

A TEST FOR DETECTING CHANGES IN MEAN*

WEI BIAO WU[†]

Abstract. In the classical time series analysis, a process is often modeled as three additive components: long-time trend, seasonal effect and background noise. Then the trend superimposed with the seasonal effect constitute the mean part of the process. The issue of mean stationarity, which is generically called *change-point* problem, is usually the first step for further statistical inference. In this paper we develop testing theory for the existence of a long-time trend. Applications to the global temperature data and the Darwin sea level pressure data are discussed. Our results extend and generalize previous ones by allowing dependence and general patterns of trends.

1. Introduction. Many time series models assume the form

$$(1) \quad X_k = \psi_k + \theta_k + Z_k, \quad k = 1, \dots, n,$$

where ψ_k is the long-time trend, θ_k is the seasonal component and Z_k is the random noise. For example, in the study of global warming, the temperature data consists of the long-time warming trend, seasonal variations within years and random fluctuations (cf. Figure 5). A fundamental issue associated with (1) is to test the existence of a long-time trend ψ based on the observations X_1, \dots, X_n . We refer this problem as *the change problem* instead of the widely accepted term *change-point problem* since an abrupt change-point may not be well-defined in examples like global warming where gradual changes are expected. The change-point analysis has been one of the central issues of statistical inference for several decades.

As a special case of (1), the model $X_k = \psi_k + Z_k$ in which $\{Z_k\}$ are independent and $\{\psi_k\}$ only take two possible values has been widely studied; see Bhattacharya (1994) and Siegmund (1986) for reviews. Here we assume $\psi_k = f(k/n)$, where f is a piecewise continuous function with finitely many segments. Let $\{Z_k\}$ be a stationary process with mean zero and absolutely summable covariances; namely $\{Z_k\}$ is short-range dependent. We consider *a posteriori* testing in that $\mathbf{X} = (X_1, \dots, X_n)$ is already available before the analysis. The hypothesis testing problem can now be formulated as

$$(2) \quad H_0 : f = \text{constant} \quad \text{against} \quad H_A : f \text{ is not constant.}$$

This setup generalizes the one adopted in Wu, Woodroffe and Mentz (2001) where $\theta \equiv 0$ and f is assumed to be non-decreasing. An isotonic regression based test is proposed in the latter paper and this test is shown to be more powerful than some existing ones (Brillinger 1989). Our extensions allow more complicated trend patterns without imposing any parametric form such as linearity and quadratic.

*Partially supported by the U.S. Army Research Office.

[†]Department of Statistics, University of Chicago, Chicago, IL 60637.

The remainder of the paper is organized as follows. Some notation and model assumptions are introduced in Section 2. In Section 3, isotonic and Kolmogorov-Smirnov tests are proposed and a power study is performed. In Section 5 our approach is applied to the global warming data and the Darwin sea level pressure data. Proofs are given in Section 6.

2. Preliminaries. For a function H on $[0, 1]$, define its *greatest convex minorant* (GCM)

$$\underline{H} = \sup\{G : G \text{ is convex and } G(x) \leq H(x), x \in [0, 1]\}.$$

The *least concave majorant* (LCM) \overline{H} can be similarly defined. Clearly, $-\underline{H} = \overline{G}$, where $G = -H$. Denote by \underline{h} and \overline{h} the left-hand derivatives of \underline{H} and \overline{H} respectively. Let

$$(3) \quad \Omega(H; n) = \sup \left\{ \sum_{k=1}^n |H(t_k) - H(t_{k-1})| : 0 = t_0 < t_1 < \dots < t_{n-1} < t_n = 1 \right\}.$$

Then $\Omega(H; \infty)$ is the total variation of H on $[0, 1]$. Let $\gamma(k) = \mathbb{E}(Z_0 Z_k)$ and write

$$(4) \quad \sigma^2 = \sum_{k=-\infty}^{\infty} \gamma(k).$$

For $\sigma > 0$, we define the piecewise linear function \mathcal{T}_{nt} , $t \in [0, 1]$ such that $\mathcal{T}_{nt} = T_k / (\sigma\sqrt{n})$ at $t = k/n$, where $T_k = Z_1 + \dots + Z_k$. Suppose that the invariance principle

$$(5) \quad \{\mathcal{T}_{nt}, 0 \leq t \leq 1\} \implies \{W(t), 0 \leq t \leq 1\}$$

holds, where $W(\cdot)$ is a standard Brownian motion. Assumption (5) is crucial in the analysis of the asymptotic behavior of our test statistics.

Consider at the outset the special model $X_k = \psi_k + Z_k$, where $\psi \in \Xi = \{\psi = (\psi_1, \dots, \psi_n) \in \mathbb{R}^n : \psi_1 \leq \psi_2 \leq \dots \leq \psi_n\}$. Let $X_{1,r} = X_1 + r\sqrt{n}$, $X_{n,r} = X_n - r\sqrt{n}$ and $X_{i,r} = X_i$ for $2 \leq i \leq n-1$. Wu, Woodroffe and Mentz (2001) proposed the test statistic

$$(6) \quad \underline{\Lambda}(n, r) = \frac{1}{\sigma_n^2} \sum_{k=1}^n (\underline{\psi}_{k,r} - \bar{X}_n)^2,$$

where σ_n^2 is a consistent estimator for σ^2 , $\bar{X}_n = \sum_{k=1}^n X_k/n$ and

$$(7) \quad \underline{\psi}_{k,r} = \max_{i \leq k} \min_{j \geq k} \frac{X_{i,r} + \dots + X_{j,r}}{j - i + 1}.$$

If Z_k were iid $N(0, \sigma^2)$ random variables, then $\underline{\psi}_{k,r}$ maximize the penalized log-likelihood function $\ell_{n,r}(\psi|\mathbf{X}) = \ell_n(\psi|\mathbf{X}) - r\sqrt{n}(\psi_n - \psi_1)/\sigma^2$, where

$\ell(\psi|\mathbf{X}) = -(2\sigma^2)^{-1} \sum_{i=1}^n (X_i - \psi_i)^2 + C$. Penalization is introduced to overcome the spiking problem where ψ_1 and ψ_n are estimated with bias. Unfortunately, for the general case where Z_k are dependent and may have distributions other than normal, the likelihood function is complicated and the maximum likelihood estimator may not have a close form. More seriously, a numeric solution can also be very difficult to be obtained since it involves the computationally intensive high dimensional maximization problem. In such cases, under the invariance principle (5), the asymptotic distribution of the easily computable test statistic (6) has a close form (cf Theorem 1), and has a very good power. In Section 3.4, the powers of isotonic and UMP tests are compared. Thus (6) is preferable in our setting.

