Journal of Statistical Computation and Simulation
Vol. 82, No. 12, December 2012, 1799–1812

Arc length tests for equivalent autocovariances

Ferebee Tunno[a]*, Colin Gallagher[b] and Robert Lund[b]

[a]*Department of Mathematics and Statistics, Arkansas State University, PO Box 70, Jonesboro, AR 72467, USA;* [b]*Department of Mathematical Sciences, Clemson University, PO Box 340975, Clemson, SC 29634-0975, USA*

(Received 17 April 2011; final version received 6 June 2011)

This paper presents a new test for discerning whether or not two independent time series have the same dynamics. Several methods in the time and spectral domain have previously been devised to conduct such a test. Here, we explore a new approach to the problem by comparing sample arc lengths (ALs) of the two series. Because sample ALs are derived from first-order differences in the series, the proposed methods work for both stationary autoregressive moving average and non-stationary autoregressive integrated moving average processes. This robustness is advantageous when one is unsure about unit root properties in the underlying series.

Keywords: arc length; autocovariance; ARMA; ARIMA

1. Introduction

In time series analysis, we seek to understand and model the evolution of a variable over time. We typically focus on the first- and second-order properties of the sequence. The autocovariance structure of a stationary time series gives an insight into the series behaviour. If two series are known to have the same structure, then autocovariances of one series can be used as a surrogate in inferential procedures involving the other series. Knowing whether or not two series have the same autocovariances may also indicate whether a series is sampled from a faulty or healthy environment. This paper suggests a new approach to testing for autocovariance equality via sample arc length (AL) methods. The methods are compared with some of the classical techniques for the problem. An excellent summary of existing tests is contained in [1].

When the two series are known to be stationary, Coates and Diggle [2] compared the spectral densities of the series to test for autocovariance equality. These methods rely upon the fact that two short-memory stationary autocovariance functions are equivalent if and only if the spectral densities agree at all frequencies (except on a set of Lebesgue measure zero). Kakizawa *et al.* [3] also considered spectral methods, handling the multivariate non-Gaussian case via Kullback–Leibler and Chernoff information criteria. In this work, the focus was on clustering and discrimination of multiple series.

*Corresponding author. Email: ftunno@astate.edu

ISSN 0094-9655 print/ISSN 1563-5163 online

http://dx.doi.org/10.1080/00949655.2011.595717
http://www.tandfonline.com

In the time domain (TD), Galeano and Peña [4] examined differences between the sample autocorrelations to make equality conclusions. Lund *et al.* [5] furthered these methods, handling the multivariate case and devising an asymptotic distribution for a quadratic form of sample autocovariance differences. A chi-squared limit law for the test statistic is proven from Bartlett's asymptotic limit formula [6, Chapter 7]. These tests are simple to use. Quinn [7] fitted high-order autoregressive (AR) processes to the two series and compared fitted coefficients to assess equality of series dynamics.

For non-stationary series, Huang *et al.* [8] used the smooth localized complex exponential model, which has the ability to extract localized spectral features of the series. Like Kakizawa *et al.* [3], Kullback–Leibler distances drive their methods. Choi *et al.* [9] used wavelet methods to detect changes in the autocorrelation structure of non-stationary series, examining the changes in the estimated wavelet-based spectra of adjacent blocks of the series. Corduas and Piccolo [10] investigated the statistical properties of the AR-based distance between autoregressive integrated moving average (ARIMA) processes. The distance measures the dissimilarity of two time series through the corresponding forecasts.

A different approach to the autocovariance equality problem utilizes the sample ALs of the two series. If one series has a bigger variance, then it should vary more in time and its sample AL should be larger. A series with positive correlations would not vary much in time and should have a comparatively smaller sample AL than a series with negative correlations. The proposed AL test is completely non-parametric in the sense that we do not fit any models to the given series, but as we show later, an AL-based statistic works for both stationary and certain non-stationary ARIMA models.

