Biometrika (2001), **88**, 3, *pp*. 793–804

Printed in Great Britain

Isotonic regression: Another look at the changepoint problem

By WEI BIAO WU, MICHAEL WOODROOFE and GRACIELA MENTZ
Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
wbwu@umich.edu michaelw@umich.edu gmentz@umich.edu

Summary

A test based on isotonic regression is developed for monotonic trends in short range dependent sequences and is applied to Argentina rainfall data and global warming data. This test provides another perspective for changepoint problems. The isotonic test is shown to be more powerful than some existing tests for trend.

Some key words: Changepoint problem; Isotonic regression; Penalised likelihood function; Short range dependence.

1. Introduction

It is often important to be able to test stationarity of a given time series. Here interest centres on stationarity of the mean, and the process of interest is assumed to be of the form

$$X_k = \mu_k + Z_k, \tag{1}$$

where the μ_k are the means and Z_k is a stationary process with mean 0 and finite covariances $\gamma(k) = \text{cov}(Z_i, Z_{i+k}) = E(Z_0 Z_k)$. For this scenario, there has been both classical work on testing for the existence of a trend and more recent work on tests for an abrupt change. The present work falls between these two approaches in developing a test for a change, or trend, that is monotonic but otherwise arbitrary. Thus, it is assumed throughout that $\mu_1 \leqslant \mu_2 \leqslant \ldots \leqslant \mu_n$, where n is the length of the series, and the hypothesis of interest is

$$H\colon\ \mu_k = \mu_1 \tag{2}$$

for all k, to be tested against the alternative that $\mu_k < \mu_{k+1}$ for some k. Darkhovsky (1994) and Brodsky & Darkhovsky (1993) show that general changepoint problems can be reduced to the problem of testing the mean stability of some new sequences. The type of testing we consider is off-line, in that the data $\mathscr{X} = \mathscr{X}_n = (X_1, \ldots, X_n)$ have already been obtained before the analysis. The trend $\{\mu_k\}$ is sometimes called the signal and the process $\{Z_k\}$ is the background noise or errors.

The formulation is motivated in part by Argentina rainfall data provided by Eng Cesar Lamelas, a meteorologist from the Agricultural Experimental Station Obispo Colombres, Tucumán. This dataset contains monthly rainfall records collected from 1884 to 1996; see Fig. 1 below. Lamelas believes that there was a change in the mean, caused by the construction of a dam in Tucumán from 1952 to 1962.

Early work on the changepoint problem mainly concerned the simple model in which the $\{Z_k\}$ are independent and the $\{\mu_k\}$ only take two possible values. Then the alternative hypothesis has the form

$$H_1:\ \mu_1 = \mu_2 = \ldots = \mu_\tau \neq \mu_{\tau+1} = \ldots = \mu_n,$$

for some unknown changepoint τ such that $1 \leqslant \tau \leqslant n$. The classical changepoint problem is to test for the existence of a changepoint and estimate its location if it exists. Bhattacharya (1994) and Siegmund (1986) have written review articles, and Shaban (1980) has complied an annotated bibliography. Extensions of the simple changepoint model either allow more general change patterns or relax the independence assumption of $\{Z_k\}$; dependence is inevitable in the study of time series. This complicates the testing procedures. Lombard & Hart (1994) consider the simple abrupt change model with dependent errors; see also Brodsky & Darkhovsky (1993, Ch. 3), where strong mixing conditions are imposed. Lavielle & Moulines (1997) consider estimation and testing of multiple changes in the mean of strong mixing random processes. The assumed mean function is piecewise constant with an unknown number of pieces.

Woodward & Gray (1993) perform a test of the existence of a linear trend in autoregressive moving average models with applications to global warming. For examples like the Argentina rainfall data, linearity is not expected, so that a nonparametric test is desirable; Brodsky & Darkhovsky (1993) present a systematic account of nonparametric methods, and Brillinger (1989) develops a test for monotonic trends with dependent errors. The latter two papers contain many further references to test for trend. Inference based on the monotonicity assumption is discussed in Robertson et al. (1988). In the context of density estimation, Woodroofe & Sun (1999) establish a test for uniformity versus a monotonicity. We show that our test statistic based on isotonic regression is asymptotically more powerful than the one proposed by Brillinger (1989).

The test statistic is presented in § 2 and illustrated by examples in § 3. Section 4 contains a power study and a comparison with Brillinger's test. Proofs are given in the Appendix.

2. The test statistic

2·1. *Preamble*: *Independent identically distributed normal errors*

In § 2·1, we assume that the $\{Z_k\}$ in (1) are independent normal random variables with mean 0 and known variance σ^2. Then the loglikelihood function is

$$l(\mu|\mathscr{X}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu_i)^2 + C, \tag{3}$$

where C denotes a generic constant that does not depend on μ. The parameter μ takes values in the space

$$\Omega = \{\mu = (\mu_1, \ldots, \mu_n) \in \mathbb{R}^n : \mu_1 \leqslant \mu_2 \leqslant \ldots \leqslant \mu_n\}. \tag{4}$$

Then the maximum likelihood estimator $\hat{\mu} = \arg\max_{\mu\in\Omega} l(\mu|\mathscr{X})$ is given by

$$\hat{\mu}_k = \max_{i\leqslant k}\min_{j\geqslant k}\frac{X_i + \ldots + X_j}{j-i+1} \tag{5}$$