2.1. A Geometric interpretation. Let $F(x) = \int_0^x f(t)dt$ and $G(x) = F(x) - xF(1)$. Then G is convex under the condition $\psi \in \Xi$. The null hypothesis H_0 corresponds to $G \equiv 0$; and the departure of alternative hypothesis to the null can be geometrically interpreted as the distance between G and 0. This distance can be naturally measured by $d(G, 0) = [\int_0^1 g^2(t)dt]^{1/2}$, where g is the left-handed derivative of G .

Set $H_{n,r}(t) = \sqrt{n}\{G_{n,r}(t) - \bar{X}t\}/\sigma_n$, where $G_{n,r}(t)$ is a continuous, piecewise linear function such that $G_{n,r}(k/n) = \sum_{i=1}^k X_{i,r}/n$ for $k = 1, \dots, n$. Then $d^2(\underline{H}_{n,r}, 0) = \int_0^1 \underline{h}_{n,r}^2(t)dt$, where $\underline{h}_{n,r}$ is the left-hand derivative of $\underline{H}_{n,r}$, the LCM of $H_{n,r}$. Interestingly enough, $d^2(\underline{H}_{n,r}, 0)$ is exactly the log-likelihood ratio $\ell_{n,r}(\psi|\mathbf{X}) - \ell_{n,r}(\bar{X}_n|\mathbf{X}) = \underline{\Lambda}(n, r)$ (cf. Equation (16) in Wu, Woodroffe and Mentz, 2001).

3. Testing for a long-time trend. In this section we assume throughout that $X_k = \psi_k + Z_k$, namely $\theta_k \equiv 0$. The constancy of a function f can be characterized by $\bar{F} = \underline{F}$, or equivalently $\bar{G} = \underline{G} = 0$, where $G(x) = F(x) - xF(1)$. The geometric consideration in Section 2.1 motivates that the characterization of constancy can be utilized to test for a change of general pattern. To this end, it is natural to consider the distance

$$(8) \quad d(G, 0) = \left\{ \int_0^1 [\bar{g}^2(t) + \underline{g}^2(t)]dt \right\}^{1/2},$$

where \bar{g} and \underline{g} are the left-handed derivatives of \bar{G} and \underline{G} respectively. As shown in the proof of Proposition 3, the distance d has the property that $d(G, 0) \leq d(H, 0)$ if $H(0) = H(1) = 0$ and $0 \leq G(x) \leq H(x)$ holds for all $0 \leq x \leq 1$. Namely d preserves the order. Another characterization of constancy is given by $d'(G, 0) = \sup_{t \leq 1} |G(t)|$, which entails the Kolmogorov-Smirnov test (see Section 3.3).

Based on the distance (8), we propose the test statistic $d^2(\underline{H}_{n,r}, 0) + d^2(\bar{H}_{n,-r}, 0)$, or equivalently

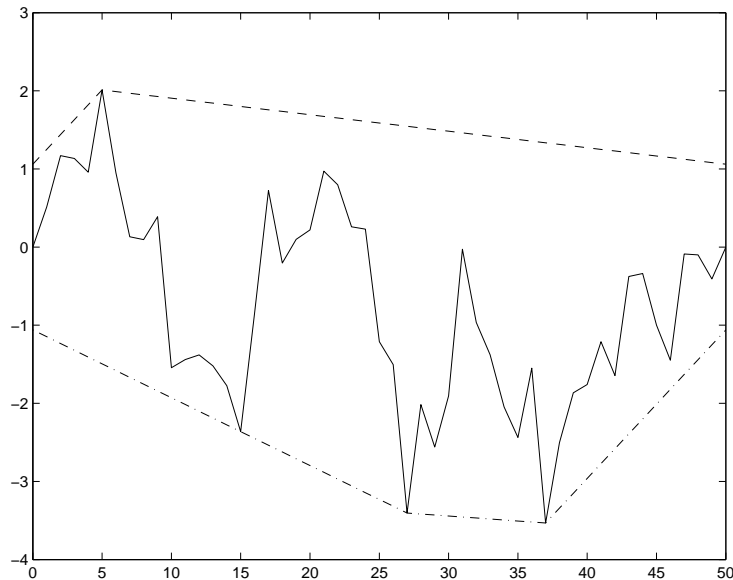


FIG. 1. The LCM \overline{Q}_n (---) and the GCM \underline{Q}_n (-·-) form a convex hull that envelopes Q_n .

$$\begin{aligned}
 (9) \quad \Lambda(n, r) &= \frac{1}{\sigma_n^2} \sum_{k=1}^n (\psi_{k,r} - \bar{X}_n)^2 + \frac{1}{\sigma_n^2} \sum_{k=1}^n (\bar{\psi}_{k,r} - \bar{X}_n)^2 \\
 &=: \underline{\Lambda}(n, r) + \overline{\Lambda}(n, r)
 \end{aligned}$$

where

$$(10) \quad \bar{\psi}_{k,r} = \min_{i \leq k} \max_{j \geq k} \frac{X_{i,-r} + \dots + X_{j,-r}}{j - i + 1}.$$

Let $Q_n(t)$ be piecewise linear function such that $Q_n(i/n) = (T_i - i\bar{X}_n)/(\sigma\sqrt{n})$ at $i = 0, \dots, n$ (Figure 1). Let $c = r/\sigma > 0$ and \overline{Q}_n be the LCM of the $n + 1$ points $(0, c), (i/n, Q_n(i/n)), 1 \leq i \leq n - 1$ and $(1, c)$; let \underline{Q}_n be the GCM of $n + 1$ points $(0, -c), (i/n, Q_n(i/n)), 1 \leq i \leq n - 1$ and $(1, -c)$. Then \overline{Q}_n and \underline{Q}_n form a convex hull that envelopes Q_n . This convex hull is expected to be thin if the trend is constant.

Consider the local alternative

$$(11) \quad f(t) = \mu + \sigma\phi(t)/\sqrt{n},$$

where ϕ is a right-continuous function on $[0, 1]$ for which $\int_0^1 \phi(t)dt = 0$. Recall that W is a standard Brownian motion. For $c > 0$ let $\underline{B}_c^\phi(t) = W(t) - tW(1) + \int_0^t \phi(s)ds + c\mathbf{1}_{(0,1)}(t)$, $\underline{b}_c^\phi(t)$ and $\overline{b}_{-c}^\phi(t)$ be the left-hand derivatives of $\underline{B}_c^\phi(t)$ and $\overline{B}_{-c}^\phi(t)$ respectively.