In the next section, AL is described and assumptions and notation are established. Section 3 then presents a formal AL test and quantifies its asymptotic distribution. In Section 4, simulations are run on ARIMA series. These simulations compare the type I error and power of the AL test to those of the SD and TD tests of Coates and Diggle [2] and Lund *et al.* [5], respectively. Section 5 presents an application, and Section 6 closes the paper with some remarks.

2. AL test

Let $\{X_t\}$ and $\{Y_t\}$ be two time series observed at times $t = 1, 2, \ldots, n$. The sample ALs of these series are the lengths of the polygonal loci connecting the $(t, X_t)_{t=1}^n$ and $(t, Y_t)_{t=1}^n$ sample points:

$$\sum_{t=2}^{n} \sqrt{1 + (X_t - X_{t-1})^2}, \quad \sum_{t=2}^{n} \sqrt{1 + (Y_t - Y_{t-1})^2}.$$

For example, the AL of the process in Figure 1 (left) is 278.8367 units, while the AL of the process in Figure 1 (right) is 539.104 units.

$$S_t^X = \sqrt{1 + (X_t - X_{t-1})^2} \quad \text{and} \quad S_t^Y = \sqrt{1 + (Y_t - Y_{t-1})^2}.$$

The sample ALs of $\{X_t\}$ and $\{Y_t\}$ are then $\sum_{t=2}^n S_t^X$ and $\sum_{t=2}^n S_t^Y$, respectively.

Our main concern lies with zero mean series. Henceforth, we take $E(X_t) = E(Y_t) \equiv 0$. We also assume that $\{X_t\}$ and $\{Y_t\}$ are independent.

2.1. *ARMA(p,q) processes*

Suppose that $\{X_t\}$ and $\{Y_t\}$ are causal and invertible stationary autoregressive moving average (ARMA) series. We assume that the innovations governing the ARMA models are zero mean i.i.d.

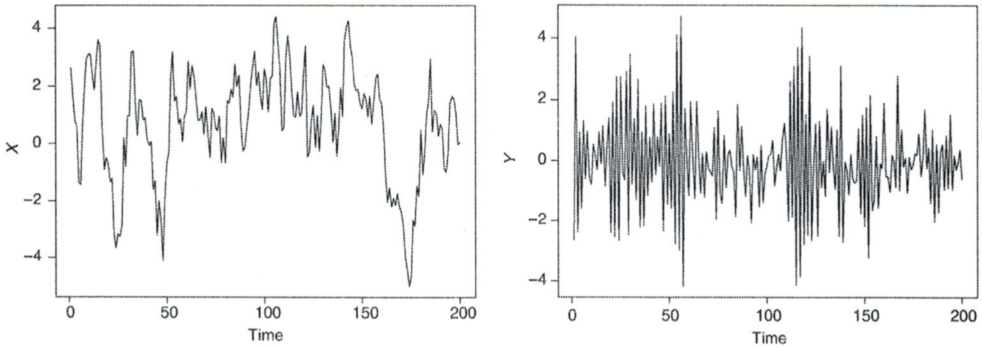

Figure 1. The process on the left has an AL of 278.8367 units, while the process on the right has an AL of 539.104 units.

and have a finite fourth moment. Let $\gamma_X(h) = E(X_t X_{t+h})$ and $\gamma_Y(h) = E(Y_t Y_{t+h})$ denote lag h autocovariances. We wish to test

$$\begin{aligned} &H_0: \gamma_X(h) = \gamma_Y(h) \quad \text{for all } h \\ \text{versus} \quad &H_1: \gamma_X(h) \neq \gamma_Y(h) \quad \text{for at least one } h. \end{aligned} \tag{1}$$

