Ann. Inst. Statist. Math.
Vol. 48, No. 2, 315–336 (1996)

NONPARAMETRIC TESTS FOR BOUNDS ON THE DERIVATIVE OF A REGRESSION FUNCTION

NANCY E. HECKMAN[1]* AND BING LI[2]**

[1] *Statistics Department, University of British Columbia, Vancouver BC, Canada V6T 1Z2*
[2] *Statistics Department, Pennsylvania State University, University Park, PA 16802, U.S.A.*

(Received August 17, 1994; revised July 20, 1995)

Abstract. We consider two tests of the null hypothesis that the k-th derivative of a regression function is uniformly bounded by a specified constant. These tests can be used to study the shape of the regression function. For instance, we can test for convexity of the regression function by setting $k = 2$ and the constant equal to zero. Our tests are based on k-th order divided difference of the observations. The asymptotic distribution and efficacies of these tests are computed and simulation results presented.

Key words and phrases: Derivative of a regression function, convexity, divided differences.

1. Introduction

In some regression analyses, determining certain characteristics of the regression function is more important than obtaining an accurate estimate of the function. For instance, in evolutionary ecology, it is important to determine if the dependence of the probability of survival upon some physical trait is monotone, convex, or concave. The shape of the function determines how the physical trait evolves in the population (see, e.g. Schluter (1988)). Another example of research involving the shape of a regression function is in the analysis of human growth data. The "pubertal spurt" is a rapid period of growth typically occurring at twelve years of age in girls and fourteen years of age in boys. Determining the age span of this spurt involves studying the convexity of height as a function of age. (See, e.g. Gasser *et al.* (1984), for analyses of these types of data.) Here, we study testing that the k-th derivative of the regression function is bounded by a specified number.

* Research supported by Natural Sciences and Engineering Research Council of Canada Grant OGP0007969.

** Research supported by National Science Foundation Grant DMS-9306738.

Suppose that the regression function f has k continuous derivatives on a compact interval, which we take to be $[0,1]$. We wish to test the hypothesis

$$H_0 : f^{(k)}(t) \leq \gamma \quad \text{for all} \quad t \in [0,1]. \tag{1.1}$$

The test is based on observations

$$X_i = f(t_i) + \epsilon_i, \qquad 1 \leq i \leq n, \tag{1.2}$$

where the ϵ_i's are independent and identically distributed with mean zero and $t_i = i/(n+1)$. (If one wishes to test $f^{(k)}(t) \geq \gamma$ instead of (1.1), one uses the proposed statistics with X_i replaced by $-X_i$. If one wishes to test that $f^{(k)}(t) \leq g(t)$ for all t, for some specified g, one uses the proposed statistics, but with X_i replaced by $X_i - G(t_i)$, where G is any function with $G^{(k)}(t) \equiv g(t)$.) Two test statistics, T and S, are defined in Section 2. Both statistics count the number of divided differences of the X_i's which are greater than or equal to a specified threshold, and H_0 is rejected if this number is too large. If $k = 1$ and the specified threshold is chosen appropriately, S is the Mann T-test (Mann (1945)) which is used to test the hypothesis that f is constant, versus the alternative that f is non-decreasing. The Mann test rejects the null hypothesis if the number of times that $X_i - X_j > 0$, $i > j$, is too large. The statistic T is equivalent to a test proposed by Sen (1965) in a time series setting. Sen studied T's null distribution, but not its efficacy, as we do here.

In Section 2 we give the asymptotic distributions of S and T assuming f is on the "boundary" of the null hypothesis, $f^{(k)}(t) \equiv \gamma$, and assuming that $f = f_n$, a sequence of regression functions local to the null hypothesis. These results can be used to calculate the asymptotic significance levels of the tests and the efficacies. The efficacies are discussed in Section 4. Unfortunately, when $k \neq 1$, the asymptotic variances of the statistics depend upon unknown parameters. In Section 3, we present estimates of these parameters. Section 5 contains simulation results. The simulation results and the efficacy calculations indicate that the test statistic S is a good one, particularly if the data contain outliers.

It is possible to construct statistics to test $H_0 : \gamma_1 \leq f^{(k)}(t) \leq \gamma_2$ for all $t \in [0,1]$ by counting the number of divided differences of the X_i's that fall within a specified interval. For a sequence of functions $\{f_n\}$ approaching a fixed function, the asymptotic distributions of the test statistics are easily derived by the methods of this paper. However, it is very difficult, if not impossible, to determine the significance levels of the tests, that is, to calculate the supremum, over H_0, of the probability of rejecting H_0. Therefore, we do not analyse these tests here.

2. Definitions and main results

Suppose that (1.2) holds. We test the null hypothesis $H_0 : f \in \mathcal{F}_{\gamma,k}$, where

$$\mathcal{F}_{\gamma,k} \equiv \mathcal{F} = \{f : [0,1] \to \mathcal{R} : f \text{ has } k \text{ continuous derivatives and } f^{(k)}(t) \leq \gamma \text{ for all } t \in [0,1]\}.$$

Our test statistics count the number of k-th order divided differences of the X_i's that are larger than a specified value. For instance, when $k = 1$ and the specified value is c, S is equal to the number of times that $(X_i - X_j)/(t_i - t_j) \geq c$ for $i > j$. The statistic T depends on an integer m, chosen by the user. When, say, $k = 1$ and $m = 3$, T is equal to the number of times that $(X_i - X_{i-3})/(t_i - t_{i-3}) \geq c$. When $c = 0$ and $k = 1$, S is equivalent to the statistic studied by Mann (1945). When $c = 0$, the statistic T is equivalent to one proposed by Sen (1965).

To define S and T for general k, we recursively define $(\Delta^{m,k}\boldsymbol{X})_i$, the k-th order divided difference with spacing m evaluated at i of the n-vector $\boldsymbol{X} = (X_1, \ldots, X_n)^t$ as follows. (The dependence upon n is suppressed in the notation.) For $k = 0$ and $1 \leq i \leq n$

$$(\Delta^{m,0}\boldsymbol{X})_i = X_i.$$

For $k = 1$ and $m + 1 \leq i \leq n$

$$(\Delta^{m,1}\boldsymbol{X})_i = \frac{n+1}{m}(X_i - X_{i-m}) = \frac{n+1}{m}((\Delta^{m,0}\boldsymbol{X})_i - (\Delta^{m,0}\boldsymbol{X})_{i-m}).$$

For general k and $km + 1 \leq i \leq n$

$$(\Delta^{m,k}\boldsymbol{X})_i = \frac{n+1}{m}((\Delta^{m,k-1}\boldsymbol{X})_i - (\Delta^{m,k-1}\boldsymbol{X})_{i-m}).$$

To count the number of times that a k-th order divided difference exceeds the specified value c let

$$W_i^m = W_i^{m,k}(c) = I\{(\Delta^{m,k}\boldsymbol{X})_i \geq c\}$$

for $km + 1 \leq i \leq n$. Then our two test statistics are

$$T_c^n = T_c^n(m,k) = \sum_{i=km+1}^{n} W_i^m$$

and

$$S_c^n = S_c^n(k) = \sum_{m=1}^{[(n-1)/k]} \sum_{i=km+1}^{n} W_i^m = \sum_{m=1}^{[(n-1)/k]} T_c^n(m,k).$$

We will reject H_0 if either T_c^n or S_c^n is too large.

The asymptotic significance levels and the efficacies of the test statistics are given in Theorems 2.1 to 2.3 below. We assume that the distribution of ϵ_1 is continuous. The Appendix contains the proofs of these theorems, in addition to proofs of the asymptotic normality of the test statistics under more general conditions.

First consider the statistic S_c^n. To define its asymptotic mean and variance, we introduce random variables U, $U_1(l,l')$, and $U_2(l,l')$, $l, l' = 0, \ldots, k$. Let

$$U = \left(\frac{m}{n+1}\right)^k (\Delta^{m,k}\boldsymbol{\epsilon})_i, \tag{2.1}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^t$. Notice that the distribution of U depends only upon k and the distribution of ϵ_1. For fixed l, l', m, m', i, and i' with $\{i-km, \ldots, i-m, i\} \cap \{i'-km', \ldots, i'-m', i'\} = \{i-lm\} = \{i'-l'm'\}$, $(\Delta^{m,k}\epsilon)_i$ and $(\Delta^{m',k}\epsilon)_{i'}$ involve exactly one common ϵ, namely $\epsilon_{i-lm} = \epsilon_{i'-l'm'}$. Thus the joint distribution of $(m/(n+1))^k(\Delta^{m,k}\epsilon)_i$ and $(m'/(n+1))^k(\Delta^{m',k}\epsilon)_{i'}$ does not depend upon n, m, m', i, or i': it depends only upon the position of the "overlap" determined by l, l', and k and upon the distribution of ϵ_1. Therefore, we can define

$$(U_1(l,l'), U_2(l,l')) = \left(\left(\frac{m}{n+1}\right)^k (\Delta^{m,k}\epsilon)_i, \left(\frac{m'}{n+1}\right)^k (\Delta^{m',k}\epsilon)_{i'}\right).$$

The dependence of the distributions of U and $(U_1(l,l'), U_2(l,l'))$ on k and the distribution of ϵ_1 is suppressed.