TABLE 1
Critical values for $\lambda_{.15}(\alpha)$.

n	$\lambda_{.15}(.05)$	$\lambda_{.15}(.01)$
10	5.95	8.94
20	7.02	10.28
40	7.73	11.05
80	8.51	11.95
160	8.90	12.17
320	9.32	12.79
640	9.68	13.19
1000	9.76	13.46
2000	9.95	13.60
4000	10.08	13.69

THEOREM 1. Assume (11) and let $\hat{\sigma}_n^2$ be a consistent estimator of σ^2 . Then (5) implies

$$(12) \quad \Lambda(n, r) \implies \int_0^1 [\underline{b}_c^\phi(t)]^2 + [\bar{b}_{-c}^\phi(t)]^2 dt.$$

Proof. See Theorem 1 in Wu, Woodroffe and Mentz (2001) for a proof of the convergence of $\underline{\Lambda}(n, r)$ in (9) to $\int_0^1 [\underline{b}_c^\phi(t)]^2$ in (12). It is easily seen that the arguments therein lead to (12). To see this heuristically, observe that $\underline{h}_{n,r}$ and $\bar{h}_{n,-r}$ are left-hand side derivatives of \underline{Q}_n and \bar{Q}_n respectively. Therefore by Robertson, Wright and Dykstra (p. 7, 1988),

$$\Lambda(n, r) = \int_0^1 \underline{h}_{n,r}^2(t) dt + \int_0^1 \bar{h}_{n,-r}^2(t) dt,$$

which intuitively converges to the right side of (12) in view of (5) and (11). \blacksquare

3.1. Critical values. The limiting distribution in (12) does not seem to have a close form even under the null hypothesis $f \equiv \text{constant}$. Table 1 gives the simulated critical values $\lambda_c(\alpha)$ for $c = .15$ and two levels $\alpha = .05$ and $.01$. The simulation is based on 50,000 Brownian motions for each $n = 10, 20, 40, 80, 160, 320, 640, 1000, 2000$ and 4000.

3.2. Monotonic trends. Proposition 1 implies the coherence property of the test statistics Λ_n in that if the alternative is a strictly increasing function, then $\int_0^1 [\bar{b}_{-c}^\phi(t)]^2 dt$, which corresponds to the LCM part of $\Lambda(n, r)$, vanishes asymptotically.

PROPOSITION 1. Let $\psi_k = \psi + \delta_n \phi(k/n)\sigma/\sqrt{n}$, where ϕ is a nondecreasing function such that $\phi(0) < 0$, $\phi(1) > 0$ and $\int_0^1 \phi(t)dt = 0$. Then as $\delta_n \rightarrow \infty$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \int_0^1 [\bar{b}_c^{\phi \delta_n}(t)]^2 dt = 0 \right\} = 1.$$

Proof of Proposition 1. It suffices to show that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{0 \leq t \leq 1} \left[W(t) - tW(1) + \delta_n \int_0^t \phi(u)du \right] \leq c \right\} = 1$$

since $\bar{B}_c^\phi(t) \equiv c$ under the event inside the probability measure. Observing that $\phi(0) < 0$ and $\phi(1) > 0$, there exists $C > 0$ such that for any $x \in [1/\sqrt{\delta_n}, 1 - 1/\sqrt{\delta_n}]$, $\delta_n \int_0^t \phi(u) \leq -C\sqrt{\delta_n}$. Hence

$$\begin{aligned} & \sup_{0 \leq t \leq 1} \left[W(t) - tW(1) + \delta_n \int_0^t \phi(u)du \right] \\ & \leq \max \left\{ \sup_{0 \leq t \leq 1/\sqrt{\delta_n}} [W(t) - tW(1)], \right. \\ & \quad \sup_{1/\sqrt{\delta_n} \leq t \leq 1 - 1/\sqrt{\delta_n}} [W(t) - tW(1) - C\sqrt{\delta_n}], \\ & \quad \left. \sup_{1 - 1/\sqrt{\delta_n} \leq t \leq 1} [W(t) - tW(1)] \right\}. \end{aligned}$$

Then almost surely the second term diverges to $-\infty$ and the first and the third terms converges to 0 as $\delta_n \rightarrow \infty$. Thus the proof is completed. \blacksquare

3.3. Test of the Kolmogorov-Smirnov type. Another characterization of the constancy of f is $\max_{0 \leq t \leq 1} |G(t)| = 0$, which leads to the cumulative sum (CUSUM) test of the Kolmogorov-Smirnov type

$$(13) \quad K_n = \frac{1}{\sigma_n \sqrt{n}} \max_{k \leq n} \left| \sum_{i=1}^k X_i - \frac{k}{n} \sum_{i=1}^n X_i \right|.$$

If σ_n is a consistent estimator of σ , then K_n converges to $\sup_{0 \leq t \leq 1} |W(t) - tW(1)|$. The well-known blocking technique can be employed to estimated σ^2 (cf (17) and (18) in Theorem 2). The limiting distribution is well-known (Shorack and Wellner, 1986):

$$\mathbb{P} \left[\sup_{0 \leq t \leq 1} |W(t) - tW(1)| > x \right] = 2 \sum_{j=1}^{\infty} (-1)^{j+1} e^{-2j^2 x^2}.$$

In Section 3.4, the powers of the isotonic-based test (9) and the Kolmogorov-Smirnov test (13) are compared.

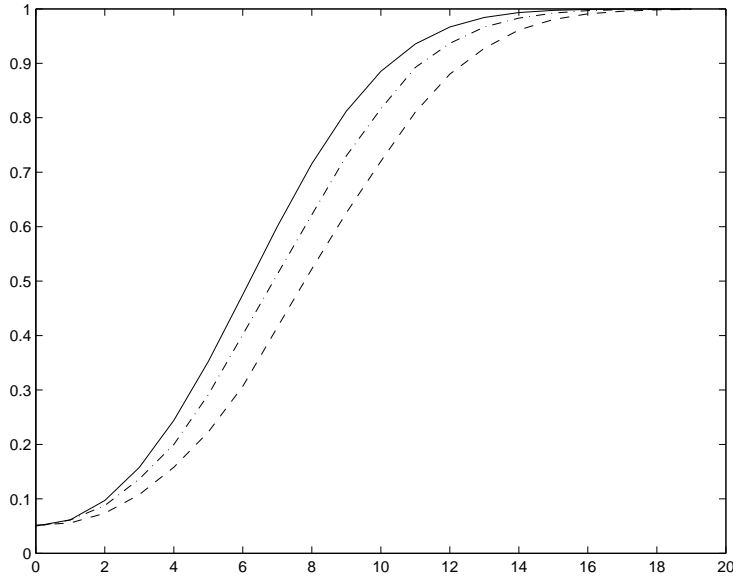


FIG. 2. Power curves for the uniformly most powerful test (solid line), the isotonic test (dashed line) and the Kolmogorov-Smirnov test (dashdot line) for the step function $f_1(u) = \mathbf{1}(u \leq 1/2)$.