A normalized AL-based statistic for this test is

$$T = \frac{(\sum_{t=2}^{n} S_t^X - \sum_{t=2}^{n} S_t^Y) - (n-1)(\mu(S_t^X) - \mu(S_t^Y))}{\sqrt{\text{Var}(\sum_{t=2}^{n} S_t^X - \sum_{t=2}^{n} S_t^Y)}},$$

where $\mu(S_t^X) \equiv E(S_t^X)$ and $\mu(S_t^Y) \equiv E(S_t^Y)$ are the theoretical means of the AL segments of $\{X_t\}$ and $\{Y_t\}$, respectively. The assumptions imply that $\{X_t\}$ and $\{Y_t\}$ are strictly stationary; hence, $\{S_t^X\}$ and $\{S_t^Y\}$ are also strictly stationary. Thus, under the null hypothesis, $E(S_t^X)$ and $E(S_t^Y)$ are equal and constant in t and

$$T = \frac{\sum_{t=2}^{n} S_t^X - \sum_{t=2}^{n} S_t^Y}{\sqrt{\text{Var}(\sum_{t=2}^{n} S_t^X) + \text{Var}(\sum_{t=2}^{n} S_t^Y)}}. \tag{2}$$

For variances of the sample ALs, the stationarity of $\{S_t^X\}$ gives

$$\text{Var}\left(\sum_{t=2}^{n} S_t^X\right) = (n-1)\gamma_{S^X}(0) + 2\sum_{h=1}^{n-2}(n-1-h)\gamma_{S^X}(h),$$

where the subscript S^X signifies that $\{S_t^X\}$ is being considered. We will estimate this variance by truncating the sum at the greatest integer less than or equal to $\sqrt[3]{n}$:

$$\widehat{\text{Var}}\left(\sum_{t=2}^{n} S_t^X\right) = (n-1)\hat{\gamma}_{S^X}(0) + 2\sum_{h=1}^{[\sqrt[3]{n}]}(n-1-h)\hat{\gamma}_{S^X}(h).$$

This estimator is consistent and avoids the bias associated with large lags when the non-negative definite autocovariance estimator

$$\widehat{\gamma}_X(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_t - \overline{X})(X_{t+h} - \overline{X})$$

is used. Berkes *et al.* [11] discuss this and more general truncation schemes. Analogous expressions can be derived for $\text{Var}(\sum_{t=2}^{n} S_t^Y)$.

THEOREM 2.1 *If $\{X_t\}$ and $\{Y_t\}$ are independent causal and invertible ARMA series whose innovations have a finite fourth moment, then T in Equation (2) converges in distribution (under H_0) to a standard normal random variable. That is,*

$$T \xrightarrow{D} N(0, 1).$$

Proof First, we prove asymptotic normality for the sample AL of $\{X_t\}$. It is sufficient to verify Theorem 4.1 from Ho and Hsing [12], which requires their conditions (3.3) and $C(t, \tau, \lambda)$ to hold (for some τ and λ) when $t = 0, 1$. We will verify these conditions with $\tau = 0$ and $\lambda = 0$.

Let $K(x) = \sqrt{1 + x^2}$. Then, Ho and Hsing's [12] functions $K_{0,0}^{(0)}(x)$ and $K_{0,0}^{(1)}(x)$ are $K_{0,0}^{(0)}(x) = K(x)$ and $K_{0,0}^{(1)}(x) = |K'(x)| = |x|/\sqrt{1 + x^2}$, both of which are well-defined and continuous for all x. Thus, Part 1 of condition $C(t, \tau, \lambda)$ is satisfied for $t = 0, 1$. The causal ARMA assumption allows us to write $X_t - X_{t-1} = \sum_{i=1}^{\infty} a_i Z_{t-i}$, where $\sum_{i=1}^{\infty} |a_i| < \infty$. The Cauchy–Schwarz inequality implies

$$E\left[\left(\sum_{i=1}^{\infty} |a_i Z_i|\right)^4\right] \leq \left(\sum_{i=1}^{\infty} |a_i|\right)^4 E(Z_0^4) < \infty,$$

since for all i, j, k, l, we have $E(|Z_i Z_j Z_k Z_l|) \leq E(Z_0^4)$. This gives

$$\sup_{I \subset \{1,2,\ldots\}} E\left[\left(\sqrt{1 + \left(x + \sum_{i \in I} a_i Z_i\right)^2}\right)^4\right] \leq E\left[\left(1 + \left(|x| + \sum_{i=1}^{\infty} |a_i Z_i|\right)^2\right)^2\right]$$
$$\leq 16\left[1 + x^4 + E\left(\left(\sum_{i=1}^{\infty} |a_i Z_i|\right)^4\right)\right] < \infty.$$

