The Annals of Statistics
1995, Vol. 23, No. 1, 282–304

TESTING FOR A CHANGE IN THE PARAMETER VALUES AND ORDER OF AN AUTOREGRESSIVE MODEL

BY RICHARD A. DAVIS,[1] DAWEI HUANG[2] AND YI-CHING YAO

Colorado State University, Queensland University of Technology and Colorado State University

The problem of testing whether or not a change has occurred in the parameter values and order of an autoregressive model is considered. It is shown that if the white noise in the AR model is weakly stationary with finite fourth moments, then under the null hypothesis of no changepoint, the normalized Gaussian likelihood ratio test statistic converges in distribution to the Gumbel extreme value distribution. An asymptotically distribution-free procedure for testing a change of either the coefficients in the AR model, the white noise variance or the order is also proposed. The asymptotic null distribution of this test is obtained under the assumption that the third moment of the noise is zero. The proofs of these results rely on Horváth's extension of Darling–Erdös' result for the maximum of the norm of a k-dimensional Ornstein–Uhlenbeck process and an almost sure approximation to partial sums of dependent random variables.

1. Introduction. The problem of detecting a change in the distributional structure of an underlying process has been extensively studied in the literature of quality control, time series, signal processing and dynamical systems. [See, e.g., Bagshaw and Johnson (1977), Basseville and Benveniste (1983, 1986), Picard (1985), Siegmund (1985), Telksnys (1986), Tsay (1988), Willsky (1976) and references therein, where various settings and formulations are considered.] Often, in a fixed sample setting, the primary interest is to test that a change has occurred in the level and/or the covariance structure of the process, and if a change has been detected, then it would be desirable to estimate the location of the changepoint(s). In a sequential setting, the objective is to stop the process and take corrective action as soon as possible after a change occurs while keeping the false alarm rate low.

In this paper, we consider the problem of testing whether or not a change has occurred in the parameter values of an autoregressive model. This incorporates both possibilities of either a shift in the level of the process or a change in the autocovariance structure.

Received April 1993; revised October 1993.
[1]Research partially supported by NSF Grant DMS-91-00392.
[2]Research supported by QUT Meritorious Project Grant 181970010.
AMS 1991 *subject classifications.* Primary 62F05; secondary 62M10, 60G10.
Key words and phrases. Likelihood ratio statistic, changepoint, autoregressive process, strong mixing.

To set the problem up, let $X_1, \ldots, X_n$ be n consecutive observations from the model

$$(1.1) \qquad \begin{aligned} X_t &= \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \qquad -\infty < t \le \tau, \\ &= \alpha_0 + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \varepsilon_t, \qquad t \ge \tau + 1, \end{aligned}$$

where $\tau \in (p, n]$, $\{\varepsilon_t\}$ is a fourth-order white noise sequence, that is, for all $i \le j \le k \le l$,

$$(1.2)(\text{i}) \qquad E(\varepsilon_i) = 0,$$

$$(1.2)(\text{ii}) \qquad E(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2, & \text{if } i = j, \\ 0, & \text{if } i < j, \end{cases}$$

$$(1.2)(\text{iii}) \qquad E(\varepsilon_i \varepsilon_j \varepsilon_k) = \begin{cases} \mu_3, & \text{if } i = j = k, \\ 0, & \text{otherwise}, \end{cases}$$

$$(1.2)(\text{iv}) \qquad E(\varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l) = \begin{cases} \mu_4, & \text{if } i = j = k = l, \\ \sigma^4, & \text{if } i = j < k = l, \\ 0, & \text{otherwise}, \end{cases}$$

and $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ satisfies the causality condition $\phi(z) \ne 0$ for all $|z| \le 1$. While we have assumed that p is known, it need not be the "true" order of the model (i.e., the largest p for which $\phi_p \ne 0$). One possibility is to let p be an upper bound on the true order of the model.

The principal objective of this paper is to study the asymptotic behavior of the likelihood ratio statistic in testing H_0: $\tau = n$ (no change has occurred) versus H_1: $\tau < n$ (a change has occurred) under the null hypothesis. Here likelihood is defined in terms of the Gaussian likelihood based on the observed data. It is common practice in time series analysis to use the Gaussian likelihood for inference-based procedures even though the underlying process may not be Gaussian. Of course, if the noise is Gaussian, then the Gaussian likelihood is the correct likelihood of the process.

The asymptotic operating characteristics of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates was examined by Yao and Davis (1986). Using a result of Darling and Erdös (1956), the asymptotic distribution of the likelihood ratio statistic under the null hypothesis of no change was derived. By extending the Darling–Erdös result, Horváth (1993) considered the likelihood ratio statistic in testing for a change in both the mean and variance of normal variates. Using Horváth's extension of the Darling–Erdös result, the limit distribution of the likelihood ratio, under H_0, is derived in Section 2 for the autoregressive model (1.1). By taking $p = 0$, this extends the Yao–Davis distributional result for testing a change in the mean in a sequence of independent nonnormal observations. (Of course, if the distribution of observations is known up to a location change, then a test based on the actual likelihood rather than the *Gaussian* likelihood should be used.)

In Section 3, we investigate the null asymptotic behavior of the likelihood ratio statistic when the orders of the AR model in (1.1) are permitted to be different before and after the changepoint. Also, the case when the white noise variance shifts at the changepoint is considered under some moment constraints on ε_t. Additionally, an asymptotically distribution-free test is proposed whose asymptotic null distribution is obtained under the assumption that the third moment of ε_t is zero.

A summary of a simulation study comparing the approximation of the limit distribution to the null distribution of the test statistics of Sections 2 and 3 is contained in Section 4. Overall, it was found that the limit distribution provides a reasonable approximation to the distribution of the test statistics for a variety of sample sizes and parameter values of the autoregressive model.

The more lengthy proofs of the results in Sections 2 and 3 are contained in the Appendix.

2. Gaussian likelihood ratio. In this section, we consider the limiting behavior of the likelihood ratio statistic for testing whether or not a change has occurred in an autoregressive process. Let $X_1, \dots, X_n$ be n consecutive observations from the model (1.1) and (1.2).

For the present we shall assume that $\sigma^2 = 1$. Extensions to the case when σ^2 is unknown are discussed in Remark 2.3. In testing H_0: $\tau = n$ (no change has occurred) versus H_1: $\tau = k$ (a change has occurred at time k), the Gaussian likelihood ratio, conditional on the first p observations and after taking $-2\ln$, is given by

$$\begin{aligned}\Lambda_n(k) := -2\ln\left(\frac{L(n)}{L(k)}\right) &= \min_{\phi} \sum_{t=p+1}^{n} (X_t - \phi_0 - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p})^2 \\ &\quad - \min_{\phi} \sum_{t=p+1}^{k} (X_t - \phi_0 - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p})^2 \\ &\quad - \min_{\alpha} \sum_{t=k+1}^{n} (X_t - \alpha_0 - \alpha_1 X_{t-1} - \cdots - \alpha_p X_{t-p})^2 \\ &= Q_1 - Q_2 - Q_3,\end{aligned}$$

where

$$\begin{aligned} Q_1 &= \mathbf{X}_n'\mathbf{X}_n - \mathbf{X}_n' M_n (M_n' M_n)^{-1} M_n' \mathbf{X}_n, \\ Q_2 &= \mathbf{X}_k'\mathbf{X}_k - \mathbf{X}_k' M_k (M_k' M_k)^{-1} M_k' \mathbf{X}_k, \\ Q_3 &= \tilde{\mathbf{X}}_k'\tilde{\mathbf{X}}_k - \tilde{\mathbf{X}}_k' \tilde{M}_k \left(\tilde{M}_k' \tilde{M}_k\right)^{-1} \tilde{M}_k' \tilde{\mathbf{X}}_k, \\ \mathbf{X}_k &= (X_{p+1}, \dots, X_k)', \\ \tilde{\mathbf{X}}_k &= (X_{k+1}, \dots, X_n)', \end{aligned}$$

$$M_k = \begin{bmatrix} 1 & X_p & X_{p-1} & \cdots & X_1 \\ 1 & X_{p+1} & X_p & \cdots & X_2 \\ \vdots & \vdots & \cdots & \vdots & \\ 1 & X_{k-1} & X_{k-2} & \cdots & X_{k-p} \end{bmatrix}$$

and

$$\tilde{M}_k = \begin{bmatrix} 1 & X_k & X_{k-1} & \cdots & X_{k-p+1} \\ 1 & X_{k+1} & X_k & \cdots & X_{k-p+2} \\ \vdots & \vdots & \cdots & \vdots & \\ 1 & X_{n-1} & X_{n-2} & \cdots & X_{n-p} \end{bmatrix}$$

Writing $\boldsymbol{\varepsilon}_n = (\varepsilon_{p+1}, \dots, \varepsilon_n)'$ we have under H_0,

$$\mathbf{X}_n = M_n \boldsymbol{\phi} + \boldsymbol{\varepsilon}_n,$$

so that, with an obvious notation,

$$Q_1 = \boldsymbol{\varepsilon}'_n \boldsymbol{\varepsilon}_n - \boldsymbol{\varepsilon}'_n M_n (M'_n M_n)^{-1} M'_n \boldsymbol{\varepsilon}_n,$$

$$Q_2 = \boldsymbol{\varepsilon}'_k \boldsymbol{\varepsilon}_k - \boldsymbol{\varepsilon}'_k M_k (M'_k M_k)^{-1} M'_k \boldsymbol{\varepsilon}_k,$$

$$Q_3 = \tilde{\boldsymbol{\varepsilon}}'_k \tilde{\boldsymbol{\varepsilon}}_k - \tilde{\boldsymbol{\varepsilon}}'_k \tilde{M}_k \left(\tilde{M}'_k \tilde{M}_k\right)^{-1} \tilde{M}'_k \tilde{\boldsymbol{\varepsilon}}_k.$$