3.4. Power study. To study the powers of the isotonic and the Kolmogorov-Smirnov tests, we consider the model

$$X_k = \frac{\delta}{\sqrt{n}} \phi\left(\frac{k}{n}\right) + Z_k, \quad 1 \leq k \leq n,$$

where Z_k are iid standard normal random variables, ϕ is a nonzero function defined on $[0, 1]$ and $\delta \geq 0$ measures the distance to the null hypothesis. The simulated power curves for three different functions (i) $\phi_1(u) = \mathbf{1}(u \leq 1/2)$ (ii) $\phi_2(u) = \mathbf{1}(u \leq 1/16) + \mathbf{1}(u \geq 15/16)$ and (iii) $\phi_3(u) = \sin(\pi u)$ are displayed in Figures 2, 3 and 4 respectively. In all figures $n = 160$ and $\delta = (l - 1)/20$, $l = 1, \dots, 20$. Since Z_i are iid normal and the parameter of interest δ is one-dimensional, the uniformly most powerful (UMP) test exists. The simulation study shows that the power of the isotonic test is not far away from that of the UMP test. The powers for all tests approach 1 if and only if $\delta = \delta_n \rightarrow \infty$. Figure 2 suggests that the Kolmogorov-Smirnov test has a slightly higher power than the isotonic one, while in Figure 3 the former has a severely less power than the latter.

To explain the above-mentioned phenomena, observe that for large δ , the major term in the Kolmogorov-Smirnov test is $\delta \sup_{0 \leq u \leq 1} |G(u)|$ while in the isotonic test the term $\delta^2 \int_0^1 [\bar{g}^2(u) + \underline{g}^2(u)] du$ has a dominated contribution, where $G(x) = \Phi(x) - x\Phi(1)$, $\Phi(x) = \int_0^x \phi(t) dt$. Proposition 3

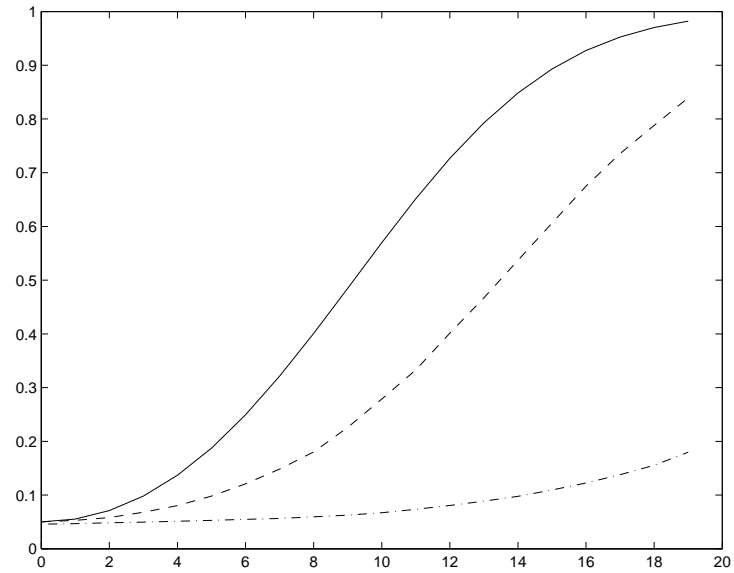


FIG. 3. Power curves for the uniformly most powerful test (solid line), the isotonic test (dashed line) and the Kolmogorov-Smirnov test (dashdot line) for the step function $f_2(u) = \mathbf{1}(u \leq 1/16) + \mathbf{1}(u \geq 15/16)$.

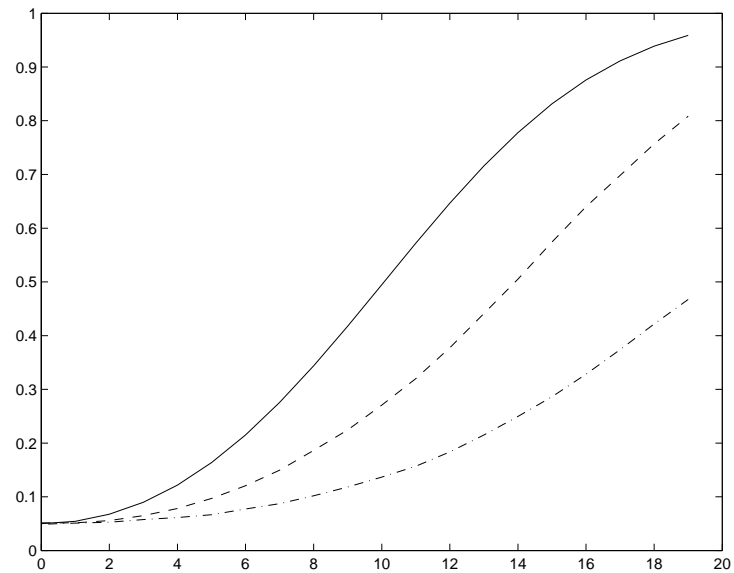


FIG. 4. Power curves for the uniformly most powerful test (solid line), the isotonic test (dashed line) and the Kolmogorov-Smirnov test (dashdot line) for the sine function $f_3(u) = \sin(\pi u)$.

asserts that in any case the power of the isotonic test will not be significantly inferior to that of the Kolmogorov-Smirnov test. On the other hand, since inequality (14) cannot be reversed even if a constant multiple is allowed, the Kolmogorov-Smirnov test may have very low power for certain alternatives. For example, for $\lambda \in (0, 1)$ let $G(x) = \int_0^x g(u)du$, where

$$g(u) = C(\lambda) \left[-\frac{1}{\lambda} \mathbf{1}(u \leq \lambda) + \frac{1}{1-\lambda} \mathbf{1}(u > \lambda) \right] \text{ and } C(\lambda) = [\lambda(1-\lambda)]^{\frac{1}{3}}.$$

Then it is easily seen that $\max_{0 \leq u \leq 1} |G(u)| = C(\lambda) \downarrow 0$ and $\int_0^1 [\bar{g}^2(u) + \underline{g}^2(u)]du = 1/C(\lambda) \uparrow \infty$ as $\lambda \downarrow 0$. To summarize, the isotonic test has a uniformly reasonable power in all circumstances.