Thus, Part 2 of condition $C(t, \tau, \lambda)$ is satisfied for $t = 0$. Because $|z|/\sqrt{1 + z^2} \leq 1$ for any z,

$$\sup_{I \subset \{1,2,\ldots\}} E\left[\left(\frac{|x + \sum_{i \in I} a_i Z_i|}{\sqrt{1 + (x + \sum_{i \in I} a_i Z_i)^2}}\right)^4\right] \leq 1$$

and Part 2 of condition $C(t, \tau, \lambda)$ is also satisfied for $t = 1$.

Now let

$$\eta = \sqrt{1 + \left(\sum_{i=1}^{\infty} |a_i Z_{1-i}|\right)^2}, \quad \xi_n = \sqrt{1 + \left(\sum_{i=1}^{n} a_i Z_{1-i}\right)^2}, \quad \text{and} \quad \xi = \sqrt{1 + \left(\sum_{i=1}^{\infty} a_i Z_{1-i}\right)^2}.$$

Then, $|\xi_n| \leq \eta$, $\xi_n \xrightarrow{\text{a.s.}} \xi$, and $E(\eta^2) < \infty$. By dominated convergence, $E(\xi - \xi_n)^2 \to 0$ as $n \to \infty$. Thus, condition (3.3) in [12] is satisfied.

Theorem 4.1 in [12] now gives

$$L_X = \frac{1}{\sqrt{n}}\left(\sum_{t=2}^{n} S_t^X - \sum_{t=2}^{n} \mu(S_t^X)\right) \xrightarrow{D} N(0, \sigma_X^2),$$

where $\sigma_X^2 = \lim_{n\to\infty} \mathrm{Var}(L_X)$. An analogous result holds for the $\{Y_t\}$ process. Invoking the independence of $\{X_t\}$ and $\{Y_t\}$ gives

$$L_X - L_Y \stackrel{H_0}{=} \frac{1}{\sqrt{n}}\left(\sum_{t=2}^{n} S_t^X - \sum_{t=2}^{n} S_t^Y\right) \xrightarrow{D} N(0, \sigma_X^2 + \sigma_Y^2).$$

The theorem is now proven when Slutsky's theorem is used with the usual consistent estimate of $\sigma_X^2 + \sigma_Y^2$. ■

To test Equation (1) at significance level α, one rejects H_0 when $|T| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the standard normal critical value with area $\alpha/2$ to its right. From the assumed invertibility of the ARMA series, $\gamma_X(h) = \gamma_Y(h)$ for all $h \geq 0$ if and only if all ARMA parameters of $\{X_t\}$ and $\{Y_t\}$ agree. Hence, the above test can also be used to test for equivalent ARMA parameters.

2.2. *ARIMA(p,1,q) processes*

AL methods are somewhat robust against departures from stationarity. To see this, suppose that $\{X_t\}$ and $\{Y_t\}$ are independent causal and invertible ARIMA series with ARIMA order $d = 1$. Then, $\{X_t\}$ and $\{Y_t\}$ are non-stationary. However, one could still test for equivalent time series dynamics by noting that ALs depend only on first-order differences of the series, and the first-order differences are stationary and ARMA. Phrased another way, two ARIMA series can also be discriminated via Theorem 2.1. Since one can again make the representation

$$X_t - X_{t-1} = \sum_{i=0}^{\infty} a_i Z_{t-i}$$

with $\sum_{i=0}^{\infty} |a_i| < \infty$, the same proof carries through.

