16.2
	$\sqrt{x_1 + x_2}$	30.9	22.4	12.4	7.0	28.9	18.7
200	10	46.6	34.9	12.4	6.8	47.5	34.7
	20	20.3	12.2	10.6	6.0	20.8	12.0
	20u	54.3	40.7	10.5	5.8	37.7	26.6
	$\sqrt{x_1 + x_2}$	50.0	36.5	13.8	8.4	47.1	34.8
400	10	71.2	59.2	16.6	10.4	72.7	61.6
	20	29.8	18.5	10.7	5.6	29.9	19.9
	20u	79.1	69.0	13.6	7.8	62.8	50.2
	$\sqrt{x_1 + x_2}$	72.5	62.3	14.7	8.6	72.5	62.0

of $\boldsymbol{\beta}$ and the function f, the asymptotic properties of the second-stage estimates are unaffected by the first-stage estimation of $\boldsymbol{\beta}$ (see, e.g., Yatchew 1997 for a brief discussion).

Note that $y_i - \mathbf{z}_i'\boldsymbol{\beta} = f(\mathbf{x}_i) + \epsilon_i$, which means that the test statistics developed for the model in (1) could be applied to the partially linear model if $\boldsymbol{\beta}$ were known. Because $\boldsymbol{\beta}$ is unknown, the natural approach is to replace it with an estimate. In particular, the simplex statistic is based on $\{(y_i(\hat{\boldsymbol{\beta}}), \mathbf{x}_i) : i = 1, \dots, n\}$, where $y_i(\mathbf{b}) \equiv y_i - \mathbf{z}_i'\mathbf{b}$ and $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$-consistent estimate of $\boldsymbol{\beta}$. The simplex statistic, then, depends on the value of the coefficient vector for $\mathbf{z}$,

$$U_n(\mathbf{b}) \equiv \binom{n}{m+2}^{-1} \sum_{t_1 < \cdots < t_{m+2}} \mathbb{1}(\{\mathbf{x}_1, \dots, \mathbf{x}_{m+2}\} \in \mathcal{S}) \times \operatorname{sign}\big(a_1 y_{[t_1]}(\mathbf{b}) + \cdots + a_{m+1} y_{[t_{m+1}]}(\mathbf{b}) - y_{[t_{m+2}]}(\mathbf{b})\big). \tag{42}$$

The notation in (42) is the same as that in Section 2, with the exception that $y_i(\mathbf{b})$ values replace y_i values. Because the simplex statistic is now a function of the coefficient vector, additional notation is needed for the its limiting value,

$$\mu(\mathbf{b}) \equiv \operatorname*{plim}_{n\to\infty} U_n(\mathbf{b}). \tag{43}$$

Then $\mu(\boldsymbol{\beta})$ corresponds to θ from Section 2, with the results therein applying to $\mu(\boldsymbol{\beta})$.

Letting $\mathbf{v}_i \equiv (y_i, \mathbf{z}_i, \mathbf{x}_i)$ and $\mathbf{v}_i(\mathbf{b}) \equiv (y_i(\mathbf{b}), \mathbf{x}_i)$, the analogous definitions of the functions h and $\tilde{h}$ [see (11) and (12) in Sec. 2] are

$$h(\mathbf{v}_1, \dots, \mathbf{v}_{m+2}; \mathbf{b}) \equiv \mathbb{1}(\{\mathbf{x}_1, \dots, \mathbf{x}_{m+2}\} \in \mathcal{S}) \times \operatorname{sign}\big(w(\mathbf{v}_1(\mathbf{b}), \dots, \mathbf{v}_{m+2}(\mathbf{b}))\big) \tag{44}$$

and

$$\tilde{h}(\mathbf{v}; \mathbf{b}) \equiv E\big(h(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m+2}; \mathbf{b}) \,|\, \mathbf{v}\big). \tag{45}$$

The following assumptions are made.

Assumption 5. An iid sample $\{(\mathbf{x}_i, \mathbf{z}_i, \epsilon_i)\}_{i=1}^n$ is drawn from the random variables $\{\mathbf{x}, \mathbf{z}, \epsilon\}$, with $\epsilon|\mathbf{x}, \mathbf{z}$ symmetrically distributed about 0, and the support of $\mathbf{x}$ and $\mathbf{z}$ is bounded. The observed sample is $\{(y_i, \mathbf{x}_i, \mathbf{z}_i)\}_{i=1}^n$, where y_i is generated according to the model in (41).

Assumption 6. The density function of $\epsilon|\mathbf{x}, \mathbf{z}$ is bounded by some constant $M > 0$ for all $(\mathbf{x}, \mathbf{z})$.

The boundedness assumptions simplify the extension of the asymptotic theory for the simplex statistics, and it may be possible to relax these. Because the estimators of $\boldsymbol{\beta}$ proposed in the literature require somewhat different assumptions for $\sqrt{n}$-consistency, we merely assume that a $\sqrt{n}$-consistent estimate has been obtained (with the understanding that the technical assumptions related to the chosen estimator must also be satisfied). The first-stage estimator of $\boldsymbol{\beta}$ and the simplex statistic converge at the same $\sqrt{n}$ rate, meaning that the asymptotic distribution of the first-stage estimator will potentially affect the asymptotic distribution of the (second-stage) simplex statistic $U_n(\hat{\boldsymbol{\beta}})$. Let $\nabla_{\mathbf{b}}$ denote the gradient with respect to $\mathbf{b}$. Then the asymptotic distribution is given by the following theorem.

Theorem 8. Suppose that $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$ satisfying

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = n^{-1/2} \sum_{i=1}^{n} \boldsymbol{\psi}(\mathbf{v}_i) + o_p(1) \quad \text{s.t.} \quad E[\boldsymbol{\psi}(\mathbf{v})] = \mathbf{0} \quad \text{and} \quad E[\boldsymbol{\psi}(\mathbf{v})\boldsymbol{\psi}(\mathbf{v})'] \text{ exists.} \tag{46}$$

If Assumptions 2, 5, and 6 hold, then

$$\sqrt{n}\big(U_n(\hat{\boldsymbol{\beta}}) - \mu(\boldsymbol{\beta})\big) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\delta}'\mathbf{V}\boldsymbol{\delta}), \tag{47}$$

where

$$\boldsymbol{\delta} \equiv \begin{pmatrix} 1 \\ \nabla_{\mathbf{b}}\mu(\boldsymbol{\beta}) \end{pmatrix},$$

$$\boldsymbol{\phi}(\mathbf{v}) \equiv \begin{pmatrix} (m+2)(\tilde{h}(\mathbf{v}; \boldsymbol{\beta}) - \mu(\boldsymbol{\beta})) \\ \boldsymbol{\psi}(\mathbf{v}) \end{pmatrix},$$

and

$$\mathbf{V} \equiv E[\boldsymbol{\phi}(\mathbf{v})\boldsymbol{\phi}(\mathbf{v})'].$$

The main results used to prove Theorem 8 are from Randles (1982), who considered the effect of estimated parameters on the asymptotic distributions of U-statistics. The gradient in the asymptotic distribution for the simplex statistic can be consistently estimated by numerical differentiation of $U_n(\hat{\boldsymbol{\beta}})$ with respect to $\mathbf{b}$. The other quantities can be consistently estimated by their sample analogs.