PROPOSITION 2. *If \bar{H} or \underline{H} is not identically 0, then as $\delta_n \rightarrow \infty$,*

$$\lim_{\delta_n \rightarrow \infty} \mathbb{P} \left\{ \int_0^1 [\bar{b}_c^{\rho_n \phi}(t)]^2 dt + \int_0^1 [\underline{b}_c^{\rho_n \phi}(t)]^2 dt > \lambda_c(\alpha) \right\} = 1,$$

where $\rho_n = \delta_n/\sigma$.

Proof. See Wu, Woodroffe and Mentz (2001). ■

PROPOSITION 3. *Let G be a function defined on $[0, 1]$ such that $G(0) = G(1) = 0$. Then*

$$(14) \quad 4 \sup_{0 \leq u \leq 1} G^2(u) \leq \int_0^1 [\bar{g}^2(u) + \underline{g}^2(u)]du,$$

where the equality holds if and only if $|G(1/2)| = \sup_{0 \leq u \leq 1} |G(u)|$ and the graph of G is contained in the triangle by points $(0, 0)$, $(1, 0)$ and $(1/2, G(1/2))$.

4. Estimating σ^2 . In Wu, Woodroffe and Mentz (2001), the model $X_k = \psi_k + Z_k$ is considered in which the trend is assumed to be nondecreasing. Then lag-windows type of estimators are constructed based on the the estimated residuals $\widehat{Z}_k = X_k - \widehat{\psi}_k$, where $\widehat{\psi}$ is the isotonic regression estimator. Here monotonicity assumption is not imposed and we shall estimate σ^2 in the presence of θ_k . Recall that $T_k = \sum_{i=1}^k Z_i$ and $f(k/n) = \psi_k$. Let $S_k = \sum_{i=1}^k X_i$, $\Theta_k = \sum_{i=1}^k \theta_i$ and $\Psi_k = \sum_{i=1}^k \psi_i$.

THEOREM 2. *Let $m \rightarrow \infty$, $m = \mathcal{O}(n^{1/3})$, $b = \lfloor n/m \rfloor$. Assume that*

$$(15) \quad \sup_{k \geq 0} |\Theta_{k+m} - \Theta_k| = o(\sqrt{m})$$

and

$$(16) \quad \Omega(f; b) = o(\sqrt{b}).$$

Then

$$(17) \quad \sigma_{\text{Block}}^2(Z) = \frac{1}{2n} \sum_{k=2}^b [T_{km} - T_{(k-1)m}]^2 \rightarrow_{\mathbb{P}} \sigma^2,$$

implies

$$(18) \quad \sigma_{\text{Block}}^2(X) = \frac{1}{2n} \sum_{k=2}^b [S_{km} - S_{(k-1)m}]^2 \xrightarrow{\mathbb{P}} \sigma^2.$$

We say that a function f is Hölder continuous with index $h > 0$ if there exist $L > 0$ such that for all $0 \leq x, y \leq 1$, $|f(x) - f(y)| \leq L|x - y|^h$. Clearly (16) holds for piecewise Hölder continuous functions with index $h > 1/2$. In the case that Z_k are iid, Hall, Kay and Titterington (1990) considered the difference-based estimation of $\sigma^2 = \mathbb{E}(Z_1^2)$ from the model $Y_j = f(j/n) + Z_j$, $j = 1, \dots, n$ by assuming f is Hölder continuous with $h > 1/2$. Our concise estimator $\sigma_{\text{Block}}^2(X)$ uses first order differences when Z_k are allowed to be dependent. To reduce bias, estimators based on higher order differences can be similarly constructed as in Hall *et al.*

REMARK 1. For the commonly used seasonal model, $\theta_k = \sum_{i=1}^I A_i \cos(k\omega_i + \alpha_i)$, where $0 < \omega_i < 2\pi$ are frequencies and A_i are amplitudes, it is easily seen that $\sup_{k \geq 0} |\Theta_{k+m} - \Theta_k| = \mathcal{O}(1)$ and hence (15) holds.

5. A separation principle. In this section we shall consider the testing problem proposed in the Introduction, namely we test for “ $f = \text{constant}$ ” in the model $X_k = \psi_k + \theta_k + Z_k$. For the seasonal component, let $\theta_k = \sum_{i=1}^I A_i \cos(k\omega_i + \alpha_i)$, where $0 < \omega_i < 2\pi$ are frequencies and $A_i > 0$ are amplitudes. Let $Y_k = \psi_k + Z_k$ be the process without seasonal components and analogously $V_k = \theta_k + Z_k$ be the process without long-time trend. Let $\kappa_n(\omega) = \sum_{k=1}^n \exp(\omega k \iota)$; $S_{n,X}(\omega) = \sum_{k=1}^n X_k \exp(\omega k \iota)$ and $S_{n,V}(\omega) = \sum_{k=1}^n V_k \exp(\omega k \iota)$, where ι is the imaginary unit. Then for a fixed $\omega \in (0, 2\pi)$, $\sup_{n \geq 0} |\kappa_n(\omega)| \leq 2/|1 - \exp(\omega \iota)| = \mathcal{O}(1)$. So if $\Omega(f, n) = o(\sqrt{n})$, then

$$\begin{aligned} \frac{|S_{n,X}(\omega) - S_{n,V}(\omega)|}{\sqrt{n}} &= \frac{1}{\sqrt{n}} \left| \sum_{k=1}^n f(k/n) \exp[\omega k \iota] \right| \\ &= \frac{1}{\sqrt{n}} \left| \sum_{k=2}^n \{f(k/n) - f((k-1)/n)\} \kappa_k(\omega) \right| + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{\mathcal{O}[\Omega(f; n)]}{\sqrt{n}} \\ &= o(1), \end{aligned}$$

which suggests an interesting feature of the spectral analysis: the periodograms of X_k and V_k have asymptotically negligible differences. Clearly, $S_{n,V}(\omega)$ has a magnitude of order n if ω is one of the frequencies ω_i . The identification of ω_i will require the asymptotic distribution of periodograms (see, for example, Chapter 10 in Brockwell and Davis, 1991). Wu (2002)

obtain central limit theorems for the Fourier transform $S_{n,Z}(\omega)$ under mild conditions on Z_k .