3. Modified AL test

One can ask if AL tests perform the same if the unity term and the square root in $\sqrt{1+x^2}$ are ignored. Let $V_t^X = (X_t - X_{t-1})^2$ and $V_t^Y = (Y_t - Y_{t-1})^2$. Then, $\sum_{t=2}^{n} S_t^X$ and $\sum_{t=2}^{n} S_t^Y$ are significantly different if and only if $\sum_{t=2}^{n} V_t^X$ and $\sum_{t=2}^{n} V_t^Y$ are significantly different. To test Equation (1), S_t^X and S_t^Y are replaced with V_t^X and V_t^Y in Equation (2). The following theorem establishes the convergence result of the modified AL test.

THEOREM 3.1 *If $\{X_t\}$ and $\{Y_t\}$ are independent causal and invertible ARMA series whose innovations have a finite fourth moment, then T in Equation (2) converges in distribution (under H_0) to a standard normal random variable when S_t^X and S_t^Y are replaced by V_t^X and V_t^Y, respectively. That is,*

$$T \xrightarrow{D} N(0, 1).$$

Proof Suppose that $\{X_t\}$ is a causal and invertible ARMA series whose innovations are i.i.d. and have a finite fourth moment. Bartlett's classical asymptotic normality for sample autocovariances

[6, Proposition 7.3.3] gives

$$\begin{bmatrix} \hat{\gamma}_X(0) \\ \hat{\gamma}_X(1) \end{bmatrix} \sim AN\left(\begin{bmatrix} \gamma_X(0) \\ \gamma_X(1) \end{bmatrix}, \frac{M_X}{n}\right),$$

where for $p, q \in \{0, 1\}$, the (p, q)th entry of covariance matrix M_X is

$$(\beta - 3)\gamma_X(p)\gamma_X(q) + \sum_{k=-\infty}^{\infty} (\gamma_X(k)\gamma_X(k - p + q) + \gamma_X(k + q)\gamma_X(k - p))$$

and $E(Z_t^X)^4 = \beta(\mathrm{Var}(Z_t^X))^2$. After some algebra, we conclude that

$$\begin{aligned} \frac{1}{n}\sum_{t=2}^{n} V_t^X &= \frac{2}{n}\left(\sum_{t=2}^{n} X_t^2 - \sum_{t=2}^{n} X_t X_{t-1}\right) + \frac{X_1^2 - X_n^2}{n} \\ &= 2(\widetilde{\gamma}_X(0) - \widetilde{\gamma}_X(1)) + \frac{X_1^2 - X_n^2}{n} \\ &\sim AN\left(2(\gamma_X(0) - \gamma_X(1)), \frac{W_X}{n}\right), \end{aligned}$$

where $\widetilde{\gamma}_X(h) = n^{-1}\sum_{t=2}^{n} X_t X_{t+h}$ and

$$\begin{aligned} W_X &= 4((\beta - 3)(\gamma_X^2(0) + \gamma_X^2(1) - \gamma_X(0)\gamma_X(1))) \\ &\quad + 4\sum_{k=-\infty}^{\infty} [3\gamma_X^2(k) + \gamma_X(k+1)\gamma_X(k-1) - 2\gamma_X(k)\gamma_X(k+1)]. \end{aligned}$$

We estimate W_X via

$$\begin{aligned} \widehat{W}_X &= 4((\beta - 3)(\hat{\gamma}_X^2(0) + \hat{\gamma}_X^2(1) - \hat{\gamma}_X(0)\hat{\gamma}_X(1))) \\ &\quad + 4\sum_{k=-[\sqrt[3]{n}]}^{[\sqrt[3]{n}]} [3\hat{\gamma}_X^2(k) + \hat{\gamma}_X(k+1)\hat{\gamma}_X(k-1) - 2\hat{\gamma}_X(k)\hat{\gamma}_X(k+1)]. \end{aligned}$$