Finally, setting $\mathbf{S}_k = M'_k \boldsymbol{\varepsilon}_k$, $P_k = (M'_k M_k)^{-1}$ and $\tilde{P}_k = (\tilde{M}'_k \tilde{M}_k)^{-1}$, we have

$$\Lambda_n(k) = -\mathbf{S}'_n P_n \mathbf{S}_n + \mathbf{S}'_k P_k \mathbf{S}_k + (\mathbf{S}_n - \mathbf{S}_k)' \tilde{P}_k (\mathbf{S}_n - \mathbf{S}_k). \tag{2.1}$$

To study the asymptotic behavior of the likelihood ratio statistic under H_0, we rescale the time axis and consider the process $Q_n(t) := \Lambda_n([nt])$ ($[s]$ = integer part of s) on the set $t \in [0, 1]$. For the remainder of this discussion we assume that $\sup_t E|\varepsilon_t|^{4+\delta} < \infty$ for some $0 < \delta \le 1$ and the process $\{X_t\}$ is strongly mixing with a mixing function $\rho(n)$ satisfying $\rho(n) \ll n^{-(1+\varepsilon)(1+4/\delta)}$ for some $\varepsilon > 0$ (see Remark 2.1 for sufficient conditions). The asymptotic behavior of $Q_n(t)$ is essentially governed by that of $S_{[nt]}$ and $P_{[nt]}$. First note that, by the causality assumption and (1.2), $\mathbf{S}_n$ is the partial sum of $n - p$ uncorrelated terms from a $p + 1$-dimensional weakly stationary strongly mixing sequence with covariance matrix $\Gamma_{p+1} = [\gamma_{ij}]_{i,j=1,\dots,p+1}$, where

$$\gamma_{ij} = \begin{cases} 1, & \text{if } i = j = 1, \\ \mu, & \text{if } i = 1, j > 1 \text{ or } i > 1, j = 1, \\ E(X_i X_j), & \text{if } i > 1, j > 1, \end{cases}$$

and $\mu = EX_i$. Applying Theorem 4 in Kuelbs and Philipp (1980) to the sequence $\xi_n = \mathbf{S}_n - \mathbf{S}_{n-1}$, there exists a sequence $\{\mathbf{Z}_n\}$ of iid Gaussian random vectors, defined on possibly a new probability space, with mean zero and covariance matrix Γ_{p+1} such that

$$\mathbf{S}_k - \mathbf{U}_k = O(k^{1/2-\lambda}) \quad \text{as } k \to \infty \text{ a.s.}, \tag{2.2}$$

for some $\lambda > 0$, where $\mathbf{U}_k = \sum_{i=p+1}^{k} \mathbf{Z}_i$. It is immediate that

$$\Gamma_{p+1}^{-1/2} \mathbf{S}_{[n\cdot]}/\sqrt{n} \to_d \mathbf{W}(\cdot) \tag{2.3}$$

in $D^{p+1}[0,1]$, where $\mathbf{W}(\cdot)$ is $p+1$-dimensional Brownian motion with covariance matrix I_{p+1}.

Turning to P_n, the (i,j) ($i > 1$ and $j > 1$) component of P_n^{-1} is $\sum_{s=1}^{n-p} X_{p+1-i+s} X_{p+1-j+s}$. Now, applying Theorem 4 in Kuelbs and Philipp (1980) once again to the strongly mixing sequence $\{X_{p+1-i+s}X_{p+1-j+s} - E(X_i X_j),\ s = 1, 2, \dots\}$, which has the same mixing rate as $\{X_t\}$, we have

$$\sum_{s=1}^{n-p} \left(X_{p+1-i+s} X_{p+1-j+s} - E(X_i X_j) \right) - V(n-p) = O(n^{1/2-\lambda}) \tag{2.4}$$

as $n \to \infty$ a.s.,

for some $\lambda > 0$, where $V(t)$ is Brownian motion with variance

$$\operatorname{Var}(X_i X_j) + 2 \sum_{s=1}^{\infty} \operatorname{Cov}(X_{-i} X_{-j}, X_{s-i} X_{s-j}).$$

Since this argument can be applied to all of the components of P_n^{-1}, it follows easily that

$$nP_{[nt]} \to t^{-1} \Gamma_{p+1}^{-1} \tag{2.5}$$

uniformly on $[t_1, t_2]$, $0 < t_1 < t_2 \le 1$. We conclude from (2.3), (2.5), the continuous mapping theorem and the observation, $\tilde{P}_{[nt]}^{-1} = P_n^{-1} - P_{[nt]}^{-1}$, that

$$\begin{aligned} Q_n(t) &\to_d \frac{\|\mathbf{W}(t)\|^2}{t} + \frac{\|\mathbf{W}(1) - \mathbf{W}(t)\|^2}{1-t} - \|\mathbf{W}(1)\|^2 \\ &= \frac{\|\mathbf{W}(t) - t\mathbf{W}(1)\|^2}{t(1-t)} \end{aligned}$$

in $D[t_1, t_2]$ ($\|\cdot\|$ = Euclidean length of a vector). We record this result as the following proposition.

PROPOSITION 2.1. *Let $\{X_t\}$ be a causal $AR(p)$ process satisfying the difference equations*

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \qquad -\infty < t < \infty,$$

where $\{\varepsilon_t\}$ satisfies (1.2) *with $\sigma^2 = 1$ and $\sup_t E|\varepsilon_t|^{4+\delta} < \infty$ for some $0 < \delta \le 1$. Further assume that $\{X_t\}$ is strongly mixing with mixing function $\rho(n) \ll n^{-(1+\varepsilon)(1+4/\delta)}$ for some $\varepsilon > 0$. Then for the transformed and rescaled likelihood ratio statistic $Q_n(t) = \Lambda_n([nt])$, where $\Lambda_n(k)$ is defined in* (2.1), *we have for $0 < t_1 < t_2 < 1$,*

$$Q_n(t) \to_d \frac{\|\mathbf{W}(t) - t\mathbf{W}(1)\|^2}{t(1-t)} \tag{2.6}$$

in $D[t_1, t_2]$, where $\mathbf{W}(t)$ is standard $p+1$-dimensional Brownian motion.

REMARK 2.1. There are many sufficient conditions on the distribution of the noise in order to ensure that $\{X_t\}$ is strongly mixing. One such condition is for $\{\varepsilon_t\}$ to be iid with a common distribution function which has a nontrivial absolutely continuous component [see Athreya and Pantula (1986a, b)]. Under this condition, it can be shown, using Theorems 16.0.1 and 16.1.5 in Meyn and Tweedie (1993) with $V(x)$ equal to the quantity defined in (4.4) of Feigin and Tweedie (1985), that the mixing function $\rho(n)$ decays at a geometric rate.

REMARK 2.2. To test that a change has occurred at time $\tau \in [nt_1, nt_2]$, it is natural to consider the constrained likelihood ratio test

$$\begin{aligned}\max_{k\in[nt_1,nt_2]} \Lambda_n(k) &= \max_{t\in[t_1,t_2]} Q_n(t) \\ &\to_d \max_{t\in[t_1,t_2]} \frac{\|\mathbf{W}(t) - t\mathbf{W}(1)\|^2}{t(1-t)}.\end{aligned} \tag{2.7}$$

DeLong (1981) numerically computed the tail probabilities

$$P\left[\max_{t\in[t_1,t_2]} \frac{\|\mathbf{W}(t) - t\mathbf{W}(1)\|^2}{t(1-t)} > b\right],$$

while James, James and Siegmund (1987) obtained an accurate large deviation approximation to the tail probabilities. Assuming Gaussian white noise with known variance, Picard (1985) proposed the above constrained likelihood ratio test as well as the weighted likelihood ratio test based on the statistic

$$\max_{p<k\le n} \frac{\{\phi(k/n)[\phi(1) - \phi(k/n)]\}^2}{(k/n)(1-k/n)} \Lambda_n(k),$$

where ϕ is a suitably chosen function satisfying a regularity condition at the boundaries 0 and 1 in order to avoid technical difficulties in the behavior of Λ_n at the boundaries. In practice, however, it is unclear how t_1, t_2 and ϕ should be chosen when no prior information is available about the location of the potential changepoint. This is why we believe that the (exact) likelihood ratio test $\Lambda_n = \max_{p<k\le n} \Lambda_n(k)$ is a natural and useful alternative to Picard's tests. It is also clear that compared to the constrained likelihood ratio test, Λ_n suffers a slight loss in power if a change has occurred between nt_1 and nt_2, and in return, gains a little if a change has occurred near the boundaries [cf. Yao and Davis (1986) and James, James and Siegmund (1987)]. While it is true that no test can effectively detect a change occurring in close proximity to either of the boundaries, it is possible to detect the changepoint even when the smaller of the pre- and postchange sample sizes is small compared to the total sample size. Indeed, in the case of independent observations, Yao, Huang, and Davis (1994) showed that for each member $\hat{\tau}$ in a class of nonparametric estimators, $\hat{\tau} - \tau = O(1)$ a.s. provided that $\min(\tau, n-\tau) > C \ln n$ for some constant C.

We now consider the behavior of the (exact) likelihood ratio statistic Λ_n. Since the limit process in (2.6) is equal to $+\infty$ at the two boundaries 0 and 1, Λ_n will not converge without renormalization. As in Yao and Davis (1986) and Horváth (1993), we establish a Darling–Erdös-type limit result for Λ_n. This is the content of the following theorem whose proof is relegated to the Appendix.

THEOREM 2.2. *Let $\{X_t\}$ be the process defined in* (1.1) *such that under H_0 (no change), $\{\varepsilon_t\}$ satisfies* (1.2) *with $\sup_t E|\varepsilon_t|^{4+\delta} < \infty$ for some $0 < \delta \le 1$, and $\{X_t\}$ is strongly mixing with $\rho(n) \ll n^{-(1+\varepsilon)(1+4/\delta)}$ for some $\varepsilon > 0$. Then under H_0,*

$$P\big[\big(\sigma^{-2}\Lambda_n - b_n(p+1)\big)/a_n(p+1) \le x\big] \to \exp(-2e^{-x/2}),$$

where $\Lambda_n = \max_{p<k\le n}\Lambda_n(k)$ is the likelihood ratio statistic, $b_n(d) = (2\ln\ln n + (d/2)\ln\ln\ln n - \ln\Gamma(d/2))^2/(2\ln\ln n)$ and $a_n(d) = \sqrt{b_n(d)/(2\ln\ln n)}$ are the normalizing constants and $\Gamma(\cdot)$ is the gamma function.