With Theorem 8, hypothesis tests involving the curvature of f can be conducted just as in Section 2. If one is interested in testing the null hypothesis of linearity, $H_0 : f(\mathbf{x}) = \alpha + \mathbf{x}'\boldsymbol{\gamma}$, then it is possible to avoid estimating the partially linear regression model with the techniques referenced earlier; instead, a least squares regression of $\mathbf{y}$ on $\mathbf{z}$ and $\mathbf{x}$ can be used, because the least squares estimator of $\boldsymbol{\beta}$ is consistent under the null hypothesis. Then the asymptotic distributions of the test statistics are given by Theorem 8, with $\mu(\boldsymbol{\beta}) = 0$ under H_0. To avoid numerical differentiation, the bootstrap also could be used to carry out the hypothesis tests. Finally, the localized tests from Section 3 could be adapted to the partially linear regression context. As for the global simplex statistic, testing is done using $\{y_i(\mathbf{b}), \mathbf{x}_i\}_{i=1}^n$. The asymptotic variances of the localized simplex statistics (see Thm. 4) must take into account the estimation of $\boldsymbol{\beta}$ as in Theorem 8.

6. EMPIRICAL APPLICATIONS

6.1 The Experience-Earnings Profile

In this section we apply the univariate simplex statistic to wage data taken from the Current Population Survey (CPS) to examine the curvature of the empirical age-earnings profile, or the "human capital earnings function" as it has been called since Mincer (1974) published the term. Expressing the logarithm of earnings as a quadratic function in experience is one of the most widely accepted empirical specifications in economics. Aggregate data used in this line of work tend to exhibit heteroscedasticity because of differences in the number of observations from cohort to cohort. Further, the quadratic specification has been questioned. For example, Murphy and Welch (1990), based on a dataset running from 1963 to 1986, pointed out that quadratic fitting would understate early career earnings growth by about 30–50% and overstate mid-career growth by 20–50%; therefore, they suggested cubic and quartic specifications as better alternatives. Due to the inherently discrete nature of the covariate data, we are able to consider only the global version of the test statistic from Section 2.

The data come from the CPS Annual Demographic Survey for the 3 years 1997–1999. We focus on the subsample of white males who were full-time employees during the year of survey and then classify them by age and education. We follow Murphy and Welch (1990) to trace out the wage-experience and wage-growth-experience profiles. The education groups include categories of 7–11 years ("some high school"), 12 years ("high school"), 13–15 years ("some college"), and 16 or more years of school completed ("college or more"). Within each educational level, workers are divided into 41 "pseudoexperience" groups based on age as in Murphy and Welch (1990): for those with 8–11 years of schooling, experience is age minus 18 years; for high school graduates, experience is age minus 19 years; for men with 13–15 years of schooling, experience is age minus 20 years; and, for college graduates, experience is age minus 22 years. An average log weekly wage is computed in each of the year-experience-schooling cells by taking the average of [log(annual earnings) − log(weeks worked)]. Figures 4 and 5 show the wage-experience and wage-growth-experience profiles of all four education groups averaged over the 3 years (weighted by sample size in each year). These figures indicate a

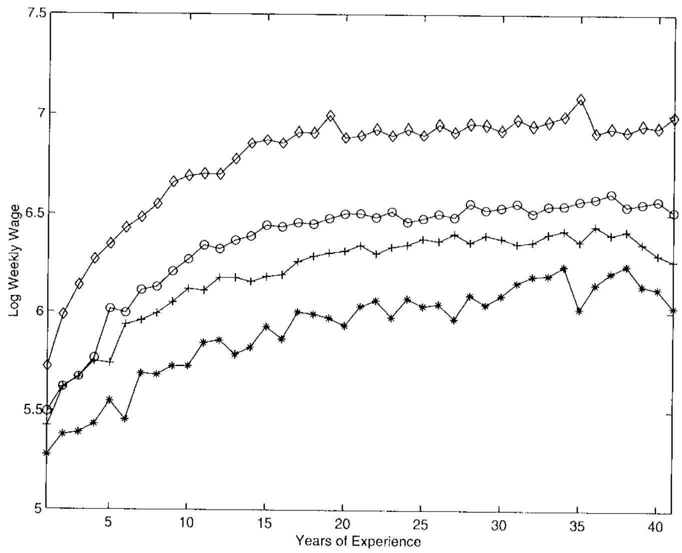

Figure 4. The Average Earnings-Experience Profiles, 1997–1999 (-- some high school; -+- high school; -o- some college; -◇- college or more).*

higher-education premium but proportionally more in the form of big early-career advances for college-educated people; once into the mid-career years, the earnings profiles of all education groups are basically parallel to each other.

The results from testing the curvature of the wage-experience and wage-growth-experience profiles are reported in Table 6. First, we test whether earning power (proxied by the average