(Robertson et al., 1988, p. 24). Under the null hypothesis H_0, the maximum likelihood estimator is obviously $\bar{\mu} = \bar{X} = n^{-1}\sum_{k=1}^{n} X_i$. The exact distribution of the likelihood ratio test statistic $2\{l(\hat{\mu}|\mathscr{X}) - l(\bar{\mu}|\mathscr{X})\}$ is given in Robertson et al. (1988, Ch. 2). As shown below, this test statistic is affected by the so-called spiking problem in large samples, in that $\hat{\mu}_1$

is too small while $\hat{\mu}_n$ is too large. Instead of (3), therefore, we shall consider the penalised loglikelihood function

$$l_{n,r}(\mu|\mathscr{X}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu_i)^2 - \frac{r\surd n}{\sigma^2}(\mu_n-\mu_1)+C,$$

where the term $r\surd n(\mu_n-\mu_1)$ penalises large $\mu_n-\mu_1$. Let $X_{1,r}=X_1+r\surd n$, $X_{n,r}=X_n-r\surd n$ and $X_{i,r}=X_i$ for $2\leqslant i\leqslant n-1$. Then

$$l_{n,r}(\mu|\mathscr{X}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_{i,r}-\mu_i)^2+C, \tag{6}$$

maximised at

$$\mu_k=\hat{\mu}_{k,r}=\max_{i\leqslant k}\min_{j\geqslant k}\frac{X_{i,r}+\ldots+X_{j,r}}{j-i+1}, \tag{7}$$

which is just formula (5) with X_i replaced by $X_{i,r}$. Furthermore,

$$\sum_{k=1}^{n}\hat{\mu}_{k,r}=\sum_{k=1}^{n}X_{k,r}=n\bar{X},$$

which suggests a test statistic of the form

$$\Lambda_{n,r}=\frac{1}{\hat{\sigma}_n^2}\sum_{k=1}^{n}(\hat{\mu}_{k,r}-\bar{X})^2, \tag{8}$$

where $\hat{\sigma}_n^2$ is a consistent estimator of σ^2. If σ^2 is known and $\hat{\sigma}_n^2=\sigma^2$, then $\Lambda_{n,r}$ is the penalised loglikelihood ratio statistic. The asymptotic null distribution of $\Lambda_{n,r}$ is presented in §§ 2·2 and 2·3.

2·2. *Short-range dependent errors*

For the remainder of § 2, suppose that the stationary sequence Z_k exhibits short range dependence in the following way. First, suppose that the covariances $\gamma(k)=E(Z_0Z_k)$ are absolutely summable:

$$\sum_{k=0}^{\infty}|\gamma(k)|<\infty. \tag{9}$$

Next let $T_k=\sum_{i=1}^{k}Z_i$, let $\mathscr{T}_n$ be a continuous piecewise linear function for which $\mathscr{T}_n(t)=T_k$ for $t=k/n$ and $k=0,1,\ldots,n$, and suppose that

$$\frac{\mathscr{T}_n}{\sigma\surd n}\to W \tag{10}$$

in distribution in $C[0,1]$, where W is a standard Brownian Motion and

$$\sigma^2=\gamma(0)+2\sum_{k=1}^{\infty}\gamma(k). \tag{11}$$

There are many families of processes for which (10) is satisfied. If Z_k is a linear process, $Z_k=\sum_{j=0}^{\infty}a_j\varepsilon_{k-j}$, say, where the ε_j are independent and identically distributed with mean 0 and finite variance and the a_j are absolutely summable, then (10) holds. Nonlinear processes that exhibit suitable mixing conditions satisfy (10); see Peligrad's (1986) review.

If we do not assume strong mixing, sufficient conditions are that $E(|Z_k|^p) < \infty$ for some $p > 2$ and that

$$\sum_{k=1}^{\infty} k^{-\alpha} E\{E(T_k|\ldots, Z_{-1}, Z_0)^2\} < \infty$$

for some $\alpha < 2$; see Maxwell & Woodroofe (2000).

The asymptotic distribution of $\Lambda_{n,r}$ is obtained in Theorems 1 and 2 below for local alternatives. For Theorem 1, suppose that

$$\mu_k = \mu + \frac{\sigma}{\sqrt{n}} \phi\left(\frac{k}{n}\right), \tag{12}$$

where ϕ is a right-continuous, nondecreasing function on $[0, 1]$ for which

$$\int_0^1 \phi(t)\, dt = 0.$$

The asymptotic null distribution is then a special case with $\phi = 0$. Let $c = r/\sigma$, let $S_{k,r} = \sum_{i=1}^{k} X_{i,r}$, let $G_{n,r}(t)$, for $t \in [0, 1]$, be a continuous, piecewise linear function for which $G_{n,r}(k/n) = S_{k,r}/n$ for $k = 1, \ldots, n$, and write S_k and G_n for $S_{k,0}$ and $G_{n,0}$. Define

$$H_{n,r}(t) = \frac{\sqrt{n}}{\sigma} \{G_{n,r}(t) - \bar{X} t\}.$$

Let functions $L_{n,r}$ and Φ_n be continuous and piecewise linear for which

$$L_{n,r}(0) = L_{n,r}(1) = 0, \quad L_{n,r}\left(\frac{1}{n}\right) = L_{n,r}\left(1 - \frac{1}{n}\right) = \frac{r}{\sqrt{n}}, \quad \Phi_n\left(\frac{k}{n}\right) = n^{-1} \sum_{p=1}^{k} \phi\left(\frac{p}{n}\right).$$