REMARK 2.3. When the white noise variance σ^2 is unknown, then the likelihood ratio statistic becomes

$$\begin{aligned}\min_{p<k\le n}\left(\frac{Q_2+Q_3}{Q_1}\right)^{(n-p)/2} &= \left(1 - \frac{1}{n-p}\max_{p<k\le n}\hat\sigma^{-2}(Q_1-Q_2-Q_3)\right)^{(n-p)/2}\\ &= \left(1 - \frac{1}{n-p}\max_{p<k\le n}\hat\sigma^{-2}\Lambda_n(k)\right)^{(n-p)/2},\end{aligned}$$

where $\hat\sigma^2 = (n-p)^{-1}Q_1$. So a test based on this likelihood ratio statistic is equivalent to a test based on

$$\tilde\Lambda_n := \frac{\sigma^2}{\hat\sigma^2}\max_{p<k\le n}\sigma^{-2}\Lambda_n(k),$$

where a large value of $\tilde\Lambda_n$ indicates significance. Under H_0, $\max_{p<k\le n}\sigma^{-2}\Lambda_n(k)$ is scale invariant and hence has the same distribution as Λ_n when $\sigma^2 = 1$. Moreover, $\sigma^2/\hat\sigma^2 = 1 + O_p(n^{-1/2})$ under H_0 so that the limit null distribution of $(\tilde\Lambda_n - b_n)/a_n$ with σ^2 unknown is the same as that specified in Theorem 2.2.

3. Extensions. We now consider two extensions of the results of Section 2. The first allows for the possibility of different orders in the autoregressive models before and after the changepoint and the second permits a change in the white noise variance before and after the changepoint.

For the first extension, suppose that the process $\{X_t\}$ follows an AR(p_0) model before the change and an AR(p_1) after the change where p_0 and p_1 are known. As remarked in Section 1, we may take p_0 and p_1 to be upper bounds on the true orders of the AR models before and after the change, respectively. If $p_0 > p_1$, then the limiting null behavior of the likelihood ratio statistic will depend on whether or not the values of $\phi_{p_1+1}, \dots, \phi_{p_0}$ vanish. Since this situation is less interesting in practice, we will not pursue this case here. Of course, by taking $p = \max(p_0, p_1)$, one could still apply Theorem 2.2 directly to this case. So assuming that $p_0 < p_1$, the likelihood ratio statistic is $\max_{p_1 < k \le n} \Lambda_n(k)$, where

$$\begin{aligned}\Lambda_n(k) = {} & \min_{\boldsymbol{\phi}} \sum_{t=p_0+1}^{n} \left(X_t - \phi_0 - \phi_1 X_{t-1} - \cdots - \phi_{p_0} X_{t-p_0}\right)^2 \\ & - \min_{\boldsymbol{\phi}} \sum_{t=p_0+1}^{k} \left(X_t - \phi_0 - \phi_1 X_{t-1} - \cdots - \phi_{p_0} X_{t-p_0}\right)^2 \\ & - \min_{\boldsymbol{\alpha}} \sum_{t=k+1}^{n} \left(X_t - \alpha_0 - \alpha_1 X_{t-1} - \cdots - \alpha_{p_1} X_{t-p_1}\right)^2.\end{aligned}$$

Note that the above quantity reduces to the $\Lambda_n(k)$ introduced in Section 2 when $p_0 = p_1 = p$. The following result is an extension of Proposition 2.1 to the case $p_0 < p_1$.

PROPOSITION 3.1. *Under the assumptions of Theorem 2.2, if $p_0 < p_1$, then under H_0, for $0 < t_1 < t_2 < 1$, we have*

$$\begin{aligned}\sigma^{-2} \max_{[nt_1] \le k \le [nt_2]} \Lambda_n(k) \to_d \max_{t_1 \le t \le t_2} \Bigg\{ & \frac{\|\mathbf{W}_1(t)\|^2}{t} + \frac{\|\mathbf{W}_1(1) - \mathbf{W}_1(t)\|^2}{1-t} \\ & - \|\mathbf{W}_1(1)\|^2 + \frac{\|\mathbf{W}_2(1) - \mathbf{W}_2(t)\|^2}{1-t} \Bigg\},\end{aligned}$$

where $\mathbf{W}_1(t)$ and $\mathbf{W}_2(t)$ are two independent Brownian motions of dimensions $p_0 + 1$ and $p_1 - p_0$, respectively. [*The scale factor σ^{-2} may be replaced by $\hat{\sigma}^{-2} = (n - p_0)/Q_1$; see Remark* 2.3.]

The proofs of all of the results in this section including the foregoing proposition are postponed to the Appendix.

The limiting behavior of the unconstrained likelihood ratio statistic for the case $p_0 < p_1$ is described in the following proposition.

PROPOSITION 3.2. *Under the assumptions of Theorem 2.2, if $p_0 < p_1$, then under H_0,*

$$P\left[\left(\hat{\sigma}^{-2} \max_{p_1 < k \le n} \Lambda_n(k) - b_n(p_1 + 1)\right) / a_n(p_1 + 1) \le x\right] \to \exp(-e^{-x/2}),$$

where $\hat{\sigma}^2$ is the estimate of σ^2 given in Remark 2.3.

We now turn to the problem when the white noise variance is allowed to be different before and after the changepoint. In this case the model becomes

$$
\begin{aligned}
X_t &= \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_{p_0} X_{t-p_0} + \varepsilon_t, \qquad t \le \tau, \\
&= \alpha_0 + \alpha_1 X_{t-1} + \cdots + \alpha_{p_1} X_{t-p_1} + \varepsilon_t, \qquad t \ge \tau + 1,
\end{aligned} \tag{3.1}
$$

where $\tau \in (p_1, n]$, $\{\varepsilon_t,\ t \le \tau\}$ is fourth-order white noise with mean 0 and variance σ_0^2, $\{\varepsilon_t,\ t > \tau\}$ is fourth-order white noise with mean 0 and variance σ_1^2 and $p_0 \le p_1$. After taking $-2\ln$, the Gaussian likelihood ratio for testing $\tau = n$ versus $\tau = k$, conditional on the first p_0 observations, is given by

$$
\Lambda'_n(k) = (n - p_0)\ln \hat{\sigma}^2 - (k - p_0)\ln \hat{\sigma}_0^2(k) - (n - k)\ln \hat{\sigma}_1^2(k), \tag{3.2}
$$

where $\hat{\sigma}^2 = Q_1/(n - p_0)$, $\hat{\sigma}_0^2(k) = Q_2/(k - p_0)$, $\hat{\sigma}_1^2(k) = Q_3/(n - k)$,

$$
Q_1 = \min_{\boldsymbol{\phi}} \sum_{t=p_0+1}^{n} \left(X_t - \phi_0 - \phi_1 X_{t-1} - \cdots - \phi_{p_0} X_{t-p_0}\right)^2,
$$

$$
Q_2 = \min_{\boldsymbol{\phi}} \sum_{t=p_0+1}^{k} \left(X_t - \phi_0 - \phi_1 X_{t-1} - \cdots - \phi_{p_0} X_{t-p_0}\right)^2
$$

and

$$
Q_3 = \min_{\boldsymbol{\alpha}} \sum_{t=k+1}^{n} \left(X_t - \alpha_0 - \alpha_1 X_{t-1} - \cdots - \alpha_{p_1} X_{t-p_1}\right)^2.
$$

THEOREM 3.3. *Let $\{X_t\}$ be the process defined in* (3.1) *such that under H_0, $\{X_t\}$ and $\{\varepsilon_t\}$ satisfy the assumptions specified in Theorem* 2.2. *In addition, assume that*

$$
\mu_3 := E(\varepsilon_t^3) = 0 \quad \textit{and} \quad \mu_4 := E(\varepsilon_t^4) = 3\sigma^4 \ (= 3\sigma_0^4)
$$

[*i.e., the first four moments of ε_t match those of a $N(0, \sigma^2)$ random variable*]. *Set* $\Lambda'_n := \max\{\Lambda'_n(k)\colon (2p_0 + 1) \vee p_1 < k \le n - p_1 - 2\}$.

(a) *If $p_0 = p_1 = p$, then under H_0,*

$$
P\big[(\Lambda'_n - b_n(p + 2))/a_n(p + 2) \le x\big] \to \exp(-2e^{-x/2}).
$$

(b) *If $p_0 < p_1$, then under H_0,*

$$
P\big[(\Lambda'_n - b_n(p_1 + 2))/a_n(p_1 + 2) \le x\big] \to \exp(-e^{-x/2}).
$$

Clearly, Theorem 3.3 applies to the case when $\{\varepsilon_t\}$ is Gaussian white noise. A modified test statistic may be formulated in case $\mu_3 = 0$, but $\mu_4 \ne 3\sigma^4$. Using a Taylor series expansion and discarding the asymptotically negligible terms, a rough approximation to $\Lambda'_n(k)$ is given by

$$
\sigma^{-2}(Q_1 - Q_2 - Q_3) + \frac{(k - p_0)(n - k)}{(n - p_0)(2\sigma^4)}\left(\hat{\sigma}_0^2(k) - \hat{\sigma}_1^2(k)\right)^2.
$$

Now the limit distribution of this quantity depends on μ_4/σ^4, which can be eliminated by replacing $2\sigma^4$ by $\mu_4 - \sigma^4$. This leads to the test statistic given by $\Lambda_n^* = \max\{\Lambda_n^*(k): \max(2p_0+1, p_1) < k \le n - p_1 - 2\}$, where

$$\Lambda_n^*(k) = \hat{\sigma}^{-2}(Q_1 - Q_2 - Q_3) \tag{3.3}$$
$$+ \frac{(k-p_0)(n-k)}{(n-p_0)(R_n - \hat{\sigma}^4)}\left(\hat{\sigma}_0^2(k) - \hat{\sigma}_1^2(k)\right)^2$$

and $\hat{\sigma}^2 = Q_1/(n-p_0)$ and R_n is any estimate of μ_4 satisfying $R_n - \mu_4 = O_p(1/\sqrt{n})$.