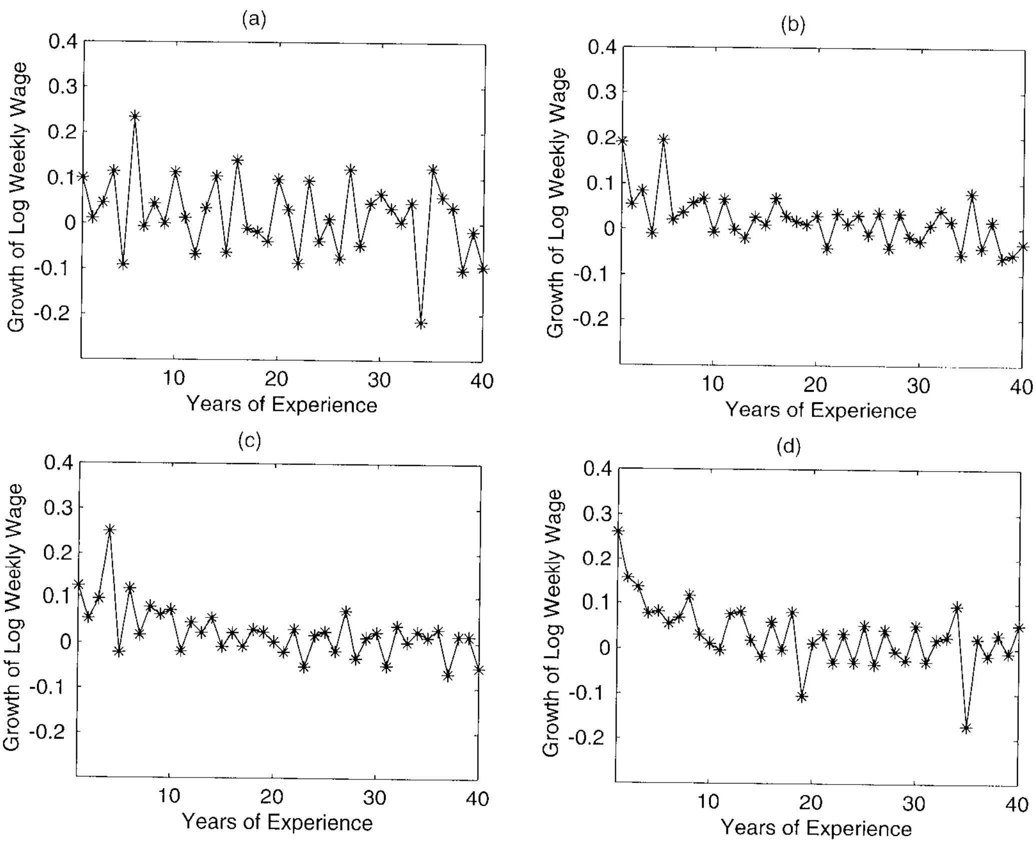

Figure 5. The Average Earnings-Growth-Experience Profiles, 1997–1999. (a) Some high school; (b) high school graduates; (c) some college; (d) college or more.

Table 6. Testing the Empirical Age-Earnings Profiles (1997–1999)

	Some high school	High school	Some college	College or more
1999				
No. of observations	2,523	8,424	6,696	7,404
Level	−.4280 (.0678)*	−.6344 (.0734)*	−.5162 (.0952)*	−.5442 (.0896)*
Growth	.0494 (.0822)	.0922 (.1066)	.0920 (.1056)	.1704 (.0880)***
1998				
No. of observations	2,605	8,402	6,639	7,222
Level	−.3440 (.0786)*	−.6082 (.0522)*	−.5188 (.0900)*	−.5348 (.0958)*
Growth	−.0098 (.0666)	.0568 (.0926)	.1262 (.0994)	.1552 (.0854)***
1997				
No. of observations	2,716	8,436	6,612	7,115
Level	−.3300 (.0784)*	−.6424 (.0576)*	−.5554 (.0842)*	−.5394 (.0718)*
Growth	−.0658 (.0762)	.0444 (.0948)	.0586 (.0776)	.0878 (.0870)
1997–1999				
No. of observations	7,884	25,262	19,947	21,741
Level	−.4770 (.0732)*	−.7280 (.0536)*	−.6610 (.0780)*	−.6296 (.0828)*
Growth	.0042 (.0798)	.0556 (.0840)	.1586 (.0752)**	.2572 (.0936)*

* Different from 0 (two-tail tests) at 1% significance.
** Different from 0 (two-tail tests) at 5% significance.
*** Different from 0 (two-tail tests) at 10% significance.

log weekly wage) is a concave function of years of experience. Workers with the same experience level were aggregated into one observation, so that there is a single data point for each value of experience. Note that the simplex statistic remains valid on aggregated data when the conditional distribution of individual disturbances is symmetric, because the average of symmetric error disturbances is still a symmetric error disturbance. The results (labeled "Level" in Table 6) indicate that linearity can be strongly rejected (at the 1% level in all cases). The values for the simplex statistics are all significantly negative, which is consistent with concavity of the wage-experience profile.

Next, we tested whether a quadratic model is appropriate for the empirical wage-experience profiles. In this setting, a quadratic specification corresponds to a linearly decreasing growth rate of wages as experience increases. Thus we conducted the linearity test using the simplex statistic on the average log growth rate of weekly wage (vs. experience again). The results (labeled "Growth" in Table 6) show that linearity of the log wage growth rate for the lowest education group (some high school) cannot be rejected. For the other three education groups, the results favor convexity, but only the statistics for the highest education group (college or more) in 1998 and 1999 are significantly different from 0 at the 10% level; a one-sided test of negative θ would be rejected at the 5% level. For the full 3-year sample, however, the statistics become significantly different from 0 at the 1% level for the college-or-more group and at the 5% level for the some-college group. Overall, based on our results, there is no compelling evidence to reject the quadratic wage-earnings specification for men without college education; for better-educated workers, however, there is evidence that earnings growth decreases at a decreasing rate.

6.2 Mutual Fund Style Timing

The research work on measuring professional money managers' performance has focused on two basic abilities: selectivity and timing. The former tests whether a fund manager is able to pick securities that outperform the market in risk-adjusted terms, and the latter tests whether a fund manager can out-guess the market by moving in and out of "timing portfolios," or portfolio proxies for risk factors. By testing the curvature of mutual fund returns vis-à-vis factor returns, we want to infer the "style timing" performance of mutual funds.

Among timing tests, most research has focused solely on the market risk factor, or "market timing" in its narrow sense. Since the work by Treynor and Mazuy (1966) and Henrikkson and Merton (1981), researchers have developed different ways to test whether fund managers successfully capitalize on market ascendancy and/or avoid downturns by changing the market exposure of their portfolios (Becker, Ferson, Myers, and Schill 1999; Goetzmann, Ingersoll, and Ivkovish 2000; Bollen and Busse 2001; Jiang 2003). Few studies, however, have examined money managers' timing on different styles of assets, even though style investing has been identified by researchers as an important strategy that can shed light on a number of style-related empirical anomalies (Barberis and Shleifer 2003). When investors entrust their money with professional managers, they hope that the managers are able to pick individual securities as well as identify categories of assets that will give high returns. For example, during the second half of the 1990s, large-cap stocks outperformed small-cap stocks, and growth stocks outperformed value stocks. Since 2000, however, these trends have been reversed. Both periods gave style timers great chances to beat the market average. Therefore, it would be interesting to see whether funds claiming to be "asset allocators" successfully allocate monies among different classes of assets to capture the ascendancy of certain styles and avoid downturns in others.

In the single-factor market timing test, the portfolio returns of a successful market timer should, on average, display convexity against the market return. That is, a successful market timer should have the fund rise significantly when the market rises and fall slightly when the market drops. The same idea can be generalized and extended to multifactor models; the returns of a successful style timer should display joint convexity against the factor returns. In this article we consider mutual funds' timing of the two most discussed risk factors (other than the market risk factor): the size and the book-to-market factors. These two factors represent the popular notions of style timing: small versus large and growth versus value. Although several market timing tests exist in the literature for the univariate case, we are not aware of any such alternatives in the multivariate case. The approach taken here is essentially a multivariate extension of the methodology of Jiang (2003).