Then

$$H_{n,r}(t) = \frac{\mathscr{T}_n(t) - t\mathscr{T}_n(1)}{\sigma\sqrt{n}} + \Phi_n(t) - t\Phi_n(1) + \frac{\sqrt{n}}{\sigma} L_{n,r}(t). \tag{13}$$

Let $\Phi(t) = \int_0^t \phi(s)\, ds$ and define

$$\mathbb{B}_c^{\phi}(t) = W(t) - tW(1) + \Phi(t) + c\mathbb{1}_{(0,1)}(t). \tag{14}$$

Also, write $\mathbb{B}_c$ for $\mathbb{B}_c^{\phi}$ when $\phi \equiv 0$. Finally, if H is a bounded function on $[0, 1]$, let $\tilde{H}$ denote the greatest convex minorant of H and let $\tilde{h}$ denote the left-hand derivative of $\tilde{H}$. Thus $\tilde{b}_c^{\phi}$ is the left-hand derivative of the greatest convex minorant of $\mathbb{B}_c^{\phi}$. Then $-\infty < \tilde{b}_c^{\phi}(0+) \leqslant \tilde{b}_c^{\phi}(1-) < \infty$ with probability 1, by Lemma 6 of Woodroofe & Sun (1999).

Theorem 1. *If* (9), (10) *and* (12) *hold, then*

$$\frac{1}{\sigma^2} \sum_{k=1}^{n} (\hat{\mu}_{k,r} - \bar{X})^2 \to \int_0^1 \{\tilde{b}_c^{\phi}(t)\}^2\, dt \tag{15}$$

in distribution, where $c = r/\sigma$.

Proof. From Robertson et al. (1988, p. 7), $\hat{\mu}_{k,r} = \tilde{g}_{n,r}(k/n)$ for $k = 1, \ldots, n$. Thus $\hat{\mu}_{k,r} - \bar{X} = \sigma \tilde{h}_{n,r}(k/n)/\sqrt{n}$ since $G_{n,r}$ and $H_{n,r}$ are linearly related, and therefore

$$\frac{1}{\sigma^2} \sum_{k=1}^{n} (\hat{\mu}_{k,r} - \bar{X})^2 = \int_0^1 \{\tilde{h}_{n,r}(t)\}^2\, dt. \tag{16}$$

Thus it is necessary to show that the right-hand side of (16) converges in distribution to the right-hand side of (15). This seems plausible in view of (10) and (13). Details may be found in the Appendix.

Remark 1. If $c=0$, then the right-hand side of (15) is infinity with probability 1; see Groeneboom & Pyke (1983). This is the spiking problem mentioned above. The left-hand side of (15) does not have a nondegenerate asymptotic distribution under null hypothesis.

Remark 2. In the context of classical changepoint analysis, there is a similar problem. Suppose we want to test

$$H_1:\ \mu_1 = \ldots = \mu_\tau < \mu_{\tau+1} = \ldots = \mu_n.$$

Let $\bar{X}_r = r^{-1}\sum_{i=1}^r X_i$ and $\bar{X}'_{n-r} = (n-r)^{-1}\sum_{i=r+1}^n X_i$. It is mentioned in § 2·4 of Bhattacharya (1994) that the test statistic based on the likelihoood ratio

$$\max_{1\leqslant r\leqslant n-1} \{r^{-1} + (n-r)^{-1}\}^{-\frac{1}{2}}(\bar{X}'_{n-r} - \bar{X}_r)$$

does not have an asymptotic distribution, since it becomes unstable at the two ends.

2·3. *Estimating* σ^2

If the Z_k were observed, then the covariance function $\gamma(k) = E(Z_0 Z_k)$ could be estimated by the sample covariances $\bar{\gamma}_n(k) = n^{-1}\sum_{i=1}^{n-k} Z_i Z_{i+k}$. Then σ^2 could be estimated by the lag window estimators

$$\bar{\sigma}_n^2 = \bar{\gamma}_n(0) + 2\sum_{k=1}^{m_n} w\left(\frac{k}{m_n}\right)\bar{\gamma}_n(k), \tag{17}$$

where the lag window w satisfies $w(0)=1$, $w(x)=0$ for $|x|\geqslant 1$ and $|w(x)|\leqslant 1$ for all x; see Brockwell & Davis (1991, eqn (10.4.8)). Under conditions of their Theorem 10.4.1, $\bar{\sigma}_n^2$ is a consistent estimator of σ^2 provided that $m_n\to\infty$ and $m_n = o(n)$.

Recall that the unpenalised isotonic regression estimator $\hat{\mu}$ is given by (5). Hence the detrended process is $\hat{Z}_k = X_k - \hat{\mu}_k$. Note that the sum of residuals $\sum_{i=1}^n \hat{Z}_k = 0$. Let

$$\hat{\gamma}_n(k) = \frac{1}{n}\sum_{i=1}^{n-k} \hat{Z}_i \hat{Z}_{i+k} \tag{18}$$

and consider estimators of the form

$$\hat{\sigma}_n^2 = \hat{\gamma}_n(0) + 2\sum_{k=1}^{m_n} w\left(\frac{k}{m_n}\right)\hat{\gamma}_n(k), \tag{19}$$

where $m_n\to\infty$ and $m_n = o(n)$.