Let

$$\mu_S^* = \mu_S^*(c,\gamma) = \int_{\alpha=0}^{1/k} (1-k\alpha)P\{U \geq \alpha^k(c-\gamma)\}d\alpha,$$

and

$$\sigma_S^{*2} = \sigma_S^{*2}(c,\gamma) = \sum_{l,l'=0}^{k} \int_{\alpha,\alpha'=0}^{1/k} [1-\max\{l\alpha, l'\alpha'\} - \max\{(k-l)\alpha, (k-l')\alpha'\}]_+$$
$$\times \operatorname{cov}(I\{U_1(l,l') \geq \alpha^k(c-\gamma)\}, I\{U_2(l,l') \geq \alpha'^k(c-\gamma)\})d\alpha d\alpha'.$$

Here u_+ is the positive part of u.

Let $f_\gamma^*(t) \equiv f^*(t) = \gamma t^k/k!$ (so $f^{*(k)} \equiv \gamma$) and let P_f be the probability assuming that f is the true regression function.

THEOREM 2.1. *If $\sigma_S^* > 0$, then*

$$\sup_{f\in\mathcal{F}} P_f\left\{\frac{S_c^n - n^2\mu_S^*}{n^{3/2}\sigma_S^*} \geq z_\beta\right\} = P_{f^*}\left\{\frac{S_c^n - n^2\mu_S^*}{n^{3/2}\sigma_S^*} \geq z_\beta\right\} \to \beta,$$

where z_β is the $1-\beta$ quantile of the standard normal distribution. If $c=\gamma$ and the distribution of ϵ_1 is symmetric, then $n^2\mu_S^*$ can be replaced by the exact expectation of S_c^n

$$E_0(S_c^n) = \frac{n^*}{2}\left(n - \frac{k}{2}(n^*+1)\right)$$

where n^* is the largest integer in $(n-1)/k$.

To study the asymptotic distribution of T_c^n, define new random variables $V_1(l)$ and $V_2(l)$, $l = 0, \ldots, k-1$:

$$(V_1(l), V_2(l)) = \left(\frac{m}{n+1}\right)^k ((\Delta^{1,k}\epsilon)_{i+k-l}, (\Delta^{1,k}\epsilon)_i).$$

Since $(\Delta^{1,k}\boldsymbol{\epsilon})_i$ and $(\Delta^{1,k}\boldsymbol{\epsilon})_{i+k-l}$ involve the $l+1$ common errors, $\epsilon_{i-l}, \ldots, \epsilon_i$, the joint distribution of $V_1(l)$ and $V_2(l)$ depends only upon l, k, and the distribution of ϵ_1. For $\alpha_0 \in (0, 1/k)$, let

$$\mu_T^* = \mu_T^*(c, \gamma, \alpha_0) = (1 - k\alpha_0)P\{U \geq \alpha_0^k(c - \gamma)\}$$

and

$$\begin{aligned}\sigma_T^{*2} &= \sigma_T^{*2}(c, \gamma, \alpha_0) \\ &= 2\sum_{l=0}^{k-1}(1 - (2k - l)\alpha_0)_+ \operatorname{cov}(I\{V_1(l) \geq \alpha_0^k(c - \gamma)\}, I\{V_2(l) \geq \alpha_0^k(c - \gamma)\}) \\ &\quad + (1 - k\alpha_0)\operatorname{var}(I\{U \geq \alpha_0^k(c - \gamma)\}).\end{aligned}$$

THEOREM 2.2. *Suppose that*

$$m = m_n = [n\alpha_0] \quad \textit{with} \quad \alpha_0 \in (0, 1/k) \tag{2.2}$$

and $[x]$ *the greatest integer less than or equal to* x. *Suppose that* $\sigma_T^* > 0$. *Then*

$$\sup_{f \in \mathcal{F}} P_f\left\{\frac{T_c^n - n\mu_T^*}{n^{1/2}\sigma_T^*} \geq z_\beta\right\} = P_{f^*}\left\{\frac{T_c^n - n\mu_T^*}{n^{1/2}\sigma_T^*} \geq z_\beta\right\} \to \beta.$$

If $c = \gamma$ *and the distribution of* ϵ_1 *is symmetric, then* $n\mu_T^*$ *can be replaced by the exact expectation of* T_c^n

$$E_0(T_c^n) = (n - km)/2.$$

We now consider the efficacies of the test statistics under local alternatives. Although results for general local alternatives are given in the Appendix, we present results here only for alternatives local to f^*, the "worst case" function in $\mathcal{F}$. We assume that $X_i^n = f_n(i/(n+1)) + \epsilon_i$ with $f_n(t) = f^*(t) + \delta_n n^{-1/2}g(t)$, where g has k continuous derivatives. To define the efficacies, we need to define a functional which is similar to $\Delta^{m,k}$, the divided difference vector. We call this new functional, $\Delta_g^{\alpha,k}$, the k-th order divided difference functional with spacing α of a function g. It is defined for g a function from the unit interval to the reals, k integer, and α in $(0, 1/k)$. $\Delta_g^{\alpha,k}$ is defined recursively:

$$\begin{aligned}&\Delta_g^{\alpha,0}(t) = g(t) \qquad t \in [0, 1] \\ &\Delta_g^{\alpha,1}(t) = \frac{1}{\alpha}[g(t) - g(t - \alpha)] = \frac{1}{\alpha}[\Delta_g^{\alpha,0}(t) - \Delta_g^{\alpha,0}(t - \alpha)] \qquad t \in [\alpha, 1] \\ &\Delta_g^{\alpha,k}(t) = \frac{1}{\alpha}[\Delta_g^{\alpha,k-1}(t) - \Delta_g^{\alpha,k-1}(t - \alpha)] \qquad t \in [k\alpha, 1].\end{aligned}$$

We define

$$\Delta_g^{0,k}(t) = g^{(k)}(t).$$

Note that, for $\boldsymbol{g} = (g(1/(n+1)), \ldots, g(n/(n+1)))^t$, $\Delta_g^{m/(n+1),k}(i/(n+1)) = (\Delta^{m,k}\boldsymbol{g})_i$.

THEOREM 2.3. *Suppose that* U, *as defined in* (2.1), *has a continuous density* f_U *and that* $\delta_n \to \delta < \infty$. *Suppose* $\sigma_S^* > 0$ *and let*

$$e_S = e_S(g) = \frac{1}{\sigma_S^*} \int_{\alpha=0}^{1/k} \int_{t=\alpha k}^{1} \alpha^k f_U(\alpha^k(c-\gamma)) \Delta_g^{\alpha,k}(t) dt d\alpha.$$

Then, under f_n,

$$\frac{S_c^n - n^2 \mu_S^*}{n^{3/2}\sigma_S^*} - \delta e_S \Longrightarrow N(0,1).$$

Suppose that $\sigma_T^* > 0$ *and let*

$$e_T = e_T(g, \alpha_0) = \frac{\alpha_0^k f_U(\alpha_0^k(c-\gamma))}{\sigma_T^*} \int_{t=\alpha_0 k}^{1} \Delta_g^{\alpha_0,k}(t) dt.$$

If $m = m_n$ *satisfies* (2.2), *then under* f_n,

$$\frac{T_c^n - n\mu_T^*}{n^{1/2}\sigma_T^*} - \delta e_T \Longrightarrow N(0,1).$$

3. Estimation of σ_S^* and σ_T^*

The results of Theorems 2.1 and 2.2 indicate how we would use the statistics S_c^n and T_c^n to construct asymptotic level α tests of $H_0 : f^{(k)}(t) \le \gamma$. If we assume that the regression error distribution is symmetric about zero and set $c = \gamma$ then the asymptotically level α tests reject H_0 if $S_c^n \ge E_0(S_c^n) + z_\alpha n^{3/2} \sigma_S^*(c,c)$ or if $T_c^n \ge E_0(T_c^n) + z_\alpha n^{1/2} \sigma_T^*(c,c,\alpha_0)$. However, the parameters σ_S^* and σ_T^* are typically unknown. In this section we consider two types of estimates of σ_S^* and σ_T^* (or equivalently, of the standard errors of S_c^n and T_c^n). One estimate is based on the asymptotics of Theorems 2.1 and 2.2. The other estimate uses bootstrapped samples. In simulation studies, we found that the asymptotic-based estimate of σ_S^{*2} performed poorly, often taking on negative values. The asymptotic-based estimate of σ_T^* performed very well. Therefore, we will only define the asymptotic based estimate of σ_T^*.