We assume that the returns of a mutual fund follow the Fama and French (1993) three-factor model,

$$r_{t+1} = \alpha + \beta_1 r_{m,t+1} + \beta_{2,t} SMB_{t+1} + \beta_{3,t} HML_{t+1} + \varepsilon_{t+1}, \tag{48}$$

where both r_{t+1} and $r_{m,t+1}$ are excess returns over the risk-free rate. Because we are testing whether fund managers are able to time the size (SMB) and book-to-market (HML) factors, the two coefficients $\beta_{2,t}$ and $\beta_{3,t}$ are random variables adapted to the information available to the manager at time t. Let $\widetilde{r}_{t+1} \equiv r_{t+1} - \beta_1 r_{m,t+1}$ be the residual return of the fund after the market factor is filtered out. Then

$$\widetilde{r}_{t+1} = \alpha + f(SMB_{t+1}, HML_{t+1}) + \varepsilon_{t+1}. \tag{49}$$

Let $\mathbf{S}_i = \binom{SMB_i}{HML_i}$ represent state i. If the fund manager does not time the factors at all, then the $\beta_{2,t}$ and $\beta_{3,t}$ remain constant, so that $E[f(\mathbf{S})] = f[E(\mathbf{S})]$; that is, the average return is the same as the return in an average state, because the relationship is linear. A successful timer, on the other hand, would take advantage of the variations in states and achieve a higher expected return, that is,

$$E[f(\mathbf{S})] > f[E(\mathbf{S})]. \tag{50}$$

Note that (50) amounts to saying that f is jointly convex in SMB and HML. Therefore, the hypothesis of style timing becomes

$$H_0: f \text{ is linear in } SMB \text{ and } HML \text{ (no timing)}$$

and

$$H_a: f \text{ is jointly convex in } SMB \text{ and } HML \text{ (good timing)}.$$

We test the hypothesis using mutual fund data from Morningstar Principia Pro Plus for Mutual Funds (1972–2001) and Fama–French factor data from French's website data library. A mutual fund is included in our sample if it invests primarily in domestic stocks and satisfies one of the following criteria: (a) it falls into the categories of "asset allocation" by its prospectus objective or (b) it falls into the category of "domestic hybrid" by Morningstar's classification. (Most funds marketing themselves as asset allocation funds are classified by Morningstar as domestic hybrid funds; therefore, we include all funds that satisfy one of the selection criteria.) Among all mutual funds, these funds are the most likely market timers. Asset allocation and hybrid funds usually do not stick to a fixed style; managers of these funds shift assets frequently based on their view about the relative returns of different classes of assets. All together, there were 222 such funds in business at the end of 2001. After 1993, Morningstar listed the name of funds that were removed from the database because they were liquidated or acquired by other funds. We were able to back out information of those funds that perished after 1993 using the same criteria from Morningstar's earlier publications of Principia Pro for Mutual Funds. We have information on 125 perished funds.

We construct the variable $\widetilde{r}$ in (49) by estimating $\widehat{\beta}_1$ from a regression of fund excess return on CRSP value-weighted market excess return and the two Fama–French factors, SMB and HML. (Under the null hypothesis that r is linear in all three factors, $\widehat{\beta}_1$ from such an estimation is consistent. To have an estimate that is consistent under both the null and the alternative, one can use one of the estimators mentioned in Sec. 5. In this application, the Robinson 1988 estimator yielded extremely similar results and are not reported.) Then we use the simplex statistic to test whether $\widetilde{r}$ is jointly convex in SMB and HML for each fund. Because $\widetilde{r}$ is formed using an estimated $\widehat{\beta}_1$, we account for the impact of estimation error in $\widehat{\beta}_1$ on the standard error of the (second-stage) simplex statistic (by applying the results of Sec. 5).

Table 7 summarizes the results for style-timing performance of all funds, live funds, and dead funds for the period 1972–2001 and for three 10-year subperiods. A fund was included in estimation only if it had at least 2 years' of monthly return data within the time window of consideration. To interpret the performance measure in probability terms, we report the scaled version of U_n' as defined in (8) instead of U_n in (7). We also report the statistics in percentage point terms, so that they are interpretable as the empirical probability that a fund moves into classes of the assets (i.e., large cap vs. small cap, and growth vs. value) at the right time (i.e., when the subsequent returns of those asset classes are higher) in excess of the

Table 7. Simplex Statistics for Measuring Style Timing

	No of funds	Quantiles 10%	25%	50%	75%	90%	% with p value < .05 Convexity test	Concavity test
All funds								
1972–2001	315	−8.97	−5.48	−.50	4.70	11.17	2.2%	5.7%
1972–1981	25	−13.69	−10.03	−5.75	3.66	9.48	20.0%	0%
1982–1991	85	−8.07	−1.50	2.48	7.51	12.31	1.2%	4.7%
1992–2001	315	−9.73	−6.13	−1.15	4.60	11.24	2.9%	5.4%
Live funds								
1972–2001	205	−8.63	−4.73	1.17	6.96	12.70	2.2%	8.3%
1972–1981	19	−14.02	−10.73	−6.51	−2.55	6.17	26.3%	0%
1982–1991	65	−7.34	−1.97	2.10	7.74	13.63	0%	6.2%
1992–2001	205	−9.77	−5.65	.55	7.14	13.08	3.9%	7.8%
Dead funds								
1972–2001	110	−9.70	−6.59	−3.31	1.52	4.70	1.0%	1.0%
1992–2001	110	−9.70	−6.90	−3.93	.85	4.70	1.0%	1.0%

probability that they do so at the wrong time. Table 7 reports various quantiles (10%, 25%, 50%, 75%, 90%) of the simplex statistics within each category. For instance, the median of the statistics for all the funds in the 1972–2001 period is −.50%.

Overall, there does not seem to be much evidence of style timing among the funds in our sample. The last two columns of Table 7 report the percentage of simplex statistics for which one-sided tests (of convexity and concavity) have a p value $< .05$. The largest rejection rate for the test of concavity is for the live funds during the 1972–2001 period, but only 8.3% of associated simplex tests have a p value $< .05$.

7. CONCLUSION

We briefly discuss some possible areas for future research. Modified versions of the simplex statistics might yield more powerful tests. First, there is nothing special about the sign(·) function. Any odd function satisfying weak regularity assumptions (i.e., existence of first and second moments of the associated kernel function) will yield a test statistic with the properties from Section 2. Second, different weights of $(m+2)$-tuples could be used in the test statistics. As presented, the simplex statistic implicitly weights each $(m+2)$-tuple equally. Different weighting would likely improve the power of tests based on the statistics. For instance, in the univariate case, slopes based on small differences in x values tend to be "noisier," suggesting that observation-triples involving such slopes should be downweighted.