THEOREM 2. *Assume that the* Z_k *satisfy* (9) *and* (10), *and that* μ_k *is of the form* (12). *If* $m_n\to\infty$, $m_n = o(\sqrt{n})$ *and* $\bar{\sigma}_n^2$ *is a consistent estimator of* σ^2, *then* $\hat{\sigma}_n^2$ *converges to* σ^2 *in probability.*

Proof. In view of (17), it suffices to show that

$$\max_{0\leqslant k\leqslant n-1} |\hat{\gamma}_n(k) - \bar{\gamma}_n(k)| = O_p(n^{-\frac{1}{2}}); \tag{20}$$

see the Appendix for details.

Meyer & Woodroofe (2000) have established the asymptotic normality of $\hat{\gamma}_n(0)$ when the errors are independent and normally distributed.

In the remainder of the paper $\Lambda_{n,r}$ is defined by (8) with $\hat{\sigma}_n^2$ defined by (19).

Corollary 1. *Suppose that H_0 is true. Then, under the conditions of Theorem 2,*

$$\Lambda_{n,r} \to \int_0^1 \{\tilde{b}_c(t)\}^2 \, dt \tag{21}$$

in distribution, where $c = r/\sigma$.

In practice, r must be $c\hat{\sigma}_n$. This does not affect the asymptotic distribution.

3. Applications

3·1. *Estimation of the asymptotic null distribution of $\Lambda_{n,r}$*

The asymptotic null distribution of $\Lambda_{n,r}$ is complicated, but percentiles can be computed by simulation. We simulated 30 000 Brownian motions and computed the right-hand side of (21) for each of them. The asymptotic null distribution was estimated from these values. Brownian motion was simulated by generating standard normal variates Z_k and forming piecewise linear functions with values $(Z_1 + \ldots + Z_k)/\sqrt{n}$ at $t = k/n$ for $k = 1, \ldots, n$. The number of grid points was $n = 20$, 50, 100, 200, 500, 1000, 2000 and 4000. The critical values $\lambda_c(\alpha)$ are given in Table 1 for $\alpha = 0{\cdot}05$ and 0·01 and $c = 0{\cdot}05$, 0·10 and 0·15.

Table 1. *Critical values $\lambda_c(\alpha)$ for $c = 0{\cdot}05$, 0·1, 0·15 and $\alpha = 0{\cdot}01$, 0·05*

n	$c = 0{\cdot}15$		$c = 0{\cdot}10$		$c = 0{\cdot}05$	
	0·05	0·01	0·05	0·01	0·05	0·01
20	5·32	8·65	6·10	9·62	7·01	10·76
50	6·01	9·47	6·99	10·68	8·16	12·27
100	6·59	10·27	7·70	11·58	9·13	13·36
200	6·95	10·74	8·17	12·23	9·82	14·30
500	7·19	11·27	8·58	12·85	10·49	15·30
1000	7·34	11·39	8·70	12·98	10·70	15·25
2000	7·45	11·35	8·88	13·03	11·02	15·56
4000	7·49	11·35	8·95	13·18	11·13	15·58

In the examples below, we used the values $c = 0{\cdot}15$ and $m_n = n^{1/3}$, and the truncated window, namely $w(x) = 1$, for $|x| \leqslant 1$, and $w(x) = 0$ otherwise. These values performed well in simulations, and 0·15 was chosen for c after comparing power curves for selected alternatives. Programs in S-Plus and Matlab and the Argentina rainfall data below are available at www.stat.lsa.umich.edu/~michaelw.

3·2. *Argentina rainfall data*

Figure 1 shows the volume of yearly rainfall in Argentina from 1884 to 1996 along with the penalised and unpenalised isotonic regression functions. There is a mild spiking problem: the year 1884 has a low value. The spiking problem is clearly suppressed in the penalised isotonic regression. From an autocorrelation plot, the residuals appear to be short range dependent. Only the eighth of the first twenty autocorrelations is significant at the 0·05 level. From (8) and (19), $\hat{\sigma}^2 = 253{\cdot}57$ and $\Lambda_{n,r} = 17{\cdot}95$, which is significant at

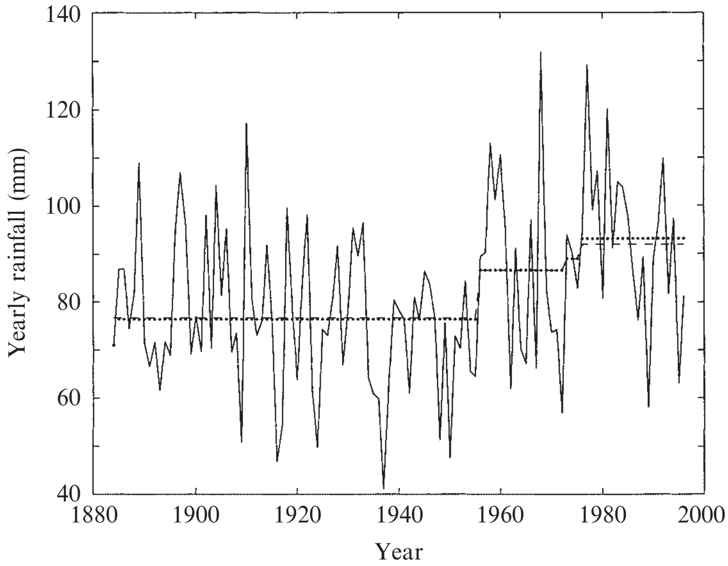

Fig. 1. Argentina rainfall data: yearly rainfall (millimetres) in Argentina from 1884 to 1996 with its penalised (dashed line) and unpenalised (dotted) isotonic regression functions.

the 0·01 level since the 0·99 quantile is between 10·27 and 10·74 from Table 1. Thus the apparent increase in the mean is highly significant.