We present an estimate of σ_T^* that satisfies

$$\lim_{n\to\infty} \sup_{f \in \mathcal{F}_n} P_f\{T_c^n \ge E_0(T_c^n) + z_\alpha n^{1/2} \hat{\sigma}_T^*\} = \alpha$$

where

$$\mathcal{F}_n = \{f \in \mathcal{F} : -\rho_n n^k \le f^{(k)}(t) \text{ for all } t\},$$

and ρ_n is any sequence of positive numbers tending to zero. As usual, T_c^n is constructed using m_n as in (2.2). We see that we must estimate

$$\sigma_T^{*2} = \sum_{l=0}^{k-1} 2(1-(2k-l)\alpha_0)_+ C(l) + (1-k\alpha_0)/4$$

where $C(l)$ is the covariance between $I\{V_1(l) \geq 0\}$ and $I\{V_2(l) \geq 0\}$. We estimate σ_T^{*2} by substituting estimates of $C(l)$. Our estimate, $\hat{C}(l)$, is the sample covariance between the collection of pairs

$$\begin{aligned}&\{(I\{(\Delta^{1,k}\boldsymbol{X})_i \geq c\}, I\{(\Delta^{1,k}\boldsymbol{X})_{i-k+l} \geq c\}), i = 2k-l+1,\ldots,n\} \\ &\quad \cup \{(I\{(\Delta^{2,k}\boldsymbol{X})_i \geq c\}, I\{(\Delta^{2,k}\boldsymbol{X})_{i-2k+2l} \geq c\}), i = 4k-2l+1,\ldots,n\}.\end{aligned}$$

Then $\hat{C}(l)$ converges to $C(l)$ in probability, uniformly on $\mathcal{F}_n$ and our estimate of σ_T^{*2} converges uniformly to σ_T^{*2}. (It is possible to use only the first set of pairs in the estimation of $C(l)$, but simulations showed that including the second set of pairs greatly increased the finite sample accuracy of the estimate.)

To estimate the standard error of S_c^n (or T_c^n) from a sample (X_i, t_i), $i = 1,\ldots,n$, we draw with replacement a bootstrap sample of size n from (X_i, t_i). We then calculate the resulting value of S_c^n (or T_c^n). This process is repeated many times, and the standard deviation of the resulting values of S_c^n (or T_c^n) is used as the estimate of the standard error. The exact theoretical properties of this procedure are unknown but simulation studies, presented in Section 5, indicate that the resulting estimates are good.

4. Efficacy

Here we study relative Pitman efficiencies of tests of $H_0 : f^{(k)}(t) \equiv \gamma$ versus $H_1 : f^{(k)}(t) > \gamma$ for some t. Three tests are considered: the two tests introduced in Section 2, which use statistics S_c^n and T_c^n, and the least squares regression test. The efficacies are calculated with respect to local alternatives of the form $f_n(t) = f^*(t) + n^{-1/2}\delta_n g(t)$, where f^* is as in Theorem 2.1, $\delta_n \to \delta$, finite, and g has k continuous derivatives. Throughout we assume that, in the construction of T_c^n, $m = m_n$ as in (2.2). No one of the three tests is clearly superior to the others and, for large classes of alternatives, all three have no power (see Theorem 4.1).

From Theorem 2.3, we see that the Pitman efficacy of S_c^n is $e_S = e_S(g)$ and the Pitman efficacy of T_c^n is $e_T = e_T(g, \alpha_0)$.

For the least squares regression test, assume that $f(t) = \beta_0 + \beta_1(t-1/2) + \cdots + \beta_k(t-1/2)^k$, with the β_j's unknown. Let $M = M(n,k)$ denote the $n \times (k+1)$ design matrix with $M_{ij} = (i/(n+1) - 1/2)^{j-1}$. Then the least squares estimate of β_k is $\hat{\beta}_k = ((M^tM)^{-1}M^t\boldsymbol{X})_{k+1}$, and the least squares regression test rejects H_0 at approximate significance level α if $(k!)\hat{\beta}_k \geq \gamma + (k!)z_\alpha \operatorname{var}^{1/2}(\hat{\beta}_k)$. (We could base our test on the parameterization $f(t) = \beta_0^* + \beta_1^* t + \cdots + \beta_k^* t^k$, rejecting H_0 when $\hat{\beta}_k^*$ is too large. However, these two parameterizations are equivalent—that is, $\hat{\beta}_k = \hat{\beta}_k^*$. The β_k parameterization is more convenient for the proof of Theorem 4.1 below.)

To calculate the efficacy of the least squares regression test, assume that the variance of ϵ_i is σ^2. Then, under the local alternative hypothesis, the variance of $\hat{\beta}_k$ is $\sigma^2((M^tM)^{-1})_{k+1,k+1}$, the mean is $\beta_k + n^{-1/2}\delta_n((M^tM)^{-1}M^t\boldsymbol{g})_{k+1}$, where $\boldsymbol{g} = (g(1/(n+1)), \ldots, g(n/(n+1)))^t$, and so the efficacy is

$$e_{LS} = \lim_n n^{-1/2} \frac{((M^tM)^{-1}M^t\boldsymbol{g})_{k+1}}{\sigma((M^tM)^{-1})^{1/2}_{k+1,k+1}}.$$

We now show that all three tests have no power against a special class of local alternatives. We assume that $n \geq k+1$, so that our estimators are well-defined.

THEOREM 4.1. *If $g^{(k)}(t) = -g^{(k)}(1-t)$ for all t, then $e_{LS} = e_S = e_T = 0$.*

PROOF. If $g^{(k)}(t) + g^{(k)}(1-t) = 0$ for all t, then $g(t) + (-1)^k g(1-t)$ is a polynomial of degree $k-1$. This fact will be used in calculating all three efficacies.

First consider e_{LS}. Let $\hat{\boldsymbol{b}} = (M^tM)^{-1}M^t\boldsymbol{g} = (\hat{b}_0, \ldots, \hat{b}_k)^t$. We will show that $\hat{b}_k = 0$. Let $\boldsymbol{g}^* = (g(n/(n+1)), g((n-1)/(n+1)), \ldots, g(1/(n+1)))^t$. Since $g(t) + (-1)^k g(1-t)$ is a polynomial of degree $k-1$, for $n \geq k+1$, $\boldsymbol{g} + (-1)^k \boldsymbol{g}^*$ can be fit perfectly via least squares by a polynomial of degree $k-1$,

$$[(M^tM)^{-1}M^t(\boldsymbol{g} + (-1)^k\boldsymbol{g}^*)]_{k+1} = 0.$$

Let $\hat{\boldsymbol{b}}^* = (M^tM)^{-1}M^t\boldsymbol{g}^*$. Thus $\hat{b}_k + (-1)^k\hat{b}^*_k = 0$. But, as argued below, $\hat{b}^*_j = (-1)^j\hat{b}_j$, $j = 0, \ldots, k$, and so $0 = \hat{b}_k + (-1)^{2k}\hat{b}_k = 2\hat{b}_k$.

The fact that $\hat{b}^*_j = (-1)^j\hat{b}_j$, $j = 0, \ldots, k$, follows directly from the fact that $\hat{\boldsymbol{b}}$ minimizes

$$\sum_i \left[g\left(\frac{i}{n+1}\right) - \sum_j b_j \left(\frac{i}{n+1} - \frac{1}{2}\right)^j \right]^2$$

and $\hat{\boldsymbol{b}}^*$ minimizes

$$\sum_i \left[g\left(\frac{n+1-i}{n+1}\right) - \sum_j b_j \left(\frac{i}{n+1} - \frac{1}{2}\right)^j \right]^2$$
$$= \sum_l \left[g\left(\frac{l}{n+1}\right) - \sum_j b_j(-1)^j \left(\frac{l}{n+1} - \frac{1}{2}\right)^j \right]^2.$$

To show that $e_S = e_T = 0$, we show that $\int_{t=\alpha k}^1 \Delta_g^{\alpha,k}(t)dt = 0$ for all α. Let $g^*(t) = g(1-t)$. Then, by induction on k, $\Delta_{g^*}^{\alpha,k}(t) = (-1)^k\Delta_g^{\alpha,k}(1+k\alpha-t)$ for $k\alpha \leq t \leq 1$. Since $g(t) + (-1)^k g(1-t) = g(t) + (-1)^k g^*(t)$ is a polynomial of degree $k-1$,

$$0 = \Delta_g^{\alpha,k}(t) + (-1)^k\Delta_{g^*}^{\alpha,k}(t) = \Delta_g^{\alpha,k}(t) + \Delta_g^{\alpha,k}(1+k\alpha-t).$$

Therefore

$$\int_{t=k\alpha}^{1} \Delta_g^{\alpha,k}(t)dt = -\int_{t=k\alpha}^{1} \Delta_g^{\alpha,k}(1+k\alpha-t)dt = -\int_{x=k\alpha}^{1} \Delta_g^{\alpha,k}(x)dx. \qquad \square$$