The issue of bandwidth choice for the consistent tests of Section 3 certainly warrants further investigation. Although Theorem 7 provides a bandwidth rate for consistency, there are no theoretical results in the literature providing guidance as to which bandwidth choices might provide more powerful tests either in the Ghosal et al. (2000) setting or the one considered in this article.

Finally, there may be other settings in which the simplex idea may prove useful. For instance, in the realm of numerical optimization, global extrema are ensured if an objective function is globally concave or convex. Perhaps simplices could be chosen appropriately to check for the required curvature.

ACKNOWLEDGMENTS

This article combines two previously separate articles, "Pairwise-Slopes Statistics for Testing Curvature" and "A Nonparametric Approach to Measuring and Testing Joint Curvature." Comments from Xiaohong Chen, Jerry Hausman, James Heckman, Robert McCulloch, two referees, and seminar participants at several institutions have improved the article. The authors are especially grateful for the detailed comments provided by the associate editor. Glenn Ellison kindly provided the computer code to implement the test of Ellison and Ellison (2000). The first author gratefully acknowledges financial support from the the National Science Foundation (SES-9986298) and the IBM Corporation Faculty Research Fund at the University of Chicago Graduate School of Business. The second author gratefully acknowledges financial support from INQUIRE and Columbia Business School. This article represents the views of the authors and not of INQUIRE.

APPENDIX: PROOFS

Proof of Theorem 1

By theorem 5.4A of Serfling (1980), almost-sure convergence (and thus convergence in probability) follows from the boundedness of the sign(·) function.

Proof of Theorem 2

The result follows from theorem 5.5.1A of Serfling (1980) (which restates the result of Hoeffding 1948), because ς is positive and the second moment Eh^2 exists. Consistency of $\widehat{\varsigma}$ follows from the Slutzky theorem and the law of large numbers.

Proof of Theorem 3

The definition of θ in (10) implies that

$$\begin{aligned}
&\operatorname{sign}(\theta)\\
&\quad= \operatorname{sign}\big[E\big[\operatorname{sign}\big(a_1 y_{[t_1]} + a_2 y_{[t_2]} + \cdots + a_{m+1} y_{[t_{m+1}]}\\
&\qquad\qquad - y_{[t_{m+2}]}\big)\big|\big\{\mathbf{v}_{t_1},\ldots,\mathbf{v}_{t_{m+2}}\big\}\in\mathcal{S}\big]\big]\\
&\quad= \operatorname{sign}\big[E\big[\operatorname{sign}\big(a_1 f(\mathbf{x}_{[t_1]}) + \cdots + a_{m+1} f(\mathbf{x}_{[t_{m+1}]})\\
&\qquad\qquad - f(\mathbf{x}_{[t_{m+2}]}) + a_1\epsilon_{[t_1]} + \cdots + a_{m+1}\epsilon_{[t_{m+1}]}\\
&\qquad\qquad - \epsilon_{[t_{m+2}]}\big)\big|\big\{\mathbf{v}_{t_1},\ldots,\mathbf{v}_{t_{m+2}}\big\}\in\mathcal{S}\big]\big]\\
&\quad= \operatorname{median}\big[\operatorname{sign}\big(a_1 f(\mathbf{x}_{[t_1]}) + \cdots + a_{m+1} f(\mathbf{x}_{[t_{m+1}]})\\
&\qquad\qquad - f(\mathbf{x}_{[t_{m+2}]}) + a_1\epsilon_{[t_1]} + \cdots + a_{m+1}\epsilon_{[t_{m+1}]}\\
&\qquad\qquad - \epsilon_{[t_{m+2}]}\big)\big|\big\{\mathbf{v}_{t_1},\ldots,\mathbf{v}_{t_{m+2}}\big\}\in\mathcal{S}\big].
\end{aligned}$$

Because each error disturbance is (conditionally) symmetrically distributed about 0, the linear combination $a_1\epsilon_{[t_1]} + \cdots + a_{m+1}\epsilon_{[t_{m+1}]} - \epsilon_{[t_{m+2}]}$ in the expression above is (conditionally) symmetrically distributed about 0. Each of the three results in the theorem then follow from Definition 1 because $a_1 f(\mathbf{x}_{[t_1]}) + \cdots + a_{m+1} f(\mathbf{x}_{[t_{m+1}]}) - f(\mathbf{x}_{[t_{m+2}]}) = 0$ for linear f, ≤ 0 for concave f, and ≥ 0 for convex f.

Proofs of Theorems 4 and 5

Because the first two moments of the kernel function for the U-statistic $U_{n,h}(\mathbf{x}^*)$ exist, the proof in the fixed h case is identical to the proof of Theorem 2, because $p_h(\mathbf{x}^*) \to \infty$. Theorem 5 follows from the consistency of $\widehat{\varsigma}(\mathbf{x}^*, h)$ and $p_h(\mathbf{x}^*) \ge 2^m h^m \eta n \to \infty$. The mean of 0 in the f linear case follows from the proof of Theorem 3 [i.e., $\theta(\mathbf{x}^*, h) = 0$], because the kernel functions K_h have no effect on the sign results in the proof of Theorem 3.

Proof of Theorem 6

Because f is linear, Theorem 5 implies that $\widetilde{U}_{n,h}(\mathbf{x}_g^*) \xrightarrow{d} \mathrm{N}(0,1)$ for each $g = 1,\ldots,G$. Although each $\widetilde{U}_{n,h}(\mathbf{x}_g^*)$ has a limiting normal distribution, it remains to be shown that (a) the series of random variables $\widetilde{U}_{n,h}(\mathbf{x}_1^*),\ldots,\widetilde{U}_{n,h}(\mathbf{x}_G^*)$ converge uniformly to standard normal random variables and (b) the random variables are asymptotically independent of each other. Once (a) and (b) are shown, results from classical extreme value theory can be applied directly. In particular, for (a), we show that

$$\lim_{n\to\infty} \max_{g=1,\ldots,G} \sup_t \left|\Pr\left(\widetilde{U}_{n,h}(\mathbf{x}_g^*) < t\right) - \Phi(t)\right| = 0, \tag{A.1}$$

where $\Phi(\cdot)$ denotes the standard normal cdf. For a given g, a Berry–Esséen type theorem for U-statistics (e.g., Serfling 1980, thm. 5.5.1B) implies that

$$\sup_t \left|\Pr\left(\widetilde{U}_{n,h}(\mathbf{x}_g^*) < t\right) - \Phi(t)\right| \leq C\nu(\mathbf{x}_g^*, h)\varsigma(\mathbf{x}_g^*, h)^{-3/2} p_h(\mathbf{x}_g^*)^{-1/2}, \tag{A.2}$$

where C is a universal constant and

$$\nu(\mathbf{x}_g^*, h) \equiv E_{\mathbf{v}_1,\ldots,\mathbf{v}_{m+2}\in V_h(\mathbf{x}_g^*)} |k(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}; \mathbf{x}_g^*, h)|^3.$$