From the penalised isotonic regression in Fig. 1, we can see there is a large jump around the years 1955–1956. This observation justifies the belief of meteorologist Eng Cesar Lamelas that the increase might be caused by the construction of a dam. The mean rainfall volume is 76·34 before 1956 and 90·11 after 1956.

Now let us consider Brillinger's test. If we use the estimated variance $\hat{\sigma}^2 = 253{\cdot}57$, the test statistic, see Brillinger (1989, eqn (2.3)), is

$$\frac{\sum_{k=1}^{n} \xi(k) X_k}{\hat{\sigma}\{\sum_{k=1}^{n} \xi^2(k)\}^{\frac{1}{2}}} = \frac{41{\cdot}0621}{29{\cdot}1645} = 1{\cdot}4079,$$

where

$$\xi(t) = [(t-1)\{1-(t/n-1/n)\}]^{\frac{1}{2}} - \{t(1-t/n)\}^{\frac{1}{2}} \quad (1 \leqslant t \leqslant n).$$

Then the p-value is $\mathrm{pr}(Z > 1{\cdot}4079) = 0{\cdot}0796$, where Z is standard normal. Hence, at the 5% level of significance, Brillinger's test is unable to detect the trend. Section 4 contains some power comparisons.

3·3. *Global warming data*

The issue of global warming has received wide attention recently within the scientific community and more generally; see Melillo (1999) and Delworth & Knutson (2000) for some scientific perspectives. The question is whether the observed monotonic trend is caused by natural variability or by greenhouse gases generated by human activity. In this paper, we interpret the former as the short range dependent background noise and the latter as the mean trend. We are therefore testing nonparametrically for an increase.

The global warming data considered in this paper are provided by P. D. Jones et al., see http://cdiac.esd.ornl.gov/trends/temp/jonescru/jones.html. They contain annual temperature anomalies from 1856 to 1998; see Fig. 2. The estimated variance and test statistic are $\hat{\sigma}^2 = 0{\cdot}01558$ and $\Lambda_{n,r} = 349{\cdot}495$; the latter is highly significant, beyond the range of Table 1. Brillinger's test also gives highly significant results in this case.

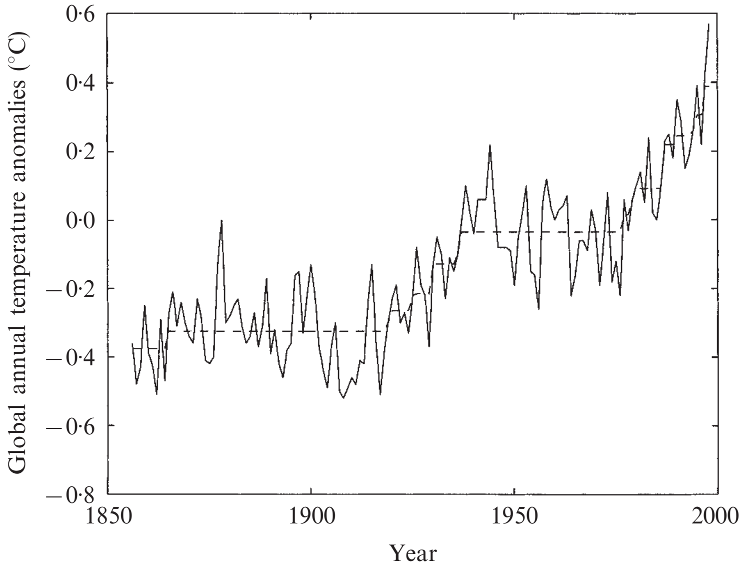

Fig. 2. Global warming data: annual temperature anomalies from 1856 to 1999 (relative to the 1961–1990 mean) with its penalised isotonic regression (dashed lines).

4. Power study

In this section, the power of our test statistic $\Lambda_{n,r}$ is compared to that of the test proposed by Brillinger. Suppose that the trend has the form

$$\mu_k = \frac{\delta}{\sqrt{n}} \phi\left(\frac{k}{n}\right) \quad (\delta > 0), \tag{22}$$

where ϕ is a nondecreasing function with $\int_0^1 \phi(t)\,dt = 0$. The test developed by Brillinger (1989) is unable to detect the trend (22) with $\delta = \delta_n \to \infty$ and $\delta_n = o\{(\log n)^{\frac{1}{2}}\}$; see Brillinger's equation (3.2).

On the other hand, suppose that ϕ is not identically 0. Then, by Theorem 1, the asymptotic power of test based on (8) is given by $\mathrm{pr}[\int_0^1 \{\tilde{b}_c^{\rho\phi}(t)\}^2\,dt > \lambda_c(\alpha)]$, where $\rho = \delta/\sigma$. By Proposition 1 below, this power converges to 1 if $\delta_n \to \infty$.