On the other hand, since $\sup_{k \geq 0} |\Theta_k| = \mathcal{O}(1)$, isotonic regressions based on X_k and Y_k produce asymptotically equivalent estimators for ψ_k . This equivalence in view of the formula (7) is implied by the fact that the invariance principle (5) still holds if we regard $Z'_k = \theta_k + Z_k$ as the new background noises. Recall $Y_k = \psi_k + Z_k$. Similarly as $X_{k,r}$, let $Y_{1,r} = Y_1 + r\sqrt{n}$, $Y_{n,r} = Y_n - r\sqrt{n}$ and $Y_{i,r} = Y_i$ for $2 \leq i \leq n-1$, and define

$$(19) \quad \underline{\nu}_{k,r} = \max_{i \leq k} \min_{j \geq k} \frac{Y_{i,r} + \dots + Y_{j,r}}{j - i + 1}, \quad \bar{\nu}_{k,r} = \min_{i \leq k} \max_{j \geq k} \frac{Y_{i,-r} + \dots + Y_{j,-r}}{j - i + 1}.$$

THEOREM 3. *Under the condition (15), we have*

$$\sum_{k=1}^n (\underline{\psi}_{k,r} - \underline{\nu}_{k,r})^2 + \sum_{k=1}^n (\bar{\psi}_{k,r} - \bar{\nu}_{k,r})^2 = o_{\mathbb{P}}(1).$$

To summarize, the spectral analysis and the isotonic regression filter ψ and θ respectively. Programs are available at <http://www.stat.uchicago.edu/faculty/wu.html>.

5.1. Global warming data. The global temperature data consists of monthly temperature anomalies from 1856 to 2000 (cf. Figure 5, <http://cdiac.esd.ornl.gov/trends/temp/jonescru/jones.html>). Now we shall apply the separation principle to the global temperature data. Wu, Woodroffe and Mentz (2001) analyzed the yearly averaged data by using the penalized isotonic regression with $c = 0.15$ (cf. Figure 6) and showed that there exists a substantial increasing trend. The estimated variance is $\hat{\sigma}^2 = 0.0158$. As shown in Figure 6, the estimated trends based on the monthly data and the yearly data are sufficiently close. Noticing that by taking yearly average is tantamount to eliminating seasonal effects, this comparison suggests the robustness of isotonic regression against seasonal components.

On the other hand, the periodogram plot for this monthly temperature data in which the long-term trend is present indicates a cyclic component with frequency $\omega_1 = 2\pi/12$ (cf. Figure 7). This observation reflects the common sense that the period is 12 months.

It is generally believed that the average global surface temperature has increased $0.4 \sim 0.8$ °C since the late 19th century (cf the report by IPCC, 2001). The IPCC report also mentioned that there are two major periods of increment: 1910-1945 and 1976-present. Based on our isotonic regression, the estimated increment is $\underline{\psi}_{145,r} - \underline{\psi}_{1,r} = 0.72$ °C, where the penalty $c = 0.15$ is used, $\underline{\psi}_{145,r}$ and $\underline{\psi}_{1,r}$ are the estimated mean temperatures of the year 2000 and 1856. Interestingly enough, from Figure 6, the isotonic regression procedure indicates that two major periods of increment are

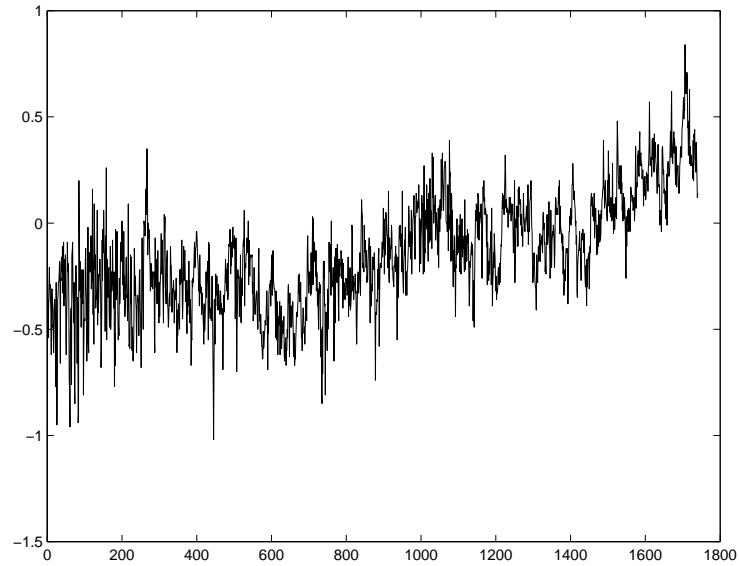


FIG. 5. *Global warming data: monthly temperature anomalies from 1856 to 2000.*

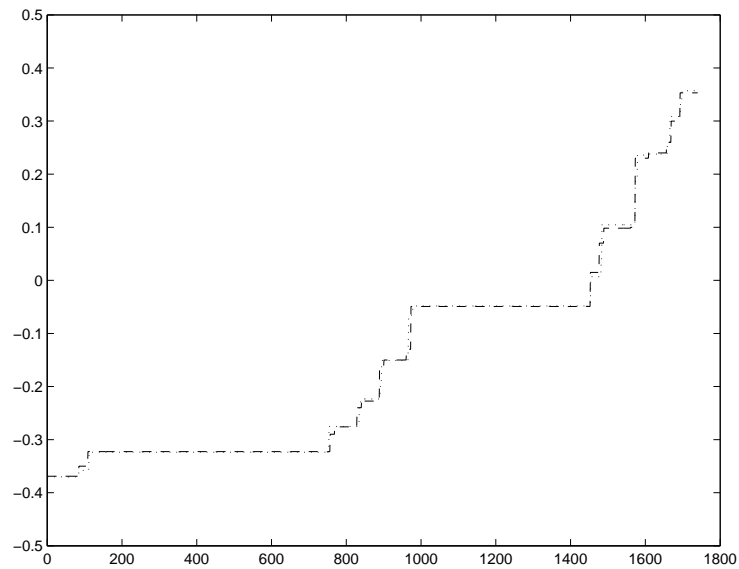


FIG. 6. *Isotonic regression estimators for monthly (dotted line) and yearly (dashed line) temperature data.*

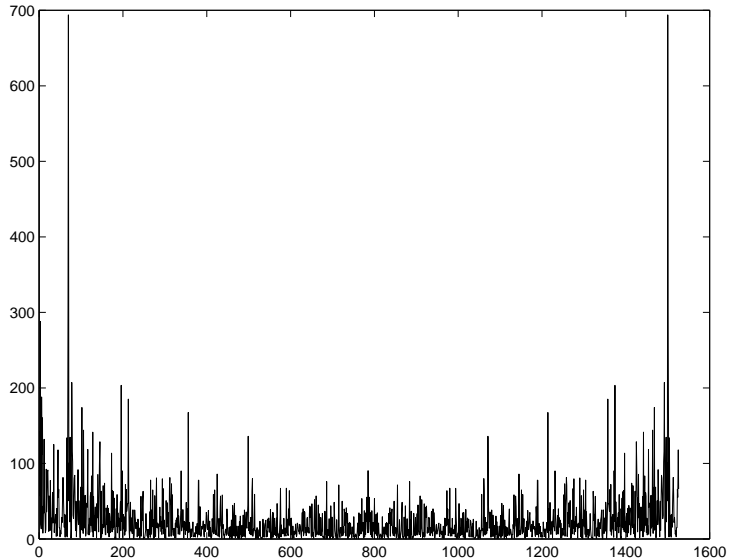


FIG. 7. Periodogram plot for the global monthly temperature data.