The quantity $\nu(\mathbf{x}_g^*, h)$ is trivially bounded because $|k|$ is bounded. Because $\varsigma(\mathbf{x}_g^*, h)$ is an expectation conditional on $\mathbf{v} \in V_h(\mathbf{x}_g^*)$ and $\tilde{k}(\mathbf{v}; \mathbf{x}_g^*, h)$ is bounded below for $\mathbf{v} \in V_h(\mathbf{x}_g^*)$, it follows that $\varsigma(\mathbf{x}_g^*, h)^{-3/2}$ is bounded above. Combining this with (A.2) yields

$$\sup_t \left|\Pr\left(\widetilde{U}_{n,h}(\mathbf{x}_g^*) < t\right) - \Phi(t)\right| \leq C_2 p_h(\mathbf{x}_g^*)^{-1/2} \tag{A.3}$$

for each $g = 1, \ldots, G$ and some constant C_2. Then

$$\begin{aligned}\max_{g=1,\ldots,G} \sup_t &\left|\Pr\left(\widetilde{U}_{n,h}(\mathbf{x}_g^*) < t\right) - \Phi(t)\right| \\ &\leq \max_{g=1,\ldots,G} C_2 p_h(\mathbf{x}_g^*)^{-1/2} \\ &= C_2\left(\min_{g=1,\ldots,G} p_h(\mathbf{x}_g^*)\right)^{-1/2}.\end{aligned} \tag{A.4}$$

Corresponding to intuition, (A.4) indicates that the rate of uniform convergence is determined by the rate of growth of the number of observations in the smallest bin. Note that $p_h(\mathbf{x}_g^*)$ is distributed as a binomial random variable with parameters n and p_g, where $p_g \equiv \Pr(\mathbf{x} \in V_h(\mathbf{x}_g^*))$. Because $p_g \geq 2^m h^m \eta$ (where $\eta > 0$ from Assumption 3), consider a binomial random variable with parameters n and $2^m h^m \eta$. Because $nh^m \to \infty$ is implied by the rate assumption in the theorem, the asymptotic normal approximation $\tilde{Z} \equiv \mathrm{N}(2^m h^m \eta n, 2^m h^m \eta(1 - 2^m h^m \eta)n)$ applies uniformly for each g. Applying the "minimum" version of theorem 1.5.3 of Leadbetter et al. (1983) to draws $\xi_1, \ldots, \xi_G$ from $(\tilde{Z} - 2^m h^m \eta n)/\sqrt{2^m h^m \eta(1 - 2^m h^m \eta)n}$, it follows that $\Pr(\min(\xi_1, \ldots, \xi_G) > t) \to 1$ (for fixed t) as long as $\sqrt{nh^m \log G}/(nh^m) \to 0$ or, equivalently, $nh^m/\log h^{-1} \to \infty$. Combined with the foregoing argument, $\min_g p_h(\mathbf{x}_g^*) \to \infty$ as $nh^m/\log h^{-1} \to \infty$. The uniform convergence in distribution in (A.1) follows directly. For (b), the same type of asymptotic argument (using a multinomial distribution) can be used. The nonoverlapping bins and bounded kernel function imply that any observation only contributes to a single simplex statistic, meaning that the only dependence across simplex statistics arises from the bin locations of the $\mathbf{x}_i$. Asymptotic independence stems from the fact that the covariance between any two simplex statistics $\widetilde{U}_n(\mathbf{x}_{g_1})$ and $\widetilde{U}_n(\mathbf{x}_{g_2})$ ($g_1 \neq g_2$) depends on $p_{g_1}p_{g_2}$, whereas the variances (from before) depend on $p_{g_1}(1 - p_{g_1})$ and $p_{g_2}(1 - p_{g_2})$. As a result, the covariances between any of the localized simplex statistics trivially approach 0 at a faster rate than the variances, yielding asymptotic independence. Then, because $G \to \infty$ and the statistics $\widetilde{U}_{n,h}(\mathbf{x}_1^*), \ldots, \widetilde{U}_{n,h}(\mathbf{x}_G^*)$ uniformly converge to independent standard normal random variables, the results in (27) and (28) follow directly from theorem 1.5.3 of Leadbetter et al. (1983) and the result in (29) follows directly from theorem 1.8.3 of Leadbetter et al. (1983).

Proof of Theorem 7

Because the proofs in the three cases are basically identical, we consider only case (a) (consistency of the concavity test) in detail. Because $h \to 0$, consider h small enough such that for some g, there exists $\mathbf{x}^* = \mathbf{x}_g^*$ such that $V_h(\mathbf{x}^*) \subset [L_1, U_1] \times \cdots \times [L_m, U_m]$. Then the eigenvalues of $H(f; \mathbf{x})$ are greater than δ for all $\mathbf{x} \in V_h(\mathbf{x}^*)$. Because f is convex on $V_h(\mathbf{x}^*)$, $\widetilde{U}_{n,h}(\mathbf{x}^*) \to \infty$ when $p_h(\mathbf{x}^*) \to \infty$. The critical result to be shown is that M_n is of order $n^{1/2}h^{(4+m)/2}$. If this is true, then the probability of M_n exceeding the critical value tends to 1 when $n^{1/2}h^{(4+m)/2}$ exceeds the order of b_n [the dominant term in (31)]. Because the order of b_n is $\sqrt{\log h^{-1}}$, this occurs if $nh^{4+m}/\log h^{-1} \to \infty$. To get the result for M_n, consider the probability limit of $U_{n,h}(\mathbf{x}^*)$,

$$\begin{aligned}\theta(\mathbf{x}^*, h) &= \Pr(\{\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}\} \in \mathcal{S}) \\ &\quad\times E\left[\prod_{j=1}^{m+2} K_h\left(\mathbf{x}_{[j]} - \mathbf{x}^*\right) \mathrm{sign}(w(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}))\right| \\ &\qquad\qquad \{\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}\} \in \mathcal{S}\Bigg].\end{aligned} \tag{A.5}$$