Proposition 1. *Let* $J_n(t) = \Phi(t) + \rho_n^{-1}\mathbb{B}_c(t)$, *for* $t \in [0, 1]$. *Then, as* $\rho_n \to \infty$,

$$\| \tilde{j}_n \|_2^2 = \| \phi \|_2^2 + O_P(\rho_n^{-\frac{1}{2}}).$$

Proof. By Marshall's lemma, $\| \tilde{J}_n - \Phi \| = O_P(\rho_n^{-1})$. Observe that

$$\tilde{j}_n(0+) = \inf_{0<t\leqslant 1} \frac{J_n(t)}{t} = O_P(1)$$

and similarly that $\tilde{j}_n(1) = O_P(1)$. Then Lemma A3 in the Appendix completes the proof. □

Figure 3(a) presents a simulated power study, when Z_k are independent and identically distributed as standard normal, $\mu_k = -\frac{1}{2}$, for $1 \leqslant k \leqslant 100$, $\mu_k = \frac{1}{2}$, for $101 \leqslant k \leqslant 200$, and $\delta = k/100$, for $k = 0, 1, \ldots, 99$. From the plot, we see that the testing procedure based on isotonic regression is more powerful than Brillinger's test. The uniformly most powerful test, of course, is the best of the three. In Fig. 3(b), the mean function is linear: $\mu(k) = k/200 - 0{\cdot}5$, for $1 \leqslant k \leqslant 200$, and $\delta = k/100$, for $k = 0, 1, \ldots, 99$. Similar conclusions hold.

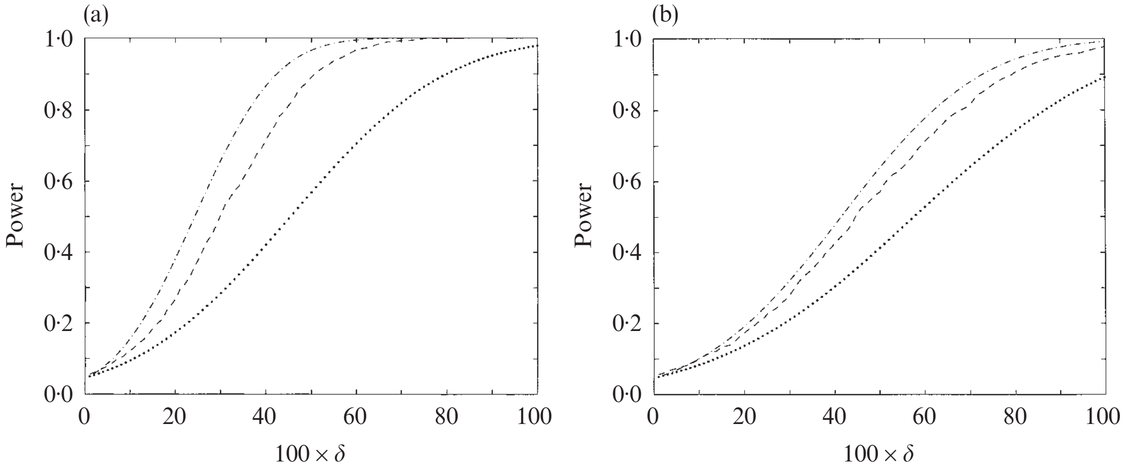

Fig. 3. Power curves for the uniformly most powerful test (dot-dashed lines), the isotonic regression test (dashed) and Brillinger's (1989) test (dotted) for (a) step function with independently distributed $N(0, 1)$ errors, and (b) linear function with independently distributed $N(0, 1)$ errors.

Acknowledgement

This research was supported by the National Science Foundation and US Army Research Office.

Appendix

Proofs

Proof of Theorem 1. The proof uses strong approximation. If the probability space is sufficiently rich, then there are versions $\mathscr{T}'_n$ of $\mathscr{T}_n$ and Brownian motions W_n such that

$$\sup_{0 \leqslant t \leqslant 1} |(\sigma\sqrt{n})^{-1} \mathscr{T}'_n(t) - W_n(t)| \to 0$$

in probability, and there is no loss of generality in supposing that $\mathscr{T}'_n = \mathscr{T}_n$ for the purposes of proving Theorem 1.

Recall the definitions (13) and (14) of H_n and $\mathbb{B}^{\phi}_c$ and relation (16), and define $\mathbb{B}^{\phi}_{n,c}$ by (14) with W replaced by W_n. Then it suffices to show that

$$\int_0^1 \{\tilde{h}_n(t)\}^2\, dt - \int_0^1 \{\tilde{b}^{\phi}_{n,c}(t)\}^2\, dt \to 0 \tag{A1}$$

in probability as $n \to \infty$, since $\int_0^1 \{\tilde{b}^{\phi}_{n,c}(t)\}^2\, dt$ has the same distribution as $\int_0^1 \{\tilde{b}^{\phi}_c(t)\}^2\, dt$ for all n.

The following lemmas are from Woodroofe & Sun (1999). In their statements, G and H are bounded functions on $[0, 1]$ and $\|G\| = \sup_{0 \leqslant t \leqslant 1} |G(t)|$.

LEMMA A1. *If G is lower semi-continuous at* 0 (*respectively,* 1), *then* $\tilde{G}(0) = G(0)$ (*respectively,* $\tilde{G}(1) = G(1)$).

LEMMA A2. *If $B \subseteq [0, 1]$, $|G(t) - H(t)| \leqslant \varepsilon$, for all $t \in B$, and $|\tilde{G}(t) - \tilde{H}(t)| \leqslant \varepsilon$, for all $t \in B'$, then* $\|\tilde{G} - \tilde{H}\| \leqslant \varepsilon$.