We now consider the case of testing for convexity, that is, with $k = 2$, and we set $c = \gamma$. We assume that the density of ϵ_i is symmetric about 0. Then we easily calculate that

$$\begin{aligned}
e_{LS} &= \frac{\sqrt{5}}{\sigma}\int_0^1 g(t)(6t^2-6t+1)dt,\\
e_S &= \frac{f_U(0)}{\sigma_S^*}\int_{\alpha=0}^{1/2}\int_{t=2\alpha}^{1}(g(t)-2g(t-\alpha)+g(t-2\alpha))dtd\alpha,\\
e_T &= \frac{f_U(0)}{\sigma_T^*}\int_{t=2\alpha_0}^{1}(g(t)-2g(t-\alpha_0)+g(t-2\alpha_0))dt,\\
\sigma_S^{*2} &= \frac{1}{4}\left(P\{\epsilon_3-2\epsilon_2+\epsilon_1\ge 0, \epsilon_5-2\epsilon_4+\epsilon_3\ge 0\}-\frac{1}{4}\right)\\
&\quad+\frac{1}{4}\left(P\{\epsilon_5-2\epsilon_4+\epsilon_3\ge 0, \epsilon_4-2\epsilon_2+\epsilon_1\ge 0\}-\frac{1}{4}\right)\\
&\quad+\frac{1}{12}\left(P\{\epsilon_5-2\epsilon_4+\epsilon_3\ge 0, \epsilon_6-2\epsilon_4+\epsilon_2\ge 0\}-\frac{1}{4}\right),
\end{aligned}$$

and

$$\begin{aligned}
\sigma_T^{*2} &= 2(1-4\alpha_0)_+\left(P\{\epsilon_3-2\epsilon_2+\epsilon_1\ge 0, \epsilon_5-2\epsilon_4+\epsilon_3\ge 0\}-\frac{1}{4}\right)\\
&\quad+2(1-3\alpha_0)_+\left(P\{\epsilon_4-2\epsilon_3+\epsilon_2\ge 0, \epsilon_3-2\epsilon_2+\epsilon_1\ge 0\}-\frac{1}{4}\right)\\
&\quad+(1-2\alpha_0)/4.
\end{aligned}$$

The test based on T with α_0 close to a half should do very well when $g(1) - g(1/2)$ is much bigger than $g(1/2) - g(0)$. For the special case $g(t) = \rho(t-1/2)^r$, $r \ge 2$,

$$\begin{aligned}
e_{LS} &= \frac{\sqrt{5}\rho}{\sigma}\frac{1}{2^{r+1}}[1+(-1)^r]\frac{r}{(r+1)(r+3)},\\
e_S &= \frac{f_U(0)\rho}{\sigma_S^*}\frac{1}{2^{r+2}}[1+(-1)^r]\frac{r}{(r+1)(r+2)},
\end{aligned}$$

and

$$\begin{aligned}
e_T(g,\alpha_0) &= \frac{f_U(0)\rho}{\sigma_T^*}\frac{1}{r+1}[1+(-1)^r]\\
&\quad\cdot\left(\frac{1}{2^{r+1}}-2\left(\frac{1}{2}-\alpha_0\right)^{r+1}+\left(\frac{1}{2}-2\alpha_0\right)^{r+1}\right).
\end{aligned}$$

Since $\sigma_T^* \to 0$ as $\alpha_0 \to 1/2$, we easily see that

$$\lim_{\alpha_0\to 1/2}\lim_{r\to\infty} e_T(g,\alpha_0)/e_S = \infty,$$

and

$$\lim_{\alpha_0 \to 1/2} \lim_{r \to \infty} e_T(g, \alpha_0)/e_{LS} = \infty$$

for r not an odd integer. However we do not recommend using the statistic T, since its efficacy is highly dependent upon the choice of α_0. Our recommendation is also supported by the simulation results of Section 5.

We now compare e_{LS} and e_S for general g in the case that the ϵ_i's are normally distributed with mean zero and variance σ^2. In this case,

$$\begin{aligned}\frac{e_S}{e_{LS}} &= \frac{1}{\sqrt{60\pi}\sigma_S^*} \frac{\int_{\alpha=0}^{1/2}\int_{t=2\alpha}^{1}(g(t)-2g(t-\alpha)+g(t-2\alpha))dtd\alpha}{\int_0^1 g(t)(6t^2-6t+1)dt} \\ &\approx 1.372\frac{\int_{\alpha=0}^{1/2}\int_{t=2\alpha}^{1}(g(t)-2g(t-\alpha)+g(t-2\alpha))dtd\alpha}{\int_0^1 g(t)(6t^2-6t+1)dt}.\end{aligned}$$

When $g(t) = \rho(t-1/2)^r$, $r \geq 2$, r not an odd integer, then

$$\frac{e_S}{e_{LS}} \approx \frac{1.372}{2}\left(1+\frac{1}{r+2}\right) \leq 0.858.$$

When $r = 2$, that is, when the least squares test is the ideal test, we see that the relative efficiency is fairly high—over 85%.

While the least squares test always performs better than S under the above assumptions, under some conditions S will perform better than the least squares test. For instance, the convergence of S does not require the existence of moments of ϵ_i. Also, for certain functions g, S will be more efficient than the least squares test statistic. In fact, for any positive constant M, we can find a function g so that $e_{LS} \leq -M$ and $e_S \geq M$. (Since we reject the null hypothesis for large values of the test statistics, a positive efficacy indicates a strong tendency to reject the null hypothesis. A negative efficacy indicates that the null hypothesis is not usually rejected.) Let δ be a small positive number and, for $C > 0$, define g with continuous second derivative such that $g''(t) = -4C$ for $t \in [0, 1/4-\delta]$, $g''(t) = C$ for $t \in [1/4+\delta, 1/2-\delta]$, $g''(1/2) = 0$, g'' linear on $[1/4-\delta, 1/4+\delta] \cup [1/2-\delta, 1/2]$, and $g''(t) = g''(1-t)$ for $t \in [1/2, 1]$. To calculate e_{LS} and e_S we first rewrite the integral expressions using integration by parts, and then approximate g'' by a step function which equals $-4C$ on $[0, 1/4] \cup [3/4, 1]$ and equals C on $(1/4, 3/4)$. (The approximation holds for δ small.) Thus

$$\begin{aligned}e_{LS} &= \frac{\sqrt{5}}{2\sigma}\int_0^{1/2}[g''(t)+g''(1-t)]t^2(1-t)^2dt \\ &\sim \frac{-3\sqrt{5}}{5120\sigma}C\end{aligned}$$

and

$$\begin{aligned}e_S &= \frac{f_U(0)}{12\sigma_S^*}\int_0^{1/2}[g''(t)+g''(1-t)]t^2(3-4t)dt \\ &\sim \frac{f_U(0)}{1536\sigma_S^*}C.\end{aligned}$$

Our claim follows by taking C sufficiently large.

5. Simulations

To study the finite sample properties of our asymptotic results and to compare the finite sample characteristics of our test statistics, we carried out simulation studies. Our focus was on testing $H_0 : f''(t) \leq 0$ for all t versus $H_1 : f''(t) > 0$ for some t. In this case, the function f^* defined prior to Theorem 2.1 is $f^*(t) \equiv 0$. This function is "on the boundary" between H_0 and H_1. Our simulated data followed model (1.2) with the ϵ_i's standard normal random variables. We considered two values of n, $n = 50$ and $n = 100$. For each function f considered and for each sample size, we conducted 1000 simulations. In all cases, the statistics T_c and S were calculated using $c = 0$. For sample size $n = 50$, the T statistics were calculated for values of $m = 5$, 10, and 15. (In the tables, these T statistics are denoted as $T(5)$, $T(10)$ and $T(15)$ respectively.) For sample size $n = 100$, the T statistics were calculated for values of $m = 10$, 20 and 30. Note that, under f^*, $W_i^{m,2} = I\{\epsilon_i - 2\epsilon_{i-m} + \epsilon_{i-2m} \geq 0\}$, and thus the null distributions of T and S do not depend on the variance of ϵ_i.

Table 1 contains summary information for the T and S statistics under the model (1.2), with $f(t) \equiv 0$. The simulated means are given with standard deviations in parentheses. The asymptotic means and accompanying parenthesized standard deviations are those calculated from the results in Theorems 2.1 and 2.2. (The asymptotic means are actually the values of $E_0(S)$ and $E_0(T)$.) The standard errors of T were estimated in the two ways described in Section 3: one using the asymptotic results of Theorem 2.2, and the other via bootstrapping, with 100 bootstrap samples taken in each simulation. The standard errors of S were calculated only via bootstrapping, again with 100 bootstrap samples taken in each simulation. The averages of the estimates are given, with the standard deviations in parentheses. The simulated means and standard deviations of the test statistics are very close to the true means and asymptotic standard deviations. In addition, the estimated standard errors are extremely accurate. In most cases, the bootstrapped estimate of the standard error is not as accurate as the asymptotic estimate. However, we still recommend the use of the bootstrapped estimate, as it will never yield negative values.