The first term, $\Pr(\{\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}\} \in \mathcal{S})$, and the K_h functions have no effect on the order of $\theta(\mathbf{x}^*, h)$, leading us to focus on the sign($\cdot$) term. For any $(m + 2)$-tuples in $V_h(\mathbf{x}^*)$ we have, conditional on $\{\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}\} \in \mathcal{S}$ [for notational ease, we that assume $\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}$ is already ordered such that $\mathbf{x}_{m+2} \in \Delta(\mathbf{x}_1, \ldots, \mathbf{x}_{m+1})$],

$$\begin{aligned}\Pr&\left(w(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}) > 0\right) \\ &= \Pr\left(a_1 f(\mathbf{x}_1) + \cdots + a_m f(\mathbf{x}_m)\right. \\ &\qquad \left. + (1 - a_1 - \cdots - a_m) f(\mathbf{x}_{m+1}) - f(\mathbf{x}_{m+2}) > \tilde{\epsilon}\right),\end{aligned} \tag{A.6}$$

where $\tilde{\epsilon} \equiv \epsilon_{m+2} - a_1\epsilon_1 - \cdots - a_m\epsilon_m - (1 - a_1 - \cdots - a_m)\epsilon_{m+1}$. Expand the left side of the inequality,

$$\begin{aligned}a_1 f(\mathbf{x}_1) &+ \cdots + a_m f(\mathbf{x}_m) \\ &+ (1 - a_1 - \cdots - a_m) f(\mathbf{x}_{m+1}) - f(\mathbf{x}_{m+2}) \\ &= a_1[f(\mathbf{x}_1) - f(\mathbf{x}_{m+1})] + \cdots + a_m[f(\mathbf{x}_m) - f(\mathbf{x}_{m+1})] \\ &\quad + [f(\mathbf{x}_{m+1}) - f(\mathbf{x}_{m+2})] \\ &= a_1(\mathbf{x}_1 - \mathbf{x}_{m+1})' f'(\mathbf{x}_1) + \cdots + a_m(\mathbf{x}_m - \mathbf{x}_{m+1})' f'(\mathbf{x}_m) \\ &\quad + (\mathbf{x}_{m+1} - \mathbf{x}_{m+2})' f'(\mathbf{x}_{m+1}) + o(h).\end{aligned}$$

The last equality follows from Taylor's rule, with the remainder term $o(h)$ because the distances $\mathbf{x}_1 - \mathbf{x}_{m+1}, \ldots, \mathbf{x}_{m+1} - \mathbf{x}_{m+2}$

are all of order h. Use the fact that $\mathbf{x}_{m+2} = a_1\mathbf{x}_1 + \cdots + a_m\mathbf{x}_m + (1 - a_1 - \cdots - a_m)\mathbf{x}_{m+1}$ and apply Taylor's rule again,

$$\begin{aligned} &a_1 f(\mathbf{x}_1) + \cdots + a_m f(\mathbf{x}_m) \\ &\quad + (1 - a_1 - \cdots - a_m) f(\mathbf{x}_{m+1}) - f(\mathbf{x}_{m+2}) \\ &\quad = a_1(\mathbf{x}_1 - \mathbf{x}_{m+1})'[f'(\mathbf{x}_1) - f'(\mathbf{x}_{m+1})] + \cdots \\ &\qquad + a_m(\mathbf{x}_m - \mathbf{x}_{m+1})'[f'(\mathbf{x}_m) - f'(\mathbf{x}_{m+1})] \\ &\quad = a_1(\mathbf{x}_1 - \mathbf{x}_{m+1})' H(f; \mathbf{x}_1)(\mathbf{x}_1 - \mathbf{x}_{m+1}) + \cdots \\ &\qquad + a_m(\mathbf{x}_m - \mathbf{x}_{m+1})' H(f; \mathbf{x}_m)(\mathbf{x}_m - \mathbf{x}_{m+1}) + o(h^2). \qquad \text{(A.7)} \end{aligned}$$

Combining this equation with the condition on the eigenvalues implies that $\text{sign}(w(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}))$ will be a positive number of order h^2 [because the elements of each $(\mathbf{x}_j - \mathbf{x}_i)$ are of order h]. Then the order of $\widetilde{U}_{n,h}(\mathbf{x}^*)$ (and, in turn, M_n) is $\sqrt{nh^m}$ times h^2 (or $n^{1/2}h^{(4+m)/2}$), because $p_h(\mathbf{x}^*)$ is of order nh^m. This result completes the proof.

Proof of Theorem 8

The basic idea is to verify that we can write $\sqrt{n}(U_n(\hat{\boldsymbol\beta}) - \mu(\boldsymbol\beta))$ as

$$\begin{aligned} \sqrt{n}\big(U_n(\hat{\boldsymbol\beta}) - \mu(\boldsymbol\beta)\big) &= \sqrt{n}\big(U_n(\boldsymbol\beta) - \mu(\boldsymbol\beta)\big) \\ &\quad + \sqrt{n}(\hat{\boldsymbol\beta} - \boldsymbol\beta)'\nabla_{\mathbf{b}}\mu(\hat{\boldsymbol\beta}) + o_p(1). \qquad \text{(A.8)} \end{aligned}$$

Then theorem 2.13 of Randles (1982) yields the desired result, using the joint asymptotic normality of $\sqrt{n}(U_n(\boldsymbol\beta) - \mu(\boldsymbol\beta))$ and $\sqrt{n}(\hat{\boldsymbol\beta} - \boldsymbol\beta)$. The conditions in (46) and the boundedness of the kernel function h allow application of the central limit theorem. To show (A.8), note that

$$\begin{aligned} &\sqrt{n}\big(U_n(\hat{\boldsymbol\beta}) - \mu(\boldsymbol\beta)\big) \\ &\quad = \sqrt{n}\big(U_n(\hat{\boldsymbol\beta}) - \mu(\hat{\boldsymbol\beta})\big) + \sqrt{n}\big(\mu(\hat{\boldsymbol\beta}) - \mu(\boldsymbol\beta)\big). \qquad \text{(A.9)} \end{aligned}$$

Theorem 2.8 of Randles (1982) gives

$$\sqrt{n}\big(U_n(\hat{\boldsymbol\beta}) - \mu(\hat{\boldsymbol\beta})\big) - \sqrt{n}\big(U_n(\boldsymbol\beta) - \mu(\boldsymbol\beta)\big) \xrightarrow{p} 0 \qquad \text{(A.10)}$$

if the following condition (condition 2.3 of Randles 1982) holds:

Suppose that there is a neighborhood of $\boldsymbol\beta$, call it $N(\boldsymbol\beta)$, and a constant $K_1 > 0$, such that if $\mathbf{b} \in N(\boldsymbol\beta)$ and $D(\mathbf{b}, d)$ is a sphere centered at $\mathbf{b}$ with radius d satisfying $D(\mathbf{b}, d) \subset N(\boldsymbol\beta)$; then

$$\begin{aligned} E\bigg[\sup_{\mathbf{b}' \in D(\mathbf{b},d)} &\big|h(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}; \mathbf{b}') \\ &- h(\mathbf{v}_1, \ldots, \mathbf{v}_{m+2}; \mathbf{b})\big|\bigg] \le K_1 d. \qquad \text{(A.11)} \end{aligned}$$

Verifying this condition is straightforward given the boundedness of the error density function and the support of $\mathbf{z}$ (from Assumptions 5 and 6) (see, e.g., the proof in sec. 2 of Randles 1984). Then a combination of (A.9) and (A.10), along with differentiability of $\mu(\cdot)$ at $\boldsymbol\beta$, yields (A.8) and the desired result.