1920–1935 and 1976–present. Thus our procedure performs well and it appears more versatile than the usual method where the trend is modeled linearly.

5.2. The Darwin sea level pressure data. The sea level pressure data were collected at Darwin, Australia (13S, 131E) from year 1882 to 2001; see the website <http://www.cpc.ncep.noaa.gov/data/indices/> (by Climate Prediction Center, National Centers for Environmental Prediction, National Oceanic and Atmospheric Administration) for more detailed information. Yearly and monthly plots are displayed in Figures 8 and 9 respectively. The unit is millibar (MB) with 1000 MB subtracted from the original observations.

For the monthly data, the estimated $\sigma_{\text{Month}} = 2.4176$ and the isotonic test statistic is 7.1789 by choosing the penalty $c = .15$. For the yearly data, $\sigma_{\text{Year}} = 0.6372$ and the test statistic is 6.9013. Both test statistics are very close to each other, and they indicate that the sea level pressure has not undergone a significant change at least in the last century.

6. Proofs.

Proof of Theorem 2. Note that $S_k = T_k + \Theta_k + \Psi_k$. By condition (17) it suffices to establish

$$\mathbb{E} \left| \sum_{k=2}^b \{ [S_{km} - S_{(k-1)m}]^2 - [T_{km} - T_{(k-1)m}]^2 \} \right| = o(n).$$

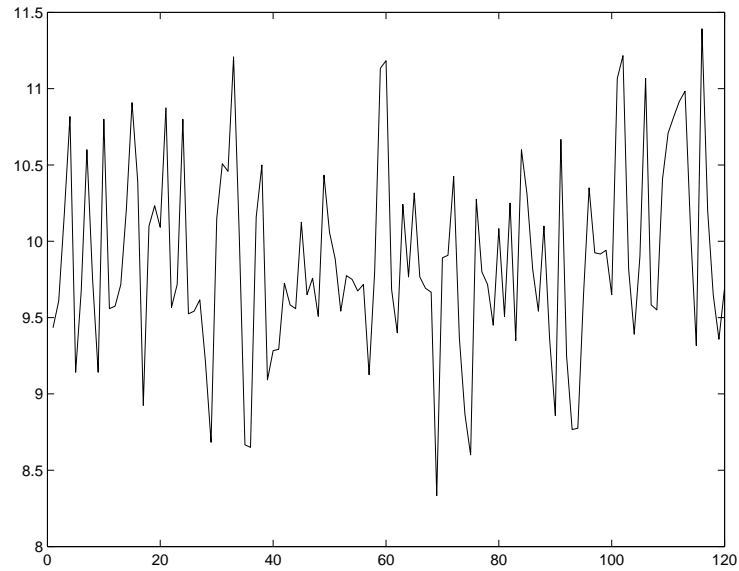


FIG. 8. Yearly sea level pressure data collected at Darwin, Australia (13S, 131E) from 1882 to 2001.

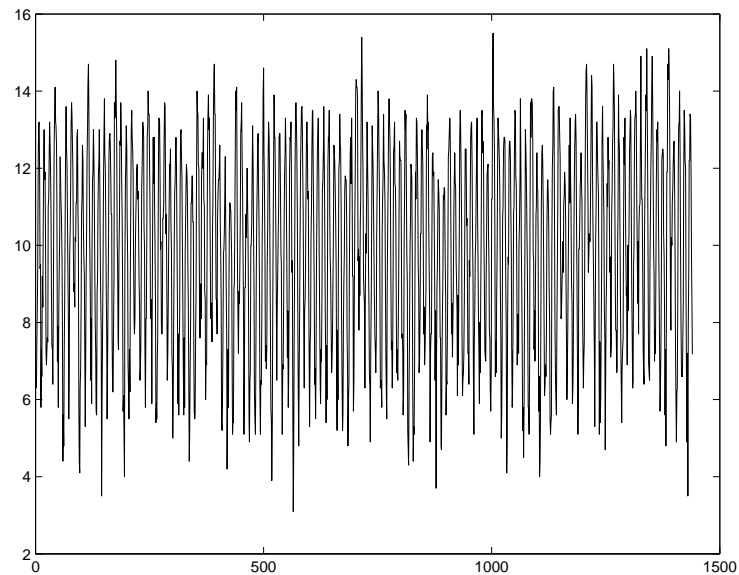


FIG. 9. Monthly sea level pressure data collected at Darwin, Australia (13S, 131E) from 1882 to 2001.

This relation clearly follows from

$$\sum_{k=2}^b \mathbb{E}\{[\Theta_{km} - \Theta_{(k-1)m}]^2 + |\Theta_{km} - \Theta_{(k-1)m}| |T_{km} - T_{(k-1)m}|\} = o(n)$$

and

$$\sum_{k=2}^b \mathbb{E}\{[\Psi_{km} - \Psi_{(k-1)m}]^2 + |\Psi_{km} - \Psi_{(k-1)m}| |T_{km} - T_{(k-1)m}|\} = o(n).$$

The former results easily from (15) and $\mathbb{E}|T_m| = \mathcal{O}(\sqrt{m})$. For the latter, let $C = \sup_{x \in [0,1]} |f(x)| < \infty$. Then by Cauchy's inequality,

$$\begin{aligned} \sum_{k=2}^b [\Psi_{km} - \Psi_{(k-1)m}]^2 &= \sum_{k=2}^b \left\{ \sum_{j=1-m}^0 [f((km+j)/n) - f((km-m+j)/n)] \right\}^2 \\ &\leq \sum_{k=2}^b m \sum_{j=1-m}^0 [f((km+j)/n) - f((km-m+j)/n)]^2 \\ &\leq Cm \sum_{j=1-m}^0 \sum_{k=2}^b |f((km+j)/n) - f((km-m+j)/n)| \\ &\leq Cm^2 \Omega(b) \\ &= o(n). \end{aligned}$$