Table 2 contains simulation results for a 0.05 level test of H_0. We considered the model (1.2) for the functions $f(t) = -t^2$, 0, t^2, and t^3. Using the statistic S, we reject H_0 at this level if $S \geq E_0(S) + 1.645\hat{\sigma}_S$, where $\hat{\sigma}_S$ is our bootstrapped estimate of the standard error of S. We considered two tests based on $T(m)$: one test rejects H_0 if $T(m) \geq n/2 - m + 1.645\hat{\sigma}_T$, where $\hat{\sigma}_T$ is the asymptotics-based estimate of the standard error of $T(m)$. The other test rejects H_0 if $T(m) \geq n/2 - m + 1.645\hat{\sigma}_{TB}$, where $\hat{\sigma}_{TB}$ is the bootstrapped estimate. These two tests are denoted $T(m)$ and $TB(m)$, respectively. The least-squares regression based test, denoted LS, rejects H_0 if $\hat{\beta}_2 \geq t_{0.05}\hat{\sigma}_{\beta_2}$, where $t_{0.05}$ is the upper 5th percentile of a t distribution with degrees of freedom $n-3$, and $\hat{\sigma}_{\beta_2}$ is the usual estimate of the standard error of $\hat{\beta}_2$. As expected, we see that the least squares test is the best, but that the test using S is comparable. The tests based on T do not perform very well.

In each of the studies of $T(5)$ with a sample size of $n = 50$, the estimate of

Table 1. Distribution of the test statistics with $f(t) \equiv 0$.

$n = 50$	$T(5)$	$T(10)$	$T(15)$	S
Simulated mean	20.00 (1.91)	14.97 (1.90)	10.03 (1.93)	299.84 (20.25)
Asymptotic mean	20.00 (1.86)	15.00 (1.84)	10.00 (1.94)	300.00 (18.77)
SE estimate/asymptotic	1.82 (0.31)	1.83 (0.13)	1.94 (0.06)	—
SE estimate/bootstrap	1.85 (0.14)	1.83 (0.12)	1.95 (0.13)	19.58 (1.46)
$n = 100$	$T(10)$	$T(20)$	$T(30)$	S
Simulated mean	40.22 (2.76)	30.07 (2.62)	19.97 (2.74)	1223.6 (52.5)
Asymptotic mean	40.00 (2.63)	30.00 (2.60)	20.00 (2.76)	1225.0 (53.1)
SE estimate/asymptotic	2.60 (0.29)	2.60 (0.12)	2.76 (0.05)	—
SE estimate/bootstrap	2.62 (0.19)	2.58 (0.19)	2.75 (0.19)	54.27 (3.99)

Simulation results are based on 1000 simulation runs, with observations $X_i = \epsilon_i$, ϵ_i standard normal.

Table 2. Percent of times H_0 is rejected in a level 0.05 test.

Sample size	Statistic	$f(t) = -t^2$	$f(t) = 0$	$f(t) = t^2$	$f(t) = t^3$
$n = 50$	S	2.4 (0.5)	5.1 (0.7)	11.4 (1.0)	14.2 (1.1)
	$T(5)$	7.2 (0.8)	6.0 (0.8)	6.4 (0.8)	6.8 (0.8)
	$TB(5)$	6.2 (0.8)	4.7 (0.7)	5.2 (0.7)	4.8 (0.7)
	$T(10)$	5.8 (0.7)	6.1 (0.8)	5.7 (0.7)	6.0 (0.8)
	$TB(10)$	5.3 (0.7)	4.8 (0.7)	5.4 (0.7)	5.9 (0.7)
	$T(15)$	4.2 (0.6)	3.8 (0.6)	4.4 (0.6)	4.4 (0.6)
	$TB(15)$	5.4 (0.7)	4.5 (0.7)	4.8 (0.7)	5.0 (0.7)
	LS	1.8 (0.4)	5.5 (0.7)	13.5 (1.1)	17.9 (1.2)
$n = 100$	S	0.6 (0.2)	4.5 (0.7)	14.0 (1.1)	22.0 (1.3)
	$T(10)$	5.3 (0.7)	7.4 (0.8)	7.1 (0.8)	7.5 (0.8)
	$TB(10)$	5.1 (0.7)	6.4 (0.8)	6.5 (0.8)	6.8 (0.8)
	$T(20)$	4.6 (0.7)	5.2 (0.7)	5.4 (0.7)	6.0 (0.8)
	$TB(20)$	4.8 (0.7)	5.9 (0.7)	5.8 (0.7)	5.8 (0.7)
	$T(30)$	5.1 (0.7)	4.7 (0.7)	4.8 (0.7)	4.1 (0.6)
	$TB(30)$	5.1 (0.7)	4.7 (0.7)	4.8 (0.7)	4.0 (0.6)
	LS	0.9 (0.3)	5.2 (0.7)	17.7 (1.2)	28.3 (1.4)

Results are based on 1000 simulation runs, with observations $X_i = f(t_i) + \epsilon_i$, where $t_i = i/(n+1)$ and ϵ_i is standard normal. The entries in the table are π = percent of times H_0 is rejected and, in parenthesis, the standard error $\sqrt{\pi(1-\pi)/1000} \times 100$.

the variance of $T(5)$ based on the asymptotics was negative in one of the thousand simulation runs. The figures given in Tables 1 and 2 for $T(5)$ are based on the 999 simulations that resulted in positive estimates of the variance.

Table 3 contains simulation results for comparing the least squares procedure with the test based on S in the presence of outliers. The sample size was $n = 50$. The simulations were identical to those presented in Table 2, except for the distribution of the ϵ_i's. Here, we generated fifty standard normal random variables, then chose N of these at random, replacing them with normal random variables with mean zero and standard deviation ten. We considered $N = 10$ and $N = 20$. As one would expect, the statistic S is much less influenced by outliers than the least squares procedure. For sample size 100 (not presented here), the differences between the two procedures were less striking.

Under all simulation configurations, the distributions of the test statistics were very close to normal.

Table 3. Percent of times H_0 is rejected in a level 0.05 test.

Number of outliers	Statistic	$f(t) = -t^2$	$f(t) = 0$	$f(t) = t^2$	$f(t) = t^3$
10	S	2.5 (0.5)	4.7 (0.7)	8.9 (0.9)	10.5 (1.0)
	LS	4.0 (0.6)	4.9 (0.7)	6.6 (0.8)	7.4 (0.8)
20	S	4.8 (0.7)	5.9 (0.7)	8.0 (0.9)	9.6 (0.9)
	LS	3.8 (0.6)	4.4 (0.6)	5.2 (0.7)	5.6 (0.7)

Results are based on 1000 simulation runs, with observations $X_i = f(t_i) + \epsilon_i$, where $t_i = i/(n+1)$ and ϵ_i is standard normal, except that outliers are normal with mean 0 and standard deviation 10. The entries in the table are π = percent of times H_0 is rejected and, in parenthesis, the standard error $\sqrt{\pi(1-\pi)/1000} \times 100$.

Appendix

Proofs of Theorems

We first prove that f^* is the "worst case" function in $\mathcal{F}$, that is that

$$\sup_{f \in \mathcal{F}} P_f\{S_c^n \geq s\} = P_{f^*}\{S_c^n \geq s\}$$

and

$$\sup_{f \in \mathcal{F}} P_f\{T_c^n \geq s\} = P_{f^*}\{T_c^n \geq s\}.$$

This is an immediate consequence of the following lemma.

LEMMA A.1. *Let $f \in \mathcal{F}$, and X_i as in (1.2) with $t_i = i/(n+1)$. Define $X_i^* = f^*(t_i) + \epsilon_i$. Then $I\{(\Delta^{m,k}\boldsymbol{X})_i \geq c\} \leq I\{(\Delta^{m,k}\boldsymbol{X}^*)_i \geq c\}$.*

PROOF. Let $\boldsymbol{f} = (f(t_1), \ldots, f(t_n))^t$, and $\boldsymbol{f}^*$ defined similarly. Since $(\Delta^{m,k}\boldsymbol{f})_i = f^{(k)}(\xi)$ for some $\xi \in [t_{i-km}, t_i]$, $(\Delta^{m,k}\boldsymbol{f})_i \le \gamma \equiv (\Delta^{m,k}\boldsymbol{f}^*)_i$. Therefore

$$\begin{aligned} I\{(\Delta^{m,k}\boldsymbol{X})_i &\ge c\} \\ &= I\{(\Delta^{m,k}\boldsymbol{f})_i + (\Delta^{m,k}\boldsymbol{\epsilon})_i \ge c\} \le I\{(\Delta^{m,k}\boldsymbol{f}^*)_i + (\Delta^{m,k}\boldsymbol{\epsilon})_i \ge c\} \\ &= I\{(\Delta^{m,k}\boldsymbol{X}^*)_i \ge c\}. \qquad \square \end{aligned}$$

We now study the distribution of S_c^n under the model

$$X_i^n = f_n\left(\frac{i}{n+1}\right) + \epsilon_i$$

where the ϵ_i's are independent and identically distributed and $\{f_n\}$ is any sequence of functions in $\mathcal{F}$ (see Theorem A.1). In Theorem A.2, we study the distribution for a fixed function f, calculating the asymptotic mean and variance. Theorem A.3 states the efficacy of S_c^n. The analogous results for the distribution of T_c^n are in Theorems A.4–A.6.