[Received November 2002. Revised May 2004.]

REFERENCES

Abel, A. B., and Eberly, J. (1994), "A Unified Model of Investment Under Uncertainty," *American Economic Review*, 84, 1369–1384.

Abrevaya, J., and Jiang, W. (2003), "Pairwise-Slopes Statistics for Testing Curvature," mimeo.

Arcones, M. A., and Giné, E. (1992), "On the Bootstrap of U and V Statistics," *The Annals of Statistics*, 20, 655–674.

Barberis, N., and Shleifer, A. (2003), "Style Investing," *Journal of Financial Economics*, 68, 161–199.

Barnett, S. A., and Sakellaris, P. (1998), "Nonlinear Response of Firm Investment to Q: Testing a Model of Convex and Non-Convex Adjustment Costs," *Journal of Monetary Economics*, 42, 261–288.

Becker, C., Ferson, W., Myers, D., and Schill, M. (1999), "Conditional Market Timing With Benchmark Investors," *Journal of Financial Economics*, 52, 119–148.

Bollen, N., and Busse, J. (2001), "On the Timing Ability of Mutual Fund Managers," *Journal of Finance*, 56, 1075–1094.

Borenstein, S., and Farrell, J. (1999), "Do Investors Forecast Fat Firms? Evidence From the Gold Mining Industry," Working Paper 7075, National Bureau of Economic Research.

Chevalier, J., and Ellison, G. (1997), "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, 105, 1167–1200.

Davis, C. E., and Quade, D. (1978), "U-statistics for Skewness or Symmetry," *Communications in Statistics, Part A—Theory and Methods*, A7, 413–418.

Diack, C. A. T. (2001), "Testing the Shape of a Regression Curve," *Comptes Rendus de a l'Académie des Sciences*, 333, 677–680.

Diack, C. A. T., and Thomas-Agnan, C. (1998), "A Nonparametric Test of the Non-Convexity of Regression," *Journal of Nonparametric Statistics*, 9, 335–362.

Dümbgen, L., and Spokoiny, V. G. (2001), "Multiscale Testing of Qualitative Hypotheses," *The Annals of Statistics*, 29, 124–152.

Ellison, G., and Ellison, S. F. (2000), "A Simple Framework for Nonparametric Specification Testing," *Journal of Econometrics*, 96, 1–23.

Fama, E. F., and French, K. R. (1993), "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Ghosal, S., Sen, A., and van der Vaart, A. W. (2000), "Testing Monotonicity of Regression," *The Annals of Statistics*, 28, 1054–1082.

Goetzmann, W., Ingersoll, J., Jr., and Ivkovich, Z. (2000), "Monthly Measurement of Daily Timers," *Journal of Financial and Quantitative Analysis*, 35, 257–290.

Hardy, G. H., Littlewood, J. E., and Polya, G. (1988), *Inequalities*, Cambridge, U.K.: Cambridge University Press.

Henriksson, R. D., and Merton, R. C. (1981), "On Market Timing and Investment Performance II: Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business*, 54, 513–534.

Hildreth, C. (1954), "Point Estimates of Ordinates of Concave Functions," *Journal of the American Statistical Association*, 49, 598–619.

Hoeffding, W. (1948), "A Class of Statistics With Asymptotically Normal Distribution," *The Annals of Mathematical Statistics*, 19, 293–325.

Hollander, M., and Wolfe, D. A. (1999), *Nonparametric Statistical Methods*, New York: Wiley.

Jiang, W. (2003), "A Nonparametric Test of Market Timing," *Journal of Empirical Finance*, 10, 399–425.

Kendall, M. G. (1938), "A New Measure of Rank Correlation," *Biometrika*, 30, 91–93.

Leadbetter, M. R., Lindgren, G., and Rootzen, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer-Verlag.

Mammen, E. (1991), "Nonparametric Regression Under Qualitative Smoothness Assumptions," *The Annals of Statistics*, 19, 741–759.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995), *Microeconomic Theory*, New York: Oxford University Press.

Matzkin, R. L. (1994), "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, Vol. 4, eds. R. F. Engle and D. L. McFadden, New York: Elsevier Science, pp. 2523–2558.

Mincer, J. (1974), *Schooling, Experience, and Earnings*, New York: Columbia University Press.

Moore, J. (1999), *Mathematic Methods for Economic Theory I*, Heidelburg, Germany: Springer-Verlag.

Murphy, K. M., and Welch, F. (1990), "Empirical Age-Earnings Profiles," *Journal of Labor Economics*, 8, 202–229.

Powell, J. L. (1987), "Semiparametric Estimation of Bivariate Latent Variable Models," Working Paper 8704, University of Wisconsin SSRI.

Randles, R. H. (1982), "On the Asymptotic Normality of Statistics With Estimated Parameters," *The Annals of Statistics*, 10, 462–474.

——— (1984), "On Tests Applied to Residuals," *Journal of the American Statistical Association*, 79, 349–354.

Randles, R. H., Fligner, M. A., Policello, G. E. II, and Wolfe, D. A. (1980), "An Asymptotically Distribution-Free Test for Symmetry versus Asymmetry," *Journal of the American Statistical Association*, 75, 168–172.

Robinson, P. M. (1988), "Root-*n* Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

Schlee, W. (1982), "Nonparametric Test of the Monotony and Convexity of Regression," in *Nonparametric Statistical Inference*, Vol. 2, eds. B. V. Gnedenko, M. L. Puri, and I. Vincze, Amsterdam: North-Holland, pp. 823–836.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Treynor, J. L., and Mazuy, K. K. (1966), "Can Mutual Funds Outguess the Market," *Harvard Business Review*, 44, 131–136.

Wang, Y. (1992), "Nonparametric Estimation Subject to Shape Restrictions," unpublished doctoral dissertation, University of California-Berkeley, Dept. of Statistics.

Yatchew, A. (1992), "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, 8, 435–451.

——— (1997), "An Elementary Estimator of the Partial Linear Model," *Economics Letters*, 57, 135–143.