LEMMA A3. *If*

$$\tilde{G}(0) = \tilde{H}(0), \quad \tilde{G}(1) = \tilde{H}(1) \quad (-\infty < \tilde{g}(0+) \leqslant \tilde{g}(1-) < \infty, \ -\infty < \tilde{h}(0+) \leqslant \tilde{h}(1-) < \infty),$$

then

$$\int_0^1 \{\tilde{h}(t) - \tilde{g}(t)\}^2 \, dt \leqslant \|\tilde{H} - \tilde{G}\| \{\tilde{g}(1-) - \tilde{g}(0+) + \tilde{h}(1-) - \tilde{h}(0+)\}.$$

LEMMA A4. *For any $c > 0$, $-\infty < \tilde{b}_c^{\phi}(0+) \leqslant \tilde{b}_c^{\phi}(1-) < \infty$ with probability* 1. *Furthermore, $\tilde{h}_n(0+)$ and $\tilde{h}_n(1-)$ are stochastically bounded.*

Proofs. Lemmas A1, A2 and A3 follow by adapting Lemmas 1, 2 and 3 of Woodroofe & Sun (1999) from the nonincreasing case to the nondecreasing one, and the first assertion of Lemma A4 is contained in Lemma 6 of Woodroofe & Sun (1999). For the stochastic boundedness, first write

$$\begin{aligned} \mathrm{pr}\{\tilde{h}_n(0+) < -\lambda\} &= \mathrm{pr}\left(\sqrt{n} \min_{k \leqslant n} \frac{S_{k,r}}{\sigma k} < -\lambda\right) \\ &\leqslant \mathrm{pr}\left(\max_{k \leqslant n\delta} \left|\frac{S_k}{\sigma(n\delta)^{\frac{1}{2}}}\right| > \frac{c}{\sqrt{\delta}}\right) + \mathrm{pr}\left(\max_{n\delta \leqslant k \leqslant n} \left|\frac{S_k}{\sigma\sqrt{n}}\right| > \lambda\sqrt{\delta}\right), \end{aligned}$$

with $\delta = 1/\sqrt{\lambda}$. Since $\max_{k \leqslant n} |S_k/\sqrt{n}|$ is stochastically bounded, it then follows that $\mathrm{pr}\{\tilde{h}_n(0+) < -\lambda\} \to 0$, as first $n \to \infty$ and then $\lambda \to \infty$. This establishes the stochastic boundedness of $\tilde{h}_n(0+)$. The right endpoint may be handled similarly.

To apply the lemmas to the problem at hand, first observe that

$$\tilde{\mathbb{B}}_{n,c}^{\phi}(0) = 0 = \tilde{H}_n(0), \quad \tilde{\mathbb{B}}_{n,c}^{\phi}(1) = 0 = \tilde{H}_n(1)$$

by Lemma A1. Thus,

$$\int_0^1 \{\tilde{h}_n(t) - \tilde{b}_{n,c}^{\phi}(t)\}^2 \, dt \leqslant \|\tilde{H}_n - \tilde{\mathbb{B}}_{n,c}^{\phi}\| \{\tilde{h}_n(1-) - \tilde{h}_n(0+) + \tilde{b}_{n,c}^{\phi}(1-) - \tilde{b}_{n,c}^{\phi}(0+)\}.$$

The second factor on the right-hand side is stochastically bounded, by Lemma A4. Thus it suffices to show that the first factor approaches 0 as $n \to \infty$. If $t \leqslant 1/n$, then

$$|\tilde{H}_n(t) - \tilde{\mathbb{B}}_{n,c}^{\phi}(t)| + |\tilde{H}_n(1-t) - \tilde{\mathbb{B}}_{n,c}^{\phi}(1-t)| \leqslant 2\,\frac{|\tilde{h}_n(1-)| + |\tilde{h}_n(0+)| + |\tilde{b}_{n,c}^{\phi}(1-)| + |\tilde{b}_{n,c}^{\phi}(0+)|}{n},$$

which approaches 0 in probability. On the other hand, if $1/n \leqslant t \leqslant 1 - 1/n$, then

$$|H_n(t) - \mathbb{B}_{n,c}^{\phi}(t)| \leqslant 2\left\|\frac{\mathscr{T}_n}{\sigma\sqrt{n}} - W_n\right\| + \|\Phi_n - \Phi\|,$$

which also approaches 0 in probability as $n \to \infty$. That $\|\tilde{H}_n - \tilde{\mathbb{B}}_{n,c}^{\phi}\| \to 0$ in probability as $n \to \infty$ now follows from Lemma A2 with $B = [1/n, 1 - 1/n]$, completing the proof of (A1) and, therefore, of Theorem 1. □

Proof of (20). The following lemma is needed.

Lemma A5. *If* $-\infty < a_1 \leqslant \ldots \leqslant a_n < \infty$ *and* $-\infty < b_1, \ldots, b_n < \infty$, *then*

$$\left| \sum_{j=1}^{n} a_j b_j \right| \leqslant \{|a_n| + (a_n - a_1)\} \max_{j \leqslant n} \left| \sum_{i=1}^{j} b_i \right|.$$

Proof. Let $B_k = b_1 + \ldots + b_k$. Then the term on the left-hand side is

$$\left| a_n B_n - \sum_{j=2}^{n} (a_j - a_{j-1}) B_{j-1} \right| \leqslant \{|a_n| + (a_n - a_1)\} \max_{j \leqslant n} |B_j|,$$

as asserted.