THEOREM A.1. *Let $\hat{S}_c^n$ be the Hájek projection of S_c^n (Hájek (1968)):*

$$\hat{S}_c^n = \sum_{j=1}^n \mathrm{E}(S_c^n \mid \epsilon_j) \equiv \sum_{j=1}^n \eta_j.$$

Suppose that there exists a sequence $\rho_n \to 0$ such that

$$\liminf_{n\to\infty} \rho_n n^{-2}\,\mathrm{var}(\hat{S}_c^n) > 0.$$

Then

$$\frac{\mathrm{var}(S_c^n)}{\mathrm{var}(\hat{S}_c^n)} \to 1$$

and

$$\frac{S_c^n - \mathrm{E}(S_c^n)}{\mathrm{var}^{1/2}(S_c^n)} \Longrightarrow N(0,1).$$

PROOF. For convenience, the superscripts and subscripts n and c will be omitted.

To prove the first statement, we show that

$$\rho n^{-2}(\mathrm{var}(S) - \mathrm{var}(\hat{S})) \to 0. \tag{A.1}$$

Write

$$\mathrm{var}(S) = \sum_{m,m'=1}^{[(n-1)/k]} \sum_{i=km+1}^{n} \sum_{i'=km'+1}^{n} \mathrm{cov}(W_i^m, W_{i'}^{m'})$$

and

$$\begin{aligned}\operatorname{var}(\hat{S}) &= \sum_{j=1}^{n} \operatorname{var}(\eta_j) \\ &= \sum_{m,m'=1}^{[(n-1)/k]} \sum_{i=km+1}^{n} \sum_{i'=km'+1}^{n} \sum_{j=1}^{n} \operatorname{cov}(\mathrm{E}(W_i^m \mid \epsilon_j), \mathrm{E}(W_{i'}^{m'} \mid \epsilon_j)).\end{aligned}$$

Since $W_i^m = I\{(\Delta^{m,k}\boldsymbol{X})_i \geq c\}$ involves $X_{i-km}, \ldots, X_i$, the calculation of the individual terms in the variances of S and $\hat{S}$ will depend upon

$$A(m,i,m',i') = \{i - km, \ldots, i - m, i\} \cap \{i' - km', \ldots, i' - m', i'\}$$

for $1 \leq m,\ m' \leq [(n-1)/k]$, $km + 1 \leq i \leq n$, and $km' + 1 \leq i' \leq n$. We partition the indices (m, i, m', i') as follows.

$$\begin{aligned} A_0 &= \{(m,i,m',i') : A(m,i,m',i') = \emptyset\}, \\ A_1 &= \{(m,i,m',i') : \text{ the cardinality of } A(m,i,m',i') = 1\}, \\ A_2 &= \{(m,i,m',i') : \text{ the cardinality of } A(m,i,m',i') \geq 2\}. \end{aligned}$$

We first study the variance of S. If $(m,i,m',i') \in A_0$, then W_i^m and $W_{i'}^{m'}$ are independent and so $\operatorname{cov}(W_i^m, W_{i'}^{m'}) = 0$. One easily shows that the cardinality of A_2 is $O(n^2)$. Therefore

$$\operatorname{var}(S) = \sum_{(m,i,m',i') \in A_1} \operatorname{cov}(W_i^m, W_{i'}^{m'}) + O(n^2).$$

To study the variance of $\hat{S}$, first note that if $j \notin A(m,i,m',i')$, then either ϵ_j and W_i^m are independent, or ϵ_j and $W_{i'}^{m'}$ are independent, or both. Thus, either $\mathrm{E}(W_i^m \mid \epsilon_j) = \mathrm{E}(W_i^m)$ or $\mathrm{E}(W_{i'}^{m'} \mid \epsilon_j) = \mathrm{E}(W_{i'}^{m'})$ or both, and so $\operatorname{cov}(\mathrm{E}(W_i^m \mid \epsilon_j), \mathrm{E}(W_{i'}^{m'} \mid \epsilon_j)) = 0$.

Therefore

$$\begin{aligned}\operatorname{var}(\hat{S}) = \Bigg[&\sum_{(m,i,m',i') \in A_1} \sum_{j \in A(m,i,m',i')} \\ &+ \sum_{(m,i,m',i') \in A_2} \sum_{j \in A(m,i,m',i')} \Bigg] \operatorname{cov}(\mathrm{E}(W_i^m \mid \epsilon_j), \mathrm{E}(W_{i'}^{m'} \mid \epsilon_j)).\end{aligned}$$

If $(m,i,m',i') \in A_1$, then there exists exactly one j in $A(m,i,m',i')$. For this value of j, W_i^m and $W_{i'}^{m'}$ are conditionally independent given ϵ_j. Therefore $\operatorname{cov}(\mathrm{E}(W_i^m \mid \epsilon_j), \mathrm{E}(W_{i'}^{m'} \mid \epsilon_j)) = \operatorname{cov}(W_i^m, W_{i'}^{m'})$. Since the cardinality of A_2 is $O(n^2)$ and the cardinality of $A(m,i,m',i')$ is at most $k+1$,

$$\operatorname{var}(\hat{S}) = \sum_{(m,i,m',i') \in A_1} \operatorname{cov}(W_i^m, W_{i'}^{m'}) + O(n^2) = \operatorname{var}(S) + O(n^2).$$

This completes the proof of (A.1).

Since

$$\frac{\mathrm{E}(S-\hat{S})^2}{\mathrm{var}(S)} = \frac{\mathrm{var}(S)-\mathrm{var}(\hat{S})}{\mathrm{var}(S)} \to 0$$

the asymptotic normality of S will follow once we show that

$$\frac{\hat{S}-\mathrm{E}(\hat{S})}{\mathrm{var}^{1/2}(\hat{S})} \Longrightarrow N(0,1).$$

By the Lindeberg-Feller Theorem (see e.g. Serfling (1980)), it suffices to show that

$$\lim_n \mathrm{var}^{-1}(\hat{S}) \sum_{j=1}^{n} \mathrm{E}((\eta_j - \mathrm{E}\eta_j)^2 I\{|\eta_j - \mathrm{E}\eta_j| \geq \phi\, \mathrm{var}^{1/2}(\hat{S})\}) = 0$$

for all $\phi > 0$. We show that $I\{|\eta_j - \mathrm{E}\eta_j| \geq \phi\,\mathrm{var}^{1/2}(\hat{S})\} = 0$ for all j, for n sufficiently large. Now

$$\eta_j - \mathrm{E}(\eta_j) = \sum_{m=1}^{[(n-1)/k]} \sum_{i=km+1}^{n} [\mathrm{E}(W_i^m \mid \epsilon_j) - \mathrm{E}(W_i^m)].$$

Since, for fixed j, $\mathrm{E}(W_i^m \mid \epsilon_j) = \mathrm{E}(W_i^m)$ unless $j = i - lm$ for some $l = 0, \ldots, k$, the inner summation over i in the above contains at most $k+1$ terms. Thus $\eta_j - \mathrm{E}\eta_j$ is a sum of $O(n)$ bounded random variables. But the variance of $\hat{S}$ grows quickly: by assumption, there exists C such that $n^{-2}\,\mathrm{var}(\hat{S}) \geq C/\rho \to \infty$. □

We now consider the asymptotic distribution of S_c^n for a fixed function f. Recall the definitions of $\Delta_f^{\alpha,k}$, U, $U_1(l,l')$, and $U_2(l,l')$ in Section 2. Let

$$\begin{aligned}
\mu_S = \mu_S(f) &= \int_{\alpha=0}^{1/k} \int_{t=k\alpha}^{1} P\{U \geq \alpha^k(c - \Delta_f^{\alpha,k}(t))\}\,dt\,d\alpha,\\
\sigma_f^2(l,l',\alpha,\alpha',t) &= \mathrm{cov}(I\{U_1(l,l') \geq \alpha^k(c-\Delta_f^{\alpha,k}(t+l\alpha))\},\\
&\qquad I\{U_2(l,l') \geq \alpha'^k(c-\Delta_f^{\alpha',k}(t+l'\alpha'))\}),\\
\beta(l,l',\alpha,\alpha',t) &= I\{\max\{(k-l)\alpha,(k-l')\alpha'\} < t < 1-\max\{l\alpha,l'\alpha'\}\},
\end{aligned}$$

and

$$\sigma_S^2 = \sigma_S^2(f) = \sum_{l,l'=0}^{k} \int_{\alpha,\alpha'=0}^{1/k} \int_{t=0}^{1} \beta(l,l',\alpha,\alpha',t)\sigma_f^2(l,l',\alpha,\alpha',t)\,dt\,d\alpha\,d\alpha'.$$

In Theorems A.2 and A.3, we require that the integrands of $\mu_S(f)$ and $\sigma_S^2(f)$ be Riemann integrable. It suffices that f has k continuous derivatives and the distributions of U and $(U_1(l,l'),U_2(l,l'))$ be continuous. One easily shows that the distributions of U and $(U_1(l,l'),U_2(l,l'))$ are continuous if the distribution of ϵ_1 is continuous.