Observe that $\mathbb{E}|T_{km} - T_{(k-1)m}| = \mathbb{E}|T_m| = \mathcal{O}(\sqrt{m})$ and $\sum_{k=2}^b |\Psi_{km} - \Psi_{(k-1)m}| = o(\sqrt{n})\sqrt{b}$, we have

$$\sum_{k=2}^b |\Psi_{km} - \Psi_{(k-1)m}| \mathbb{E}|T_{km} - T_{(k-1)m}| = \mathcal{O}(\sqrt{m})o(\sqrt{n})\sqrt{b} = o(n)$$

completes the proof. \blacksquare

Proof of Theorem 3. Recall $G_{n,r}(k/n) = \sum_{i=1}^n X_{i,r}/n$ and $H_{n,r}(t) = \sqrt{n}[G_{n,r}(t) - \bar{X}_n t]/\sigma$. Analogously, for Y_k let $P_{n,r}(t) = \sqrt{n}[R_{n,r}(t) - \bar{Y}_n t]/\sigma$, where $R_{n,r}(k/n) = \sum_{i=1}^n Y_{i,r}/n$. Let $\|F\| = \sup_{0 \leq t \leq 1} |F(t)|$. Then by Woodroffe and Sun (1999),

$$\begin{aligned} &\int_0^1 [\underline{\psi}_{k,r}(t) - \underline{\nu}_{k,r}(t)]^2 dt \\ &\leq \|\underline{P}_{n,r} - \underline{H}_{n,r}\| [\underline{\psi}_{k,r}(1-) - \underline{\psi}_{k,r}(0+) + \underline{\nu}_{k,r}(1-) - \underline{\nu}_{k,r}(0+)]. \end{aligned}$$

By Marshall's lemma,

$$\|\underline{P}_{n,r} - \underline{H}_{n,r}\| \leq \|P_{n,r} - H_{n,r}\| = \mathcal{O}\left(\sup_{1 \leq k \leq n} |\Theta_k|/\sqrt{n}\right) = o(1),$$

which entails the theorem by noticing that $\underline{\psi}_{k,r}(1-)$, $\underline{\psi}_{k,r}(0+)$, $\underline{\nu}_{k,r}(1-)$ and $\underline{\nu}_{k,r}(0+)$ are all stochastically bounded (cf. Lemma A4 in Wu, Woodroffe and Mentz (2001)). ■

Proof of Proposition 3. For two functions G_1 and G_2 defined on $[0, 1]$, denote by $G_1 \leq G_2$ if $G_1(x) \leq G_2(x)$ holds for all $x \in [0, 1]$. Let H be a concave function defined on $[0, 1]$ for which $H(0) = H(1) = 0$. Then for any triple $0 \leq x_1 < x_2 < x_3 \leq 1$, by concavity it is easily verified that

$$\begin{aligned} & \left[\frac{H(x_2) - H(x_1)}{x_2 - x_1} \right]^2 (x_2 - x_1) + \left[\frac{H(x_3) - H(x_2)}{x_3 - x_2} \right]^2 (x_3 - x_2) \\ & \geq \left[\frac{H(x_3) - H(x_1)}{x_3 - x_1} \right]^2 (x_3 - x_1). \end{aligned}$$

Let G_1 and G_2 be concave, continuous and piecewise linear (CCPL) functions with values being 0 at the endpoints $x = 0, 1$. Then $G_1 \leq G_2$ entails $d(G_1, 0) \leq d(G_2, 0)$. To see this, a chain of CCPL functions $G_1 = H_1 \leq H_2 \dots \leq H_q = G_2$ can be constructed such that H_i and H_{i+1} differ only on the interval $I = [x_1, x_3]$ (say), where $H_i(x)$ is a linear function, $H_i(x_1) = H_{i+1}(x_1)$, $H_i(x_3) = H_{i+1}(x_3)$, and H_{i+1} is linear on intervals $[x_1, x_2]$ and $[x_2, x_3]$ respectively. The above inequality on H implies that $d(H_i, 0) \leq d(H_{i+1}, 0)$. Hence $d(G_1, 0) \leq d(G_2, 0)$. Since concave and continuous functions can be approximated by CCPL functions, the general case that $G_1 \leq G_2$ implies $d(G_1, 0) \leq d(G_2, 0)$ follows by taking limits.

Let $\lambda = \operatorname{argmax}_u |G(u)|$. Without loss of generality assume $G(\lambda) > 0$. Then $\overline{G} \geq L$, where $L(0) = L(1) = 0$, $L(\lambda) = G(\lambda)$ and L is linear on intervals $[0, \lambda]$ and $[\lambda, 1]$. Using the monotonicity of $d(\cdot, 0)$, we have $d(\overline{G}, 0) \geq d(L, 0)$, or

$$\int_0^1 \overline{g}^2(u) du \geq \left[\frac{L(\lambda)}{\lambda} \right]^2 \lambda + \left[\frac{L(\lambda)}{1-\lambda} \right]^2 (1-\lambda) = \frac{L^2(\lambda)}{\lambda(1-\lambda)} \geq 4G^2(\lambda). \quad \blacksquare$$

Acknowledgment. The author thanks H. Pollack (Department of Geological Sciences, University of Michigan) for the IPCC (2001) report. The suggestions from the reviewer greatly improve the paper.

REFERENCES

- BHATTACHARYA P.K., Some aspects of change-point analysis. In *Change-Point Problems*, Ed. E. Carlstein, H. Müller and D. Siegmund, pp. 28–56. Hayward, CA: Inst. Math. Statist., 1994.
- BRILLINGER D.R. (1989). Consistent detection of a monotonic trend superposed on a stationary time series. *Biometrika*. **76**, 23–30.
- BROCKWELL P.J. AND DAVIS R.A. *Time series: theory and methods*. New York: Springer-Verlag, 1991.

- HALL P., KAY J.W., AND TITTERINGTON D.M.. Asymptotically Optimal Difference-Based Estimation of Variance in Nonparametric Regression. *Biometrika*, **77**, No. 3. (Sep., 1990), pp. 521–528.
- IPCC. *Climate change 2001: the scientific basis. Contribution of Working Group I to the third assessment report of the Intergovernmental Panel on Climate Change.* Edited by J.T. Houghton *et al.*, Cambridge University Press, 2001.
- RICE J, Bandwidth Choice for Nonparametric Regression. *Annals of Statistics*, **12**, 1215–1230, 1984.
- SHORACK G.R. AND WELLNER J.A. *Empirical processes with applications to statistics.* New York: Wiley, 1986.
- SIEGMUND D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14**, 361–404.
- WOODROOFE M. AND SUN J. (1999). Testing uniformity versus a monotone density. *Ann. Statist.* **27**, 338–360.
- WU W.B. (2002). Fourier transforms of stationary processes. *Technical Report #514, Department of Statistics, University of Chicago.*
- WU W.B., WOODROOFE M., AND MENTZ G., Isotonic regression: another look at the change point problem. *Biometrika*, **88**, 793–804, 2001.