THEOREM 3.4. *Let $\{X_t\}$ be the process defined in* (3.1) *such that under H_0, $\{X_t\}$ and $\{\varepsilon_t\}$ satisfy the assumptions specified in Theorem* 2.2. *Furthermore, assume that $\mu_3 := E(\varepsilon_t^3) = 0$. Then, under H_0,*

$$P\left[(\Lambda_n^* - b_n(p+2))/a_n(p+2) \le x\right] \to \exp(-2e^{-x/2}) \quad \textit{if } p_0 = p_1 = p,$$
$$P\left[(\Lambda_n^* - b_n(p_1+2))/a_n(p_1+2) \le x\right] \to \exp(-e^{-x/2}) \quad \textit{if } p_0 < p_1.$$

REMARK 3.1. A constrained version of Λ_n^* is $\max_{[nt_1] \le k \le [nt_2]} \Lambda_n^*(k)$ with $0 < t_1 < t_2 < 1$. It is not difficult to show that under H_0, this constrained test converges in distribution to

$$\max_{t_1 \le t \le t_2} \left\{ \frac{\|\mathbf{W}(t)\|^2}{t} + \frac{\|\mathbf{W}(1) - \mathbf{W}(t)\|^2}{1-t} - \|\mathbf{W}(1)\|^2 \right\} \quad \text{if } p_0 = p_1 = p,$$

$$\max_{t_1 \le t \le t_2} \left\{ \frac{\|\mathbf{W}_1(t)\|^2}{t} + \frac{\|\mathbf{W}_1(1) - \mathbf{W}_1(t)\|^2}{1-t} - \|\mathbf{W}_1(1)\|^2 + \frac{\|\mathbf{W}_2(1) - \mathbf{W}_2(t)\|^2}{1-t} \right\}$$
$$\text{if } p_0 < p_1,$$

where $\mathbf{W}(t)$ is standard $p+2$-dimensional Brownian motion and $\mathbf{W}_1(t)$ and $\mathbf{W}_2(t)$ are two independent standard Brownian motions of dimensions $p_0 + 2$ and $p_1 - p_0$, respectively.

4. Simulation results. A small simulation study was conducted to compare the approximation of the limit distribution to the null distribution of some of the test statistics discussed in Sections 2 and 3. The three test statistics considered were (assuming $p_0 = p_1 = p$)

$$\Lambda(1) = \hat{\sigma}^{-2} \max_{p<k\le n} \Lambda_n(k), \tag{4.1}$$

$$\Lambda(2) = \max_{2p+1<k\le n-p-2} \Lambda_n'(k), \tag{4.2}$$

$$\Lambda(3) = \max_{2p+1<k\le n-p-2} \Lambda_n^*(k), \tag{4.3}$$

where $\Lambda_n(k)$, $\Lambda_n'(k)$ and $\Lambda_n^*(k)$ are defined in (2.1), (3.2) and (3.3), respectively, and $\hat{\sigma}^2 = Q_1/(n-p)$ (see Remark 2.3). The test statistics in (4.1)–(4.3) correspond to slightly different situations. The test based on $\Lambda(1)$ was derived under the assumption of no change in the white noise variance after the changepoint, whereas the tests based on $\Lambda(2)$ and $\Lambda(3)$ permitted a change in the variance after the changepoint. The test statistic $\Lambda(2)$ assumes that the first four moments of the noise are the same as those of a $N(0, \sigma^2)$ random variable, while $\Lambda(1)$ and $\Lambda(3)$ are more distribution-free tests with $\Lambda(3)$ only requiring a zero third moment.

For the simulation study we took the noise $\{\varepsilon_t\}$ to be iid $N(0,1)$, $p_0 = p_1 = 1$, $\phi_0 = 0$ and used parameter values $\phi = -0.9, -0.5, 0, 0.5, 0.9$ and sample sizes $n = 25, 50, 100, 200, 500$. For each combination, we computed 10,000 replicates of the statistics $\Lambda(1)$, $\Lambda(2)$ and $\Lambda(3)$. The empirical distributions of $a_n^{-1}(\Lambda(1) - b_n)$ and $a_n^{-1}(\Lambda(3) - b_n)$, where the normalizing constants a_n and b_n are as specified in Theorems 2.2 and 3.4, respectively, were plotted together with the limit distribution $\exp(-2e^{-x/2})$ for $\phi_0 = 0$, $\phi_1 = 0.5$ and $n = 25, 50, 100, 200, 500$ (see Figures 1 and 2). From these figures, one clearly sees the convergence of the sampling distributions to the limit distribution as the sample size increases. Also, note that the limit distribution provides a reasonably good approximation for all values of x and is particularly good for values of $x > 6$. The approximations are slightly better for small values of x when the σ_0^2 or ϕ_0 is assumed to be known. Tables 1 and 2 provide (empirical) type I errors for the test statistics $\Lambda(1)$ and $\Lambda(3)$ using a cutoff value determined from the limit distribution. In other words, the

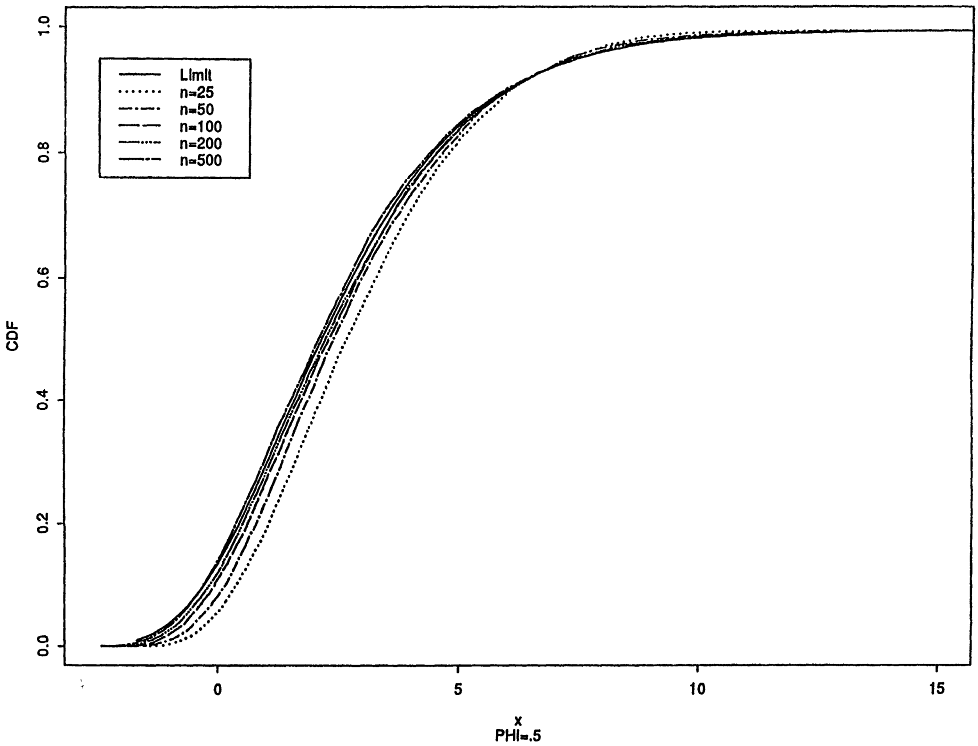

FIG. 1. *Limit approximation to test statistic* (4.1).

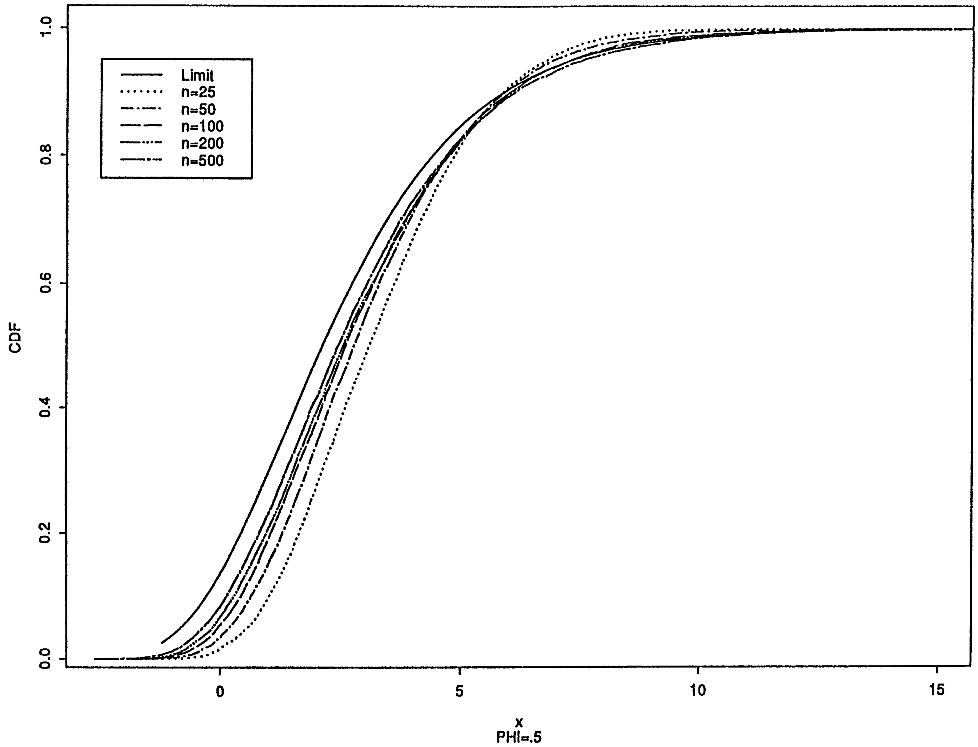

FIG. 2. *Limit approximation to test statistic* (4.3).

empirical probabilities of $P[a_n^{-1}(\Lambda(1) - b_n) > x_{1-\alpha}]$ and $P[a_n^{-1}(\Lambda(3) - b_n) > x_{1-\alpha}]$ were computed for $n = 100$, $\phi_0 = 0$ and $\phi_1 = 0, 0.5$, where $x_{1-\alpha} = -2\ln(-0.5\ln(1 - \alpha))$ is the $1 - \alpha$ quantile of $\exp(-2e^{-x/2})$.