THEOREM A.2. *Suppose that $f_n \equiv f$, $\sigma_S^2(f) > 0$, and that the integrands in $\sigma_S^2(f)$ and $\mu_S(f)$ are Riemann integrable. Then, under f,*

$$\frac{S_c^n - \mathrm{E}(S_c^n)}{\mathrm{var}^{1/2}(S_c^n)} \Longrightarrow N(0,1),$$

$$n^{-3}\,\mathrm{var}(\hat{S}_c^n) \sim n^{-3}\,\mathrm{var}(S_c^n) \to \sigma_S^2,$$

and

$$\mathrm{E}(S_c^n) = n^2\mu_S + o(n^2)$$

where $\hat{S}_c^n$ is the projection of S_c^n, as defined in Theorem A.1. *If $f = f^*$, as defined in Theorem* 2.1, *and if U has a continuous density, then*

$$\mathrm{E}(S_c^n) = n^2\mu_S + O(n).$$

PROOF. Write

$$W_i^m = I\left\{\left(\frac{m}{n+1}\right)^k (\Delta^{m,k}\boldsymbol{\epsilon})_i \ge \left(\frac{m}{n+1}\right)^k \left(c - \Delta_f^{m/(n+1),k}\left(\frac{i}{n+1}\right)\right)\right\}.$$

By Theorem A.1, the asymptotic normality of S and the asymptotic variance expression will follow if we show that $\mathrm{var}(\hat{S}) \sim n^3\sigma_S^2$. Since $\mathrm{E}(W_i^m \mid \epsilon_j) = \mathrm{E}(W_i^m)$ if $j \ne i - lm$, $l = 0, \dots, k$,

$$\eta_j - \mathrm{E}(\eta_j) = \sum_{l=0}^{k} \sum_{m=1}^{[(n-1)/k]} (\mathrm{E}(W_{j+lm}^m \mid \epsilon_j) - \mathrm{E}(W_{j+lm}^m)) I\{km + 1 \le j + lm \le n\}$$

and so

$$\begin{aligned}
\text{(A.2)}\qquad \mathrm{var}(\hat{S}) &= \sum_{j=1}^{n} \sum_{l,l'=0}^{k} \sum_{m,m'=1}^{[(n-1)/k]} [I\{\max\{(k-l)m, (k-l')m'\} \\
&\qquad\qquad < j < n + 1 - \max\{lm, l'm'\}\} \\
&\qquad\qquad \mathrm{cov}(\mathrm{E}(W_{j+lm}^m \mid \epsilon_j), \mathrm{E}(W_{j+l'm'}^{m'} \mid \epsilon_j))] \\
&= \sum_{l,l'=0}^{k} \sum_{j=1}^{n} \sum_{m,m'=1}^{[(n-1)/k]} \beta\left(l, l', \frac{m}{n+1}, \frac{m'}{n+1}, \frac{j}{n+1}\right) \\
&\qquad\qquad \cdot \sigma_f^2\left(l, l', \frac{m}{n+1}, \frac{m'}{n+1}, \frac{j}{n+1}\right) \\
&\sim n^3\sigma_S^2.
\end{aligned}$$

To study the asymptotic expected value of S, write

$$\begin{aligned}
\mathrm{E}(S) &= \sum_{m=1}^{[(n-1)/k]} \sum_{i=mk+1}^{n} P\left\{U \ge \left(\frac{m}{n+1}\right)^k \left(c - \Delta_f^{m/(n+1),k}\left(\frac{i}{n+1}\right)\right)\right\} \\
&\sim n^2\mu_S.
\end{aligned}$$

The stronger statement concerning $\mathrm{E}(S)$ is easily shown. □

THEOREM A.3. *Let $f_n(x) = f(x) + \delta_n n^{-1/2} g(x)$, where f and g have k continuous derivatives, $\sigma_S^2(f) > 0$, and $\delta_n \to \delta < \infty$. Assume that $(U_1(l,l'), U_2(l,l'))$ have continuous joint densities and that U has a continuous density, f_U. Then*

$$n^{-3/2}(\mathrm{E}_{f_n}(S_c^n) - \mathrm{E}_f(S_c^n)) \to \delta \int_{\alpha=0}^{1/k} \int_{t=k\alpha}^{1} \alpha^k \Delta_g^{\alpha,k}(t) f_U(\alpha^k(c - \Delta_f^{\alpha,k}(t)))\,dt\,d\alpha$$

and under f_n

$$\frac{S_c^n - \mathrm{E}_{f_n}(S_c^n)}{n^{3/2}\sigma_S(f)} \Longrightarrow N(0,1).$$

PROOF. To prove the statement concerning the means of S_c^n, write

$$\begin{aligned}
\mathrm{E}_{f_n}(W_i^m) - \mathrm{E}_f(W_i^m) &= \mathrm{E}(W_i^m) - \mathrm{E}(W_i^{*m}) \\
&= P\{(\Delta^{m,k}\boldsymbol{\epsilon})_i \ge c - (\Delta^{m,k}\boldsymbol{f})_i - \frac{\delta}{\sqrt{n}}(\Delta^{m,k}\boldsymbol{g})_i\} \\
&\quad - P\{(\Delta^{m,k}\boldsymbol{\epsilon})_i \ge c - (\Delta^{m,k}\boldsymbol{f})_i\} \\
&= P\left\{U \ge \left(\frac{m}{n+1}\right)^k \left(c - (\Delta^{m,k}\boldsymbol{f})_i - \frac{\delta}{\sqrt{n}}(\Delta^{m,k}\boldsymbol{g})_i\right)\right\} \\
&\quad - P\left\{U \ge \left(\frac{m}{n+1}\right)^k (c - (\Delta^{m,k}\boldsymbol{f})_i)\right\} \\
&= \frac{\delta}{\sqrt{n}} \left(\frac{m}{n+1}\right)^k \Delta_g^{m/(n+1),k}\left(\frac{i}{n+1}\right) \\
&\quad \cdot f_U\left[\left(\frac{m}{n+1}\right)^k \left(c - \Delta_f^{m/(n+1),k}\left(\frac{i}{n+1}\right)\right)\right] \\
&\quad + o(n^{-1/2})
\end{aligned}$$

uniformly in i and m. Thus the statement concerning $\mathrm{E}_{f_n}(S_c^n) - \mathrm{E}_f(S_c^n)$ holds.
Similarly

$$\sigma_{f_n}^2\left(l, l', \frac{m}{n+1}, \frac{m'}{n+1}, \frac{j}{n+1}\right) - \sigma_f^2\left(l, l', \frac{m}{n+1}, \frac{m'}{n+1}, \frac{j}{n+1}\right) = O(n^{-1/2})$$

and so, using A.2,

$$n^{-3}(\mathrm{var}(\hat{S}_c^n) - \mathrm{var}(\hat{S}_c^{*n})) \to 0$$

where $\hat{S}_c^n$ and $\hat{S}_c^{*n}$ are the projections of S_c^n and S_c^{*n} respectively. By Theorem A.2, $n^{-3}\,\mathrm{var}(\hat{S}_c^{*n}) \to \sigma_S^2(f)$, and so $n^{-3}\,\mathrm{var}(\hat{S}_c^n) \to \sigma_S^2(f)$. The theorem follows from Theorem A.1. □

We now study T_c^n. Assume that $m = m_n$ as in (2.2).

THEOREM A.4. *Suppose that*

$$\liminf_{n\to\infty} \operatorname{var}(T_c^n) = \infty.$$

Then

$$\frac{T_c^n - \mathrm{E}(T_c^n)}{\operatorname{var}^{1/2}(T_c^n)} \Longrightarrow N(0,1).$$

PROOF. T_c^n can be rewritten

$$T_c^n = \sum_{i=1}^m Z_i$$

where

$$Z_i = Z_i(n,c,m,k) = W_{i+km}^m + W_{i+(k+1)m}^m + \cdots + W_{i+(k+l(i))m}^m$$

and $l(i) = l(i,n,m,k)$ is as large as possible, that is $i + (k + l(i))m \le n < i + (k + l(i) + 1)m$. $Z_1, Z_2, \ldots, Z_m$ are independent.