To prove (20), now write

$$\hat{\gamma}_n(k) - \bar{\gamma}_n(k) = \frac{1}{n} \sum_{i=1}^{n-k} (\hat{Z}_i \hat{Z}_{i+k} - Z_i Z_{i+k}) = \mathrm{I}_n(k) + \mathrm{II}_n(k) + \mathrm{III}_n(k),$$

where

$$\mathrm{I}_n(k) = \frac{1}{n} \sum_{i=1}^{n-k} (\mu_i - \hat{\mu}_i) Z_{i+k}, \quad \mathrm{II}_n(k) = \frac{1}{n} \sum_{i=1}^{n-k} Z_i (\mu_{i+k} - \hat{\mu}_{i+k}),$$

$$\mathrm{III}_n(k) = \frac{1}{n} \sum_{i=1}^{n-k} (\mu_i - \hat{\mu}_i)(\mu_{i+k} - \hat{\mu}_{i+k}).$$

Here

$$\mathrm{I}_n(k) \leqslant 2\{|\mu_n| + |\mu_1| + |\hat{\mu}_1| + |\hat{\mu}_n|\} \max_{j \leqslant n} |Z_1 + \ldots + Z_j|/n$$

for each k, and $\hat{\mu}_n$ and $\hat{\mu}_1$ are stochastically bounded, by Lemma A4. Thus $\max_{j \leqslant n} |\mathrm{I}_n(k)| = O_P(n^{-\frac{1}{2}})$, and similarly $\max_{j \leqslant n} |\mathrm{II}_n(k)| = O_P(n^{-\frac{1}{2}})$. For $\mathrm{III}_n(k)$, let $\bar{\mu}_n = n^{-1} \sum_{k=1}^{n} \mu_k$ Then

$$\sum_{i=1}^{n} (\hat{\mu}_i - \mu_i)^2 \leqslant 4 \sum_{i=1}^{n} (\hat{\mu}_i - \bar{X}_n)^2 + 4 \sum_{i=1}^{n} (\mu_i - \bar{\mu}_n)^2 + 4n(\bar{X}_n - \bar{\mu}_n)^2,$$

which is stochastically bounded, by Theorem 1, (9) and (12). That $\max_{k \leqslant n} |\mathrm{III}_n(k)| = O_P(1/n)$ then follows from Schwarz's inequality. □

References

Bhattacharya, P. K. (1994). 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.

Brillinger, D. R. (1989). Consistent detection of a monotonic trend superposed on a stationary time series. *Biometrika*. **76**, 23–30.

Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods*. New York: Springer.

Brodsky, B. E. & Darkhovsky, B. S. (1993). *Nonparametric Methods in Change Point Problems*. Dordrecht: Kluwer.

Darkhovsky, B. S. (1994). Nonparametric methods in change-point problems: a general approach and some concrete algorithms. In *Change-Point Problems*, Ed. E. Carlstein, H. Müller and D. Siegmund, pp. 99–107. Hayward, CA: Inst. Math. Statist.

Delworth, T. L. & Knutson, T. R. (2000). Simulation of early 20th century global warming. *Science* **287**, 2246–50.

Groeneboom, P. & Pyke, R. (1983). Asymptotic normality of statistics based on the convex minorants of empirical distribution functions. *Ann. Prob.* **11**, 328–45.

Lavielle, M. & Moulines, É. (1997). Détection de ruptures multiples dans la moyenne d'un processus aléatoire. *C. R. Acad. Sci. Paris Ser. I Math.* **324**, 239–43.

Lombard, F. & Hart, J. D. (1994). The analysis of change-point data with dependent errors. In *Change-Point Problems*, Ed. E. Carlstein, H. Müller and D. Siegmund, pp. 194–209. Hayward, CA: Inst. Math. Statist.

Maxwell, M. & Woodroofe, M. (2000). Central limit theorems for additive functionals of Markov chains. *Ann. Prob.* **28**, 713–24.

MELILLO, J. M. (1999). Climate change: warm, warm on the Range. *Science* **283**, 183–4.
MEYER, M. & WOODROOFE, M. (2000). On the degrees of freedom in shape-restricted regression *Ann. Statist.* **28**, 1083–104.
PELIGRAD, M. (1986). Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey). In *Dependence in Probability and Statistics: A Survey of Recent Results*, Ed. E. Eberlein and M. Taqqu, pp. 193–224. Boston, MA: Birkhauser.
ROBERTSON, T., WRIGHT, F. T. & DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. New York: John Wiley.
SHABAN, S. A. (1980). Change point problem and two-phase regression: An annotated bibliography. *Int. Statist. Rev.* **48**, 83–93.
SIEGMUND, D. (1986). Boundary crossing probabilities and statistical applications. *Ann. Statist.* **14**, 361–404.
WOODROOFE, M. & SUN, J. (1999). Testing uniformity versus a monotone density. *Ann. Statist.* **27**, 338–60.
WOODWARD, W. A. & GRAY, H. L. (1993). Global warming and the problem of testing for trend in time-series data. *J. Climate* **6**, 953–62.

[*Received June* 2000. *Revised January* 2001]