We did not include plots of the empirical distribution of $a_n^{-1}(\Lambda(2) - b_n)$ since the limit approximation tended to be considerably larger than the empirical distribution. This is in part due to the fact that $\hat{\sigma}_k^2$ ($\hat{\sigma}_{n-k}^2$) may occasionally be very small for small k (large k), resulting in a (negatively) large value of $\ln \hat{\sigma}_k^2$ ($\ln \hat{\sigma}_{n-k}^2$). The approximation did become noticeably

TABLE 1
Simulated type I errors using tests based on $\Lambda(1)$ *and* $\Lambda(3)$ *with cutoff values* $x_{1-\alpha} = -2\ln(-0.5\ln(1 - \alpha))$ (*here* $n = 100$, $\phi_0 = 0$, $\phi_1 = 0$)

	α		
Test Statistic	**0.1**	**0.05**	**0.01**
$\Lambda(1)$	0.066	0.028	0.005
$\Lambda(3)$	0.081	0.034	0.006

TABLE 2
Simulated type I errors using tests based on $\Lambda(1)$ *and* $\Lambda(3)$ *with cutoff values* $x_{1-\alpha} = -2\ln(-0.5\ln(1-\alpha))$ *(here* $n = 100$, $\phi_0 = 0$, $\phi_1 = 0.5$*)*

	α		
Test	**0.1**	**0.05**	**0.01**
$\Lambda(1)$	0.103	0.050	0.008
$\Lambda(3)$	0.107	0.049	0.008

better as we restricted the values of k from the two ends, 1 and n, in the computation of $\Lambda(2)$.

Figure 3 compares the empirical distributions of $a_n^{-1}(\Lambda(3) - b_n)$ for $n = 100$, $\phi_0 = 0$, $\phi_1 = -0.9, -0.5, 0, 0.5, 0.9$ with the limit distribution. [The corresponding plots for $\Lambda(1)$ are nearly the same.] Note that the quality of the limit approximation is reasonably good except when $\phi_1 = 0.9$. At first glance, it might seem surprising that the two cases $\phi_1 = -0.9$ and $\phi_1 = 0.9$ produce such disparate results. However, this phenomenon may be attributable to the

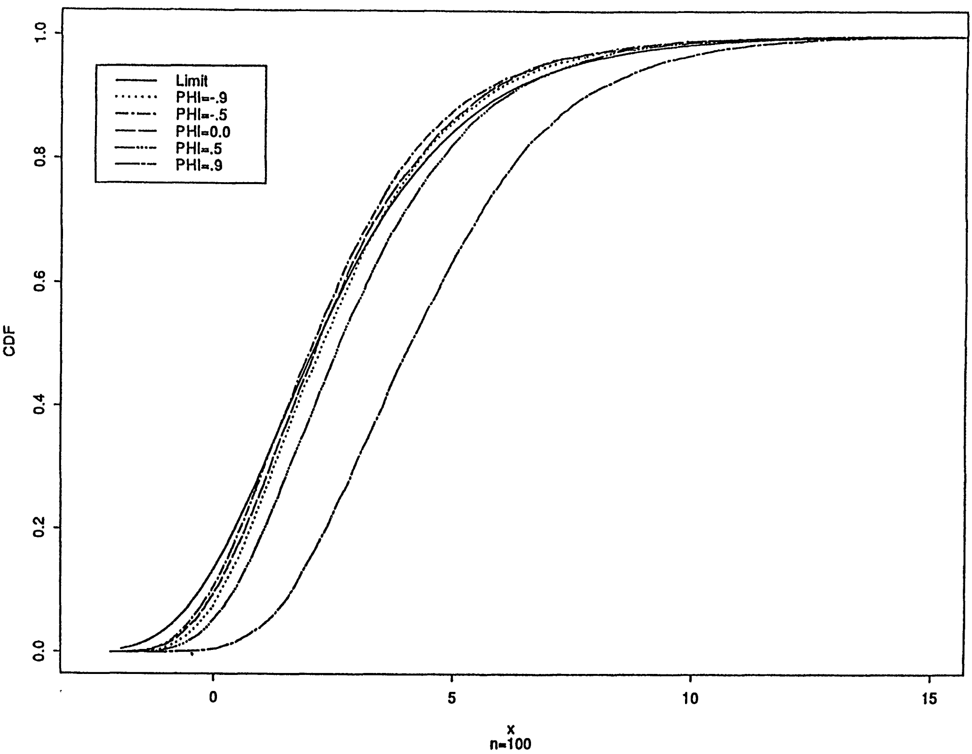

FIG. 3. *Limit approximation to test statistic* (4.3).

large variance of $\hat{\mu}$, the estimate of the mean of the process, when ϕ_1 is close to 1. [For example, under H_0, $\hat{\mu}$ is asymptotically normal with mean μ and variance $\sigma^2/(n(1-\phi_1)^2)$.] Assuming that ϕ_0 is known to be 0, the difficulty in estimating the mean disappears and the sampling distributions of $a_n^{-1}(\Lambda(3) - b_n)$, with $\Lambda(3)$ properly modified, no longer exhibit this asymmetry for different signs of ϕ_1 (see Figure 4) .

It is well known [Hall (1979)] that the maximum of an iid sequence of normal random variables converges to the double exponential distribution at a very slow rate. Our simulation study, however, shows that for moderate sample sizes, the double exponential distribution provides a reasonable approximation for the null distributions of the statistics $\Lambda(1)$ and $\Lambda(3)$. Since the test statistics $\Lambda(1)$ and $\Lambda(3)$ are asymptotically distribution-free, the double exponential distribution is a valid limit for a variety of noise distributions. In the special case that $n = 100$, $\phi_0 = 0$, $\phi_1 = 0.5$ and the noise is iid with a Laplace distribution, we found that the limit distribution is still a reasonable approximation to the empirical distribution of $a_n^{-1}(\Lambda(3) - b_n)$.

In previous work, Yao and Davis (1986) and Horváth (1993) considered the limiting distribution of the square root of the likelihood ratio rather than the

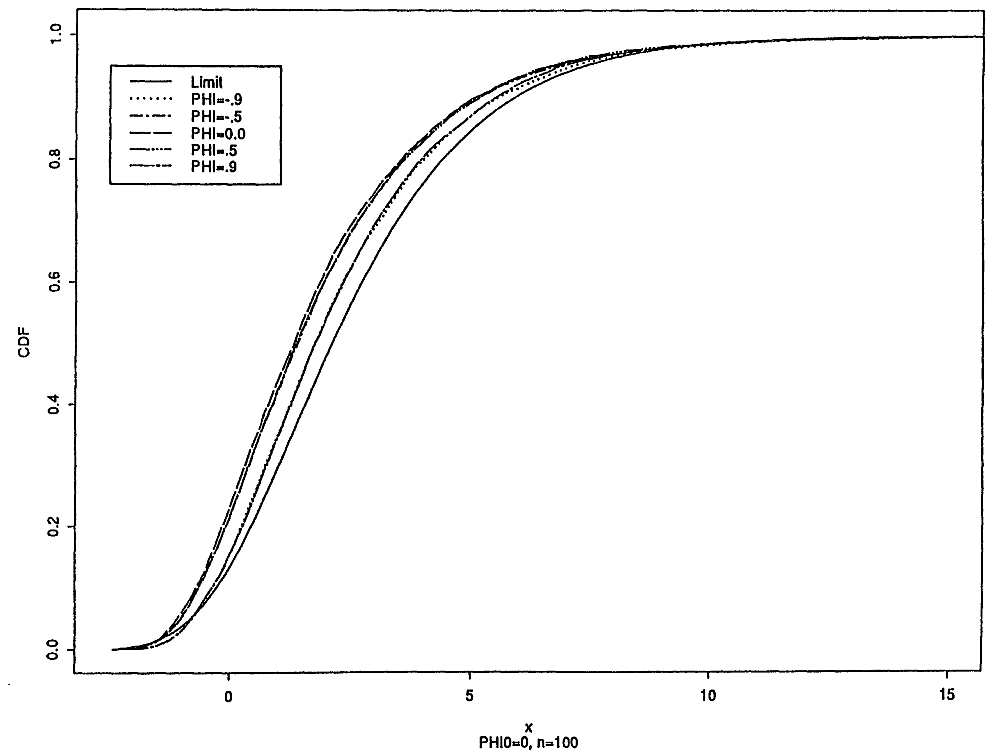

FIG. 4. *Limit approximation to test statistic* (4.3).

likelihood ratio itself. Using Proposition A.1, it is easy to show that

$$P\left[2\alpha(\ln n)\Lambda(1)^{1/2} - 2\beta_2(\ln n) \le x\right] \to \exp(-2e^{-x/2}).$$

In Figure 5, the empirical quantiles of $2\alpha(\ln n)\Lambda(1)^{1/2} - 2\beta_2(\ln n)$ and $a_n^{-1}(\Lambda(1) - b_n)$ are plotted versus the $i/10{,}000$ ($i = 1, \dots, 10{,}000$) quantiles of the limit distribution $\exp\{-2e^{-x/2}\}$ for the case $n = 100$, $\phi_0 = 0$ and $\phi_1 = 0.5$. From these Q-Q plots, it is clear that the graph corresponding to $\Lambda(1)$ is surprisingly linear for values less than 15 (the plotted line in Figure 5 has slope 1 and intercept 0). On the other hand, the graph corresponding to $\Lambda(1)^{1/2}$ exhibits a strong nonlinear shape. This suggests that the limit distribution provides a better overall approximation to the distribution of $\Lambda(1)$ than for $\Lambda(1)^{1/2}$. While the corresponding Q-Q plots for $\Lambda(1)$ were nearly linear for all of our simulations, the best fitting line to the Q-Q plot did not always have slope near 1 and intercept 0. It therefore may be possible to improve the approximation of the limit distribution by a more judicious choice of normalizing constants a_n and b_n. This will be the subject of future study.