By the Lindeberg-Feller Theorem it suffices to show that, for all $\phi > 0$,

$$\operatorname{var}^{-1}(T) \sum_1^m \mathrm{E}((Z_i - \mathrm{E}(Z_i))^2 I\{|Z_i - \mathrm{E}(Z_i)| \ge \phi \operatorname{var}^{1/2}(T)\}) \to 0.$$

This follows immediately, since Z_i is a sum of at most $n/m = O(1/\alpha_0) = O(1)$ bounded terms and $\operatorname{var}(T) \to \infty$ by assumption. Thus, for n sufficiently large, $I\{|Z_i - E(Z_i)| \ge \phi \operatorname{var}^{1/2}(T)\} = 0$ for all $i = 1, \ldots, m$. □

As in the study of S_c^n, the study of the asymptotic mean and variance of T_c^n requires the partitioning of index sets via $A(m,i,m',i')$. However, here the "overlapping" sets are much simpler since, for each n, m is fixed. Thus we need only consider $A(m,i,m,i')$:

$$A(m,i,m,i') = \{i, i-m, \ldots, i-km\} \cap \{i', i'-m, \ldots, i'-km\}$$

for $km + 1 \le i, i' \le n$. If $A(m,i,m,i') = \emptyset$ then W_i^m and $W_{i'}^m$ are independent, so we can neglect these terms when studying the variance of T_c^n. We partition $\{(i,i') : A(m,i,m,i') \ne \emptyset\}$:

$$\{(i,i') : A(m,i,m,i') \ne \emptyset\} = \left[\bigcup_{l=0}^{k-1} (B_1(l) \cup B_2(l))\right] \cup \left[\bigcup_{i=km+1}^{n} \{(i,i)\}\right]$$

where

$$\begin{aligned} B_1(l) = \{(i,i') : A(m,i,m,i') &= \{i', i'-m, \ldots, i'-lm\} \\ &= \{i-(k-l)m, \ldots, i-km\}\} \end{aligned}$$

and

$$B_2(l) = \{(i, i') : A(m, i, m, i') = \{i, i-m, \ldots, i-lm\} \\ = \{i' - (k-l)m, \ldots, i' - km\}\}.$$

The dependence of $B_1(l)$ and $B_2(l)$ upon n, k, and m is suppressed. Note that $(i, i') \in B_1(l)$ if and only if W_i^m and $W_{i'}^m$ involve the $l+1$ common observations: $X_{i'} = X_{i-(k-l)m}, \ldots, X_{i'-lm} = X_{i-km}$.

Recall the definition of $(V_1(l), V_2(l))$ in Section 2. We thus have

$$(V_1(l), V_2(l)) \sim \left(\frac{m}{n+1}\right)^k ((\Delta^{m,k}\epsilon)_i, (\Delta^{m,k}\epsilon)_{i'}), \qquad (i, i') \in B_1(l),$$

and

$$(V_2(l), V_1(l)) \sim \left(\frac{m}{n+1}\right)^k ((\Delta^{m,k}\epsilon)_i, (\Delta^{m,k}\epsilon)_{i'}), \qquad (i, i') \in B_2(l).$$

Let

$$\mu_T = \mu_T(f, \alpha) = \int_{t=k\alpha}^1 P\{U \geq \alpha^k(c - \Delta_f^{\alpha,k}(t))\}dt$$

and

$$\sigma_f^2(k, \alpha, t) = \mathrm{var}(I\{U \geq \alpha^k(c - \Delta_f^{\alpha,k}(t))\}).$$

For $l = 0$, $k-1$, let

$$\sigma_f^2(l, \alpha, t) = \mathrm{cov}(I\{V_1(l) \geq \alpha^k(c - \Delta_f^{\alpha,k}(t))\}, \\ I\{V_2(l) \geq \alpha^k(c - \Delta_f^{\alpha,k}(t - (k-l)\alpha))\})$$

and

$$b(l, \alpha, t) = I\{(2k - l)\alpha < t < 1\}.$$

Let

$$\sigma_T^2 = \sigma_T^2(f, \alpha) = 2\sum_{l=0}^{k-1} \int_{t=k\alpha}^1 \sigma_f^2(l, \alpha, t) b(l, \alpha, t)dt + \int_{t=k\alpha}^1 \sigma_f^2(k, \alpha, t)dt.$$

For the remainder of the section, assume that $\sigma_T^2(f, \alpha_0)$ is positive, and that there exists a neighborhood of α_0 such that $\sigma_f^2(l, \alpha, t)$, $l = 0, \ldots, k$, and $P\{U \geq \alpha^k(c - \Delta_f^{\alpha,k}(t))\}$ are continuous in (α, t) for all t and all α in that neighborhood. For instance, it suffices to assume that f has k continuous derivatives and that the distributions of $(V_1(l), V_2(l))$ and U are continuous. This will be true if the distribution of ϵ_1 is continuous.

THEOREM A.5. *Suppose that $f_n = f$ is fixed. Then*

$$\frac{T_c^n - \mathrm{E}(T_c^n)}{\mathrm{var}^{1/2}(T_c^n)} \Longrightarrow N(0, 1), \\ n^{-1}\,\mathrm{var}(T_c^n) \to \sigma_T^2(f, \alpha_0),$$

and

$$\mathrm{E}(T_c^n) = n\mu_T(f, \alpha_0) + o(n).$$

If $f = f^*$ and U has a continuous density then

$$\mathrm{E}(T_c^n) = n\mu_T(f, \alpha_0) + O(1).$$

PROOF. We first prove the second statement.

$$\begin{aligned}\operatorname{var}(T_c^n) &= \sum_{i,i'=km+1}^{n} \operatorname{cov}(W_i^m, W_{i'}^m) = \sum_{(i,i'):A(m,i,m,i')\neq\emptyset} \operatorname{cov}(W_i^m, W_{i'}^m) \\ &= \sum_{l=0}^{k-1} \sum_{(i,i')\in B_1(l)\cup B_2(l)} \operatorname{cov}(W_i^m, W_{i'}^m) + \sum_{i=km+1}^{n} \operatorname{var}(W_i^m).\end{aligned}$$

For $l = 0, \ldots, k-1$,

$$\begin{aligned}&\sum_{(i,i')\in B_1(l)} \operatorname{cov}(W_i^m, W_{i'}^m) \\ &= \sum_{(i,i')\in B_2(l)} \operatorname{cov}(W_i^m, W_{i'}^m) \\ &= \sum_i \operatorname{cov}\Bigg(I\left\{V_1(l) \geq \left(\frac{m}{n+1}\right)^k \left(c - \Delta_f^{m/(n+1),k}\left(\frac{i}{n+1}\right)\right)\right\}, \\ &\qquad I\left\{V_2(l) \geq \left(\frac{m}{n+1}\right)^k \left(c - \Delta_f^{m/(n+1),k}\left(\frac{i-(k-l)m}{n+1}\right)\right)\right\}\Bigg) \\ &= \sum_i \sigma_f^2\left(l, \frac{m}{n+1}, \frac{i}{n+1}\right)\end{aligned}$$

where the summation is over all i with $km+1 \leq i$, $i-(k-l)m \leq n$, that is with $(2k-l)m < i \leq n$, or equivalently, over all i with $b(l, m/(n+1), i/(n+1)) = 1$. Also,

$$\operatorname{var}(W_i^m) = \sigma_f^2\left(k, \frac{m}{n+1}, \frac{i}{n+1}\right).$$

The second statement of the theorem follows immediately, by calculating Riemann integrals.

The asymptotic normality of T follows from Theorem A.4.

The third statement, concerning the expected value of T, follows from

$$\mathrm{E}(T_c^n) = \sum_{i=mk+1}^{n} P\left\{U \geq \left(\frac{m}{n+1}\right)^k \left(c - \Delta_f^{m/(n+1),k}\left(\frac{i}{n+1}\right)\right)\right\} \sim n\mu_T.$$

The fourth statement follows easily. □

THEOREM A.6. *Let $f_n(x) = f(x) + \delta_n n^{-1/2} g(x)$, where f and g have k continuous derivatives. Suppose that U has a continuous density f_U and that $\delta_n \to \delta < \infty$. Then*

$$\frac{T_c^n - \mathrm{E}_{f_n}(T_c^n)}{n^{1/2}\sigma_T(f, \alpha_0)} \Longrightarrow N(0,1)$$

and

$$n^{-1/2}(\mathrm{E}_{f_n}(T_c^n) - \mathrm{E}_f(T_c^n)) \to \delta\alpha_0^k \int_{t=\alpha_0 k}^{1} \Delta_g^{\alpha_0,k}(t) f_U(\alpha_0^k(c - \Delta_f^{\alpha_0,k}(t)))dt.$$

The proof is similar to that of Theorem A.3 and is omitted.

REFERENCES

Gasser, T., Müller, H. G., Köhler, W., Molinari, L. and Prader, A. (1984). Nonparametric regression analysis of growth curves, *Ann. Statist.*, **12**, 210–229.

Hájek, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives, *Ann. Math. Statist.*, **39**, 325–346.

Mann, Henry B. (1945). Nonparametric tests against regression, *Econometrika*, **13**, 245–259.

Schluter, D. (1988). Estimating the form of natural selection on a quantitative trait, *Evolution*, **45**, 849–861.

Sen, P. K. (1965). Some non-parametric tests for m-dependent time series, *J. Amer. Statist. Assoc.*, **60**, 134–147.

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