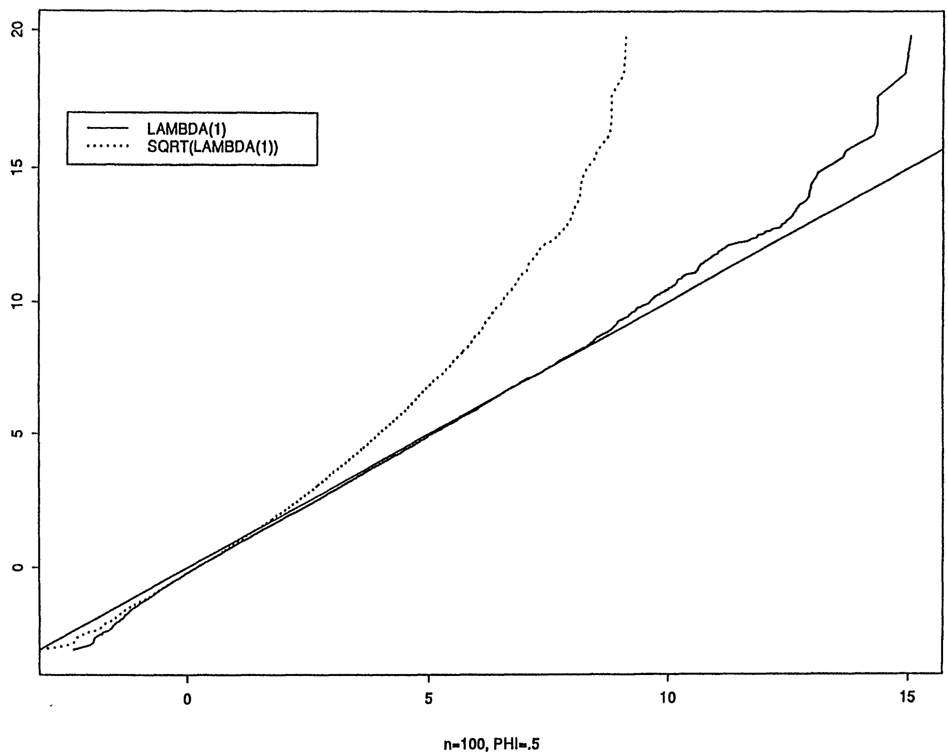

FIG. 5. *Q-Q plot.*

APPENDIX

In this Appendix we provide proofs of the main results in Sections 2 and 3. We begin with Horváth's extension of Darling–Erdös' result to random vectors.

PROPOSITION A.1 [Lemma 2.2 of Horváth (1993)]. *Let $\{\mathbf{Z}_t,\ t = 1, 2, \ldots\}$ be an iid sequence of d-dimensional random vectors with $E\mathbf{Z}_1 = \mathbf{0}$ and $E(\mathbf{Z}_1\mathbf{Z}_1') = I$. If the components of $\mathbf{Z}_1$ all have finite rth moments for some $r > 2$, then*

$$P\left[\alpha(\ln n) \max_{1 \le k \le n} \frac{\|\mathbf{S}_k\|}{k^{1/2}} - \beta_d(\ln n) \le x\right] \to \exp(-e^{-x})$$

as $n \to \infty$, where $\mathbf{S}_k = \sum_{j=1}^k \mathbf{Z}_j$, $\alpha(x) = (2 \ln x)^{1/2}$, $\beta_d(x) = 2 \ln x + (d/2)\ln_2 x - \ln \Gamma(d/2)$, $\ln_k$ is the kth iterated logarithm and Γ is the gamma function.

COROLLARY A.2. *Under the assumptions of Proposition* A.1,

$$P\left[\left(\max_{1 \le k \le n} \frac{\|\mathbf{S}_k\|^2}{k} - B_d(\ln n)\right)\Big/A_d(\ln n) \le x\right] \to \exp(-e^{-x/2}),$$

where $A_d(x) = \beta_d(x)/\alpha^2(x)$ and $B_d(x) = (\beta_d(x)/\alpha(x))^2$. Note that $A_d(x)$ and $B_d(x)$ may be replaced by 1 *and* $2 \ln x + d \ln_2 x - 2 \ln \Gamma(d/2)$.

PROOF OF THEOREM 2.2. Since the distribution of $\sigma^{-2}\Lambda_n$ does not depend on σ^2, we assume throughout the proof that $\sigma^2 = 1$. The main idea of the proof is to show that for all $\delta' > 0$,

$$\limsup_{n \to \infty} P\left[\left|\max_{p < k \le \varepsilon' n} \Lambda_n(k) - \max_{p < k \le \varepsilon' n} \mathbf{S}_k' P_k \mathbf{S}_k\right|\Big/a_n > \delta'\right] \to 0 \quad \text{as } \varepsilon' \to 0, \tag{A.1}$$

$$\begin{aligned}\limsup_{n \to \infty} P\Bigg[\Bigg|&\max_{(1-\varepsilon')n < k \le n} \Lambda_n(k) \\ &- \max_{(1-\varepsilon')n < k \le n} (\mathbf{S}_n - \mathbf{S}_k)'\tilde{P}_k(\mathbf{S}_n - \mathbf{S}_k)\Bigg|\Big/a_n > \delta'\Bigg] \to 0 \quad \text{as } \varepsilon' \to 0,\end{aligned} \tag{A.2}$$

and for all $\varepsilon' > 0$,

$$P\left[\left(\max_{p < k \le \varepsilon' n} \mathbf{S}_k' P_k \mathbf{S}_k - b_n\right)\Big/a_n \le x\right] \to \exp(-e^{-x/2}), \tag{A.3}$$

$$\begin{aligned}P\Bigg[\Bigg(&\max_{(1-\varepsilon')n < k \le n} (\mathbf{S}_n - \mathbf{S}_k)'\tilde{P}_k(\mathbf{S}_n - \mathbf{S}_k) - b_n\Bigg)\Big/a_n \le x\Bigg] \\ &\to \exp(-e^{-x/2}),\end{aligned} \tag{A.4}$$

where $a_n = a_n(p + 1)$ and $b_n = b_n(p + 1)$.

Once we establish the validity of (A.1)–(A.4), then the remainder of the proof is easy to complete. First the strong mixing assumption implies that $\max_{p<k\le\varepsilon'n}\mathbf{S}_k'P_k\mathbf{S}_k$ and $\max_{(1-\varepsilon')n<k\le n}(\mathbf{S}_n-\mathbf{S}_k)'\tilde{P}_k(\mathbf{S}_n-\mathbf{S}_k)$ are asymptotically independent (for $\varepsilon'<1/2$) and hence from (A.3) and (A.4),

$$P\Big[\Big(\max_{p<k\le\varepsilon'n}\mathbf{S}_k'P_k\mathbf{S}_k-b_n\Big)/a_n\le x,$$
$$\Big(\max_{(1-\varepsilon')n<k\le n}(\mathbf{S}_n-\mathbf{S}_k)'\tilde{P}_k(\mathbf{S}_n-\mathbf{S}_k)-b_n\Big)/a_n\le x\Big]\to\exp(-2e^{-x/2}).$$

Also, by (2.7),

$$\Big(\max_{\varepsilon'n\le k\le(1-\varepsilon')n}\Lambda_n(k)-b_n\Big)/a_n\to_P-\infty.$$

Now appealing to Slutsky's theorem and Theorem 4.2 in Billingsley (1968), the conclusion of the theorem is immediate.

To establish (A.1) observe that

$$\begin{aligned}
&\frac{\Big|\max_{p<k\le\varepsilon'n}\Lambda_n(k)-\max_{p<k\le\varepsilon'n}\mathbf{S}_k'P_k\mathbf{S}_k\Big|}{a_n}\\
&\quad\le\max_{p<k\le\varepsilon'n}\frac{|\Lambda_n(k)-\mathbf{S}_k'P_k\mathbf{S}_k|}{a_n}\\
&\quad=\max_{p<k\le\varepsilon'n}\frac{\big|(\mathbf{S}_n-\mathbf{S}_k)'\tilde{P}_k(\mathbf{S}_n-\mathbf{S}_k)-\mathbf{S}_n'P_n\mathbf{S}_n\big|}{a_n}\\
&\quad\to_d\max_{t\le\varepsilon'}\left|\frac{\|\mathbf{W}(1)-\mathbf{W}(t)\|^2}{1-t}-\|\mathbf{W}(1)\|^2\right|\quad(\text{as } n\to\infty)\\
&\quad\to 0\quad\text{a.s. as }\varepsilon'\to 0,
\end{aligned}$$

since $a_n\to 1$. This proves (A.1) and (A.2) can be handled in a similar fashion.

The proof of (A.3) will be broken up into several steps.

STEP 1. For the sequence $\{\mathbf{U}_k\}$ given in (2.2),

$$\mathbf{S}_k'\Gamma_{p+1}^{-1}\mathbf{S}_k-\mathbf{U}_k'\Gamma_{p+1}^{-1}\mathbf{U}_k=O(k^{1-\lambda'})$$

as $k\to\infty$ a.s. for some $\lambda'>0$.

PROOF. The difference is equal to

$$\mathbf{S}_k'\Gamma_{p+1}^{-1}(\mathbf{S}_k-\mathbf{U}_k)+(\mathbf{S}_k'-\mathbf{U}_k')\Gamma_{p+1}^{-1}\mathbf{U}_k\tag{A.5}$$

and since by the law of the iterated logarithm, $\mathbf{U}_k'\Gamma_{p+1}^{-1}=O((k\ln_2 k)^{1/2})$, we have $\mathbf{S}_k'\Gamma_{p+1}^{-1}=O((k\ln_2 k)^{1/2})$ from (2.2). Step 1 now follows from (A.5) and (2.2).

STEP 2. We have

$$\mathbf{S}_k' P_k \mathbf{S}_k - \mathbf{U}_k' \Gamma_{p+1}^{-1} \mathbf{U}_k / k \to 0 \quad \text{as } k \to \infty \text{ a.s.}$$

PROOF. From Step 1, it suffices to show that

$$\mathbf{S}_k' P_k \mathbf{S}_k - \frac{\mathbf{S}_k' \Gamma_{p+1}^{-1} \mathbf{S}_k}{k} = \frac{\mathbf{S}_k'}{k^{1/2}} k P_k \left(\Gamma_{p+1} - \frac{P_k^{-1}}{k} \right) \Gamma_{p+1}^{-1} \frac{\mathbf{S}_k}{k^{1/2}} \to 0 \quad \text{a.s.} \tag{A.6}$$

However, by (2.4) and the law of the iterated logarithm, $\Gamma_{p+1} - P_k^{-1}/k = O((\ln_2 k/k)^{1/2})$ as $k \to \infty$ and hence, the expression in (A.6) converges to 0 as $k \to \infty$ a.s.

STEP 3. For all $\varepsilon' > 0$, as $n \to \infty$,

$$P\left[\max_{p < k \le n} \mathbf{S}_k' P_k \mathbf{S}_k = \max_{p < k \le \varepsilon' n} \mathbf{S}_k' P_k \mathbf{S}_k \right] \to 1. \tag{A.7}$$

PROOF. From the proof of Proposition 2.1,

$$\max_{\varepsilon' n \le k \le n} \mathbf{S}_k' P_k \mathbf{S}_k \to_d \max_{t \in [\varepsilon', 1]} \frac{\| \mathbf{W}(t) \|^2}{t}.$$

Now (A.7) follows from the fact that $\max_{\varepsilon' n \le k \le n} \mathbf{S}_k' P_k \mathbf{S}_k = O_p(1)$ and $\max_{p < k \le \varepsilon' n} \mathbf{S}_k' P_k \mathbf{S}_k \to_P \infty$.

To finish the proof of (A.3), we see that Step 2 implies

$$\begin{aligned} &\left| \max_{M \le k \le n} \mathbf{S}_k' P_k \mathbf{S}_k - \max_{M \le k \le n} \mathbf{U}_k' \Gamma_{p+1}^{-1} \mathbf{U}_k / k \right| \\ &\quad \le \sup_{k \ge M} |\mathbf{S}_k' P_k \mathbf{S}_k - \mathbf{U}_k' \Gamma_{p+1}^{-1} \mathbf{U}_k / k| \to 0 \end{aligned} \tag{A.8}$$

as $M \to \infty$ a.s. Since the random vectors $\Gamma_{p+1}^{-1/2} \mathbf{Z}_k$ satisfy the assumptions of Corollary A.2, it follows that

$$P\left[\left(\max_{p < k \le n} \mathbf{U}_k' \Gamma_{p+1}^{-1} \mathbf{U}_k / k - b_n \right) / a_n \le x \right] \to \exp(-e^{-x/2}).$$

This relation together with (A.7) and (A.8) proves (A.3).

The argument for (A.4) is similar. One merely follows the same steps above but applied to the reverse-time process, which also satisfies the strong mixing condition. □

PROOF OF PROPOSITION 3.1. Following the same argument as used in the proof of Proposition 2.1, it is not difficult to show that

$$\sigma^{-2} \Lambda_n([nt]) \to_d L(t)$$

in $D[t_1, t_2]$, where

$$L(t) = \frac{\mathbf{W}_1^*(t)'\Gamma_{p_0+1}^{-1}\mathbf{W}_1^*(t)}{t} + \frac{(\mathbf{W}^*(1) - \mathbf{W}^*(t))'\Gamma_{p_1+1}^{-1}(\mathbf{W}^*(1) - \mathbf{W}^*(t))}{1-t} - \mathbf{W}_1^*(1)'\Gamma_{p_0+1}^{-1}\mathbf{W}_1^*(1)$$

and $\mathbf{W}^*(t)$ is $p_1 + 1$-dimensional Brownian motion with covariance matrix Γ_{p_1+1} and $\mathbf{W}_1^*(t)$ is the vector consisting of the first $p_0 + 1$ components of $\mathbf{W}^*(t)$. Note that Γ_{p_0+1} is the upper left $(p_0 + 1) \times (p_0 + 1)$ submatrix of Γ_{p_1+1}. Let $\tilde{\mathbf{W}}(t) = \Gamma_{p_1+1}^{-1/2}\mathbf{W}^*(t)$, so that $\tilde{\mathbf{W}}(t)$ is standard $(p_1 + 1)$-dimensional Brownian motion and

$$\begin{aligned} L(t) &= \mathbf{W}^*(t)'\begin{bmatrix} \Gamma_{p_0+1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\frac{\mathbf{W}^*(t)}{t} \\ &\quad + \frac{(\mathbf{W}^*(1) - \mathbf{W}^*(t))'\Gamma_{p_1+1}^{-1}(\mathbf{W}^*(1) - \mathbf{W}^*(t))}{1-t} \\ &\quad - \mathbf{W}^*(1)'\begin{bmatrix} \Gamma_{p_0+1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{W}^*(1) \\ &= \frac{\tilde{\mathbf{W}}(t)'\tilde{\Gamma}\tilde{\mathbf{W}}(t)}{t} + \frac{\|\tilde{\mathbf{W}}(1) - \tilde{\mathbf{W}}(t)\|^2}{1-t} - \tilde{\mathbf{W}}(1)'\tilde{\Gamma}\tilde{\mathbf{W}}(1), \end{aligned}$$

where

$$\tilde{\Gamma} = \Gamma_{p_1+1}^{1/2}\begin{bmatrix} \Gamma_{p_0+1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\Gamma_{p_1+1}^{1/2}.$$

It is easily checked that $\tilde{\Gamma}$ is symmetric and idempotent of rank $p_0 + 1$, so that there exists an orthogonal matrix U such that

$$\tilde{\Gamma} = U'\begin{bmatrix} I_{p_0+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}U.$$

Let $\mathbf{W}(t) = U\tilde{\mathbf{W}}(t)$, which remains standard $p_1 + 1$-dimensional Brownian motion. It follows that

$$L(t) = \frac{\|\mathbf{W}_1(t)\|^2}{t} + \frac{\|\mathbf{W}_1(1) - \mathbf{W}_1(t)\|^2 + \|\mathbf{W}_2(1) - \mathbf{W}_2(t)\|^2}{1-t} - \|\mathbf{W}_1(1)\|^2,$$

where $\mathbf{W}(t) = \begin{bmatrix} \mathbf{W}_1(t) \\ \mathbf{W}_2(t) \end{bmatrix}$ and $\mathbf{W}_1(t)$ and $\mathbf{W}_2(t)$ are two independent standard Brownian motions of dimensions $p_0 + 1$ and $p_1 - p_0$, respectively. This completes the proof. □

PROOF OF PROPOSITION 3.2. It suffices to show that under H_0,

$$P\left[\max_{p_1<k\le n/2}\Lambda_n(k) < \max_{n/2<k\le n}\Lambda_n(k)\right] \to 1$$

and

$$P\left[\left(\sigma^{-2}\max_{n/2<k\le n}\Lambda_n(k)-b_n(p_1+1)\right)/a_n(p_1+1)\le x\right]\to\exp(-e^{-x/2}).$$

However, these relations follow from the proofs of Theorem 2.2 and Proposition 3.1. □

PROOF OF THEOREM 3.3. (a) We shall us the notation already developed in Section 2. Furthermore, since

$$\Lambda'_n(k)=(n-p_0)\ln\frac{\hat\sigma^2}{\sigma^2}-(k-p_0)\ln\frac{\hat\sigma_0^2(k)}{\sigma^2}-(n-k)\ln\frac{\hat\sigma_1^2(k)}{\sigma^2},$$

we may assume, without loss of generality, that $\sigma^2=1$. Applying the law of the iterated logarithm to $\mathbf{U}_k$ in (2.2), we have $\|\mathbf{S}_k\|^2=O(k\ln_2 k)$ a.s. and hence $Q_2=\boldsymbol\varepsilon'_k\boldsymbol\varepsilon_k+O(\ln_2 k)$ a.s. Also, since $Q_1=\boldsymbol\varepsilon'_n\boldsymbol\varepsilon_n+O_p(1)$ and

$$\max_{p<k\le n/2}|Q_3-\tilde{\boldsymbol\varepsilon}'_k\tilde{\boldsymbol\varepsilon}_k|=O_p(1),$$

it follows that

$$\hat\sigma^2=\frac{1}{n-p}\boldsymbol\varepsilon'_n\boldsymbol\varepsilon_n+O_p\left(\frac1n\right),\tag{A.9}$$

$$\hat\sigma_0^2(k)=\frac{1}{k-p}\boldsymbol\varepsilon'_k\boldsymbol\varepsilon_k+O\left(\frac1k\ln_2 k\right)\quad\text{a.s.}\tag{A.10}$$

and

$$\max_{p<k\le n/2}\left|\hat\sigma_1^2(k)-\frac{1}{n-k}\tilde{\boldsymbol\varepsilon}'_k\tilde{\boldsymbol\varepsilon}_k\right|=O_p\left(\frac1n\right).\tag{A.11}$$

Let $T_k=\boldsymbol\varepsilon'_k\boldsymbol\varepsilon_k-(k-p)$ so that by the law of the iterated logarithm, $T_k=O(\sqrt{k\ln_2 k})$ a.s. From (A.9)–(A.11), we conclude that

$$\hat\sigma^2-1=O_p\left(\frac{1}{\sqrt n}\right),\qquad\hat\sigma_0^2(k)-1=O\left(\sqrt{\frac{\ln_2 k}{k}}\right)\quad\text{a.s.}$$

and

$$\max_{p<k\le n/2}|\hat\sigma_1^2(k)-1|=O_p\left(\frac{1}{\sqrt n}\right).$$

A Taylor series expansion yields

$$\begin{aligned}\Lambda'_n(k)&=(n-p)(\hat\sigma^2-1)-(k-p)(\hat\sigma_0^2(k)-1)-(n-k)(\hat\sigma_1^2(k)-1)\\&\quad-\frac{n-p}{2}(\hat\sigma^2-1)^2+\frac{k-p}{2}(\hat\sigma_0^2(k)-1)^2+\frac{n-k}{2}(\hat\sigma_1^2(k)-1)^2\\&\quad+\frac{n-p}{3}(\hat\sigma^2-1)^3(1+\eta)^{-3}-\frac{k-p}{3}(\hat\sigma_0^2(k)-1)^3(1+\eta_0)^{-3}\\&\quad-\frac{n-k}{3}(\hat\sigma_1^2(k)-1)^3(1+\eta_1)^{-3},\end{aligned}$$

where $|\eta| \le |\hat\sigma^2 - 1|$, $|\eta_0| \le |\hat\sigma_0^2(k) - 1|$ and $|\eta_1| \le |\hat\sigma_1^2(k) - 1|$. Clearly,

$$(n-p)(\hat\sigma^2 - 1) - (k-p)(\hat\sigma_0^2(k) - 1) - (n-k)(\hat\sigma_1^2(k) - 1)$$
$$= Q_1 - Q_2 - Q_3,$$

and using the functional central limit theorem, we obtain for every $\varepsilon > 0$,

$$\max_{\varepsilon n < k \le n/2} \left| -\frac{n-p}{2}(\hat\sigma^2 - 1)^2 + \frac{n-k}{2}(\hat\sigma_1^2(k) - 1)^2 \right| = O_p(1),$$

and for every $\delta > 0$,

$$\lim_{\varepsilon \downarrow 0} \limsup_{n\to\infty} P\left[\max_{k \le \varepsilon n} \left| -\frac{n-p}{2}(\hat\sigma^2 - 1)^2 + \frac{n-k}{2}(\hat\sigma_1^2(k) - 1)^2 \right| > \delta \right] = 0,$$

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\Bigg[\max_{m \le k \le n/2} \Bigg| \frac{n-p}{3}(\hat\sigma^2 - 1)^3 (1+\eta)^{-3} - \frac{k-p}{3}(\hat\sigma_0^2(k) - 1)^3 (1+\eta_0)^{-3} - \frac{n-k}{3}(\hat\sigma_1^2(k) - 1)^3 (1+\eta_1)^{-3} \Bigg| > \delta \Bigg] = 0.$$

Also, from (A.10),

$$\frac{k-p}{2}(\hat\sigma_0^2(k) - 1)^2 - \frac{T_k^2}{2(k-p)} = O\left(\frac{(\ln_2 k)^{3/2}}{k^{1/2}} \right) \quad \text{a.s.}$$

The above results, together with (A.1), imply that for any $\delta > 0$,

$$\lim_{\varepsilon \downarrow 0} \limsup_{n\to\infty} P\left[\left| \max_{p<k\le \varepsilon n} \Lambda'_n(k) - \max_{p<k\le\varepsilon n} \left(\mathbf{S}'_k P_k \mathbf{S}_k + \frac{T_k^2}{2(k-p)} \right) \right| > \delta \right] = 0.$$

A similar relation holds for $\max_{k \ge (1-\varepsilon)n} \Lambda'_n(k)$. Thus, to complete the proof of (a), it suffices to show that for all $-\infty < x < \infty$,

$$\text{(A.12)} \qquad \lim_{\varepsilon \downarrow 0} \limsup_{n\to\infty} \left| P\left[\left(\max_{p<k\le\varepsilon n} \left(\mathbf{S}'_k P_k \mathbf{S}_k + \frac{T_k^2}{2(k-p)} \right) - b_n(p+2) \right) \Big/ a_n(p+2) \le x \right] - \exp(-e^{-x/2}) \right| = 0.$$

First note that the sequence $\xi_k := \begin{bmatrix} \mathbf{S}_k - \mathbf{S}_{k-1} \\ T_k - T_{k-1} \end{bmatrix}$ satisfies the hypotheses of Theorem 4 in Kuelbs and Philipp (1980). By virtue of the moment conditions on ε_1,

$$\Gamma = E(\xi_k \xi'_k) = \begin{bmatrix} \Gamma_{p+1} & \mathbf{0} \\ \mathbf{0} & 2 \end{bmatrix}.$$

Therefore, there exists a sequence of iid Gaussian random vectors $\left\{\begin{bmatrix}\mathbf{Z}_{k,1}\\ Z_{k,2}\end{bmatrix}\right\}$ with mean $\mathbf{0}$ and covariance matrix Γ such that

$$\begin{bmatrix}\mathbf{S}_k\\ T_k\end{bmatrix} - \begin{bmatrix}\mathbf{U}_{k,1}\\ U_{k,2}\end{bmatrix} = O(k^{1/2-\lambda}) \quad \text{as } k \to \infty \text{ a.s.}$$

for some $\lambda > 0$, where

$$\begin{bmatrix}\mathbf{U}_{k,1}\\ U_{k,2}\end{bmatrix} = \sum_{t=p+1}^{k} \begin{bmatrix}\mathbf{Z}_{t,1}\\ Z_{t,2}\end{bmatrix}.$$

Following the proof of Theorem 2.2, we have

$$\mathbf{S}'_k P_k \mathbf{S}_k - \mathbf{U}'_{k,1}\Gamma_{p+1}^{-1}\mathbf{U}_{k,1}/(k-p) \to 0 \quad \text{a.s.}$$

and

$$\frac{T_k^2}{2(k-p)} - \frac{U_{k,2}^2}{2(k-p)} \to 0 \quad \text{a.s.}$$

Since

$$\left\{\mathbf{U}'_{k,1}\Gamma_{p+1}^{-1}\mathbf{U}_{k,1}/(k-p) + U_{k,2}^2/(2(k-p)),\, k = p+1, p+2, \dots\right\}$$
$$=_d \{\mathbf{W}(k)'\mathbf{W}(k)/k,\, k = 1, 2, \dots\},$$

where $\{\mathbf{W}(t)\}$ is a standard $(p+2)$-dimensional Brownian motion, (A.12) follows from Corollary A.2.

(b) The proof of (b) is similar to the above argument and the proof of Proposition 3.2 and hence is omitted. □

PROOF OF THEOREM 3.4. We shall outline the proof for the case $p_0 = p_1 = p$ only. By strong approximation and the law of the iterated logarithm, it can be shown [cf. Lemma 2.4 of Yao and Davis (1986)] that

$$\max_{n/\ln n \le k \le n - n/\ln n} \Lambda_n^*(k) = O_p(\ln_3 n).$$

Following the proof of Theorem 3.3, one can establish that for all $\delta > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left[\max_{m \le k \le n/\ln n} \left|\Lambda_n^*(k) - \left\{\sigma^{-2}\mathbf{S}'_k P_k \mathbf{S}_k + \frac{k-p}{\mu_4 - \sigma^4}\left(\hat{\sigma}_0^2(k) - \sigma^2\right)^2\right\}\right| > \delta\right] = 0$$

and

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left[\max_{m \le k \le n/\ln n} \left|(k-p)\left(\hat{\sigma}_0^2(k) - \sigma^2\right)^2 - \frac{T_k^2}{k-p}\right| > \delta\right] = 0$$

and similar relations hold for k/n near 1, where $T_k = \boldsymbol{\varepsilon}'_k \boldsymbol{\varepsilon}_k - (k-p)\sigma^2$. The theorem follows from these results and Corollary A.2. □

REFERENCES

ATHREYA, K. B. and PANTULA, S. G. (1986a). A note on strong mixing of ARMA processes. *Statist. Probab. Lett.* **4** 187–190.

ATHREYA, K. B. and PANTULA, S. G. (1986b). Mixing properties of Harris chains and autoregressive processes. *J. Appl. Probab.* **23** 41–56.

BAGSHAW, M. and JOHNSON, R. A. (1977). Sequential procedures for detecting parameter changes in a time series. *J. Amer. Statist. Assoc.* **72** 593–597.

BASSEVILLE, M. and BENVENISTE, A. (1983). Sequential detection of abrupt changes in the spectral characteristics of digital signals. *IEEE Trans. Inform. Theory* **IT-29** 709–723.

BASSEVILLE, M. and BENVENISTE, A., eds. (1986). *Detection of Abrupt Changes in Signals and Dynamical Systems*. Springer, New York.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

DARLING, D. A. and ERDÖS, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J.* **23** 143–155.

DELONG, D. M. (1981). Crossing probabilities for a square root boundary by a Bessel process. *Comm. Statist. Theory Methods* **10** 2197–2213.

FEIGIN, P. D. and TWEEDIE, R. L. (1985). Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Series Anal.* **6** 1–15.

JAMES, B., JAMES, K. L. and SIEGMUND, D. (1987). Tests for a change-point. *Biometrika* **74** 71–83.

HALL, P. (1979). The rate of convergence of normal extremes. *J. Appl. Probab.* **16** 433–439.

HORVÁTH, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *Ann. Statist.* **21** 671–680.

KUELBS, J. and PHILIPP, W. (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *Ann. Probab.* **8** 1003–1036.

MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.

PICARD, D. (1985). Testing and estimating change-points in time series. *Adv. in Appl. Probab.* **17** 841–867.

SIEGMUND, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer, New York.

TELKSNYS, L., ed. (1986). *Detection of Changes in Random Processes*. Optimization Software, Inc., New York.

TSAY, R. S. (1988). Outliers, level shifts, and variance changes in time series. *J. Forecasting* **7** 1–20.

WILLSKY, A. S. (1976). A survey of designed methods for failure detection in dynamical systems. *Automatica* **12** 601–611.

YAO, Y.-C. and DAVIS, R. A. (1986). The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. *Sankhyā Ser. A* **48** 339–353.

YAO, Y.-C., HUANG, D. and DAVIS, R. A. (1994). On almost sure behavior of change-point estimators. In *Change-Point Problems* (E. Carlstein, H.-G. Müller and D. Siegmund, eds.) 359–372. IMS, Hayward, CA.

RICHARD A. DAVIS
YI-CHING YAO
DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523

DAWEI HUANG
SCHOOL OF MATHEMATICS
QUEENSLAND UNIVERSITY OF TECHNOLOGY
GPO BOX 2434
BRISBANE, QUEENSLAND 4001
AUSTRALIA