An analogous result applies to $\{Y_t\}$. Slutsky's theorem shows that under H_0,

$$\frac{(1/(n-1))\left(\sum_{t=2}^{n} V_t^X - \sum_{t=2}^{n} V_t^Y\right)}{\sqrt{\widehat{W}_X/(n-1) + \widehat{W}_Y/(n-1)}} \xrightarrow{D} N(0, 1),$$

since $\{X_t\}$ and $\{Y_t\}$ are independent and $\widehat{W}_X$ and $\widehat{W}_Y$ consistently estimate W_X and W_Y, respectively. ■

Figure 2 plots type I errors and powers for the AL and modified arc length (MAL) tests when $\{X_t\}$ and $\{Y_t\}$ are drawn from an AR(1) model with varying first-order AR coefficient $\phi \in (-1, 1)$. The two methods perform almost identically. Additional simulations with other ARMA orders (not shown here) reveal the same behaviour; hence, the MAL method will not be considered further. This modified test, however, is significant in that it reveals a direct connection between the AL concept and the autocovariance function.

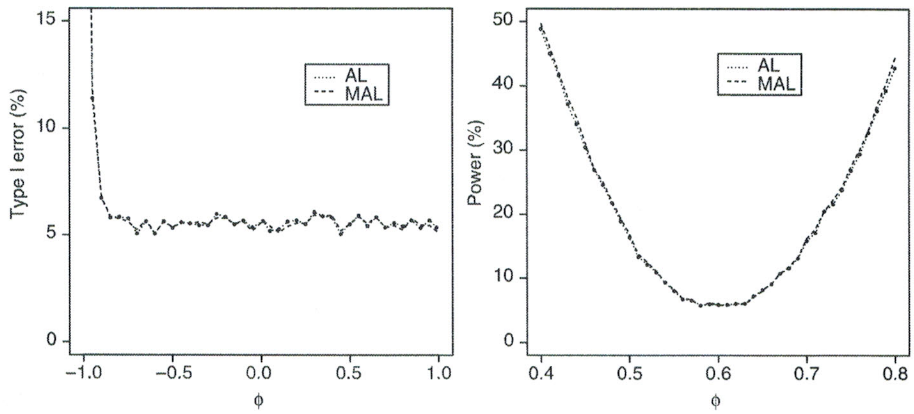

Figure 2. Type I error for both AL tests when $\{X_t\}$ and $\{Y_t\}$ follow the same AR(1) process with $-1 < \phi_X = \phi_Y < 1$ (left). Power for both AL tests when $\{X_t\}$ and $\{Y_t\}$ follow different AR(1) processes with $\phi_X = 0.6$ and $0.4 \le \phi_Y \le 0.8$ (right).

4. Simulations

4.1. *Set-up*

This section compares the type I error and power of the AL test, the TD test of Lund *et al.* [5], and the SD test of Coates and Diggle [2] for testing Equation (1). For each figure, series of length 1000 were generated and hypotheses were tested at level $\alpha = 0.05$. Ten thousand independent simulations were conducted to estimate the type I errors and powers. Below is a brief summary of the SD and TD tests.

4.2. *SD and TD tests*

Let $f_X(\cdot)$ and $f_Y(\cdot)$ be the spectral densities for two stationary ARMA series $\{X_t\}$ and $\{Y_t\}$, where

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}h\omega} \gamma_X(h) \quad \text{and} \quad f_Y(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}h\omega} \gamma_Y(h)$$

for frequency $\omega \in [0, 2\pi)$ and $\mathrm{i} = \sqrt{-1}$. Then, $\gamma_X(h) = \gamma_Y(h)$ for all $h \ge 0$ if and only if $f_X(\omega) = f_Y(\omega)$ for all $\omega \in [0, 2\pi)$. Since the spectral densities are unknown in practice, they are estimated by

$$\widehat{f_X}(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t \mathrm{e}^{-\mathrm{i}t\omega} \right|^2 \quad \text{and} \quad \widehat{f_Y}(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} Y_t \mathrm{e}^{-\mathrm{i}t\omega} \right|^2.$$

Coates and Diggle [2] noted that if

$$D_v = \ln\left(\frac{\widehat{f_X}(\omega_v)}{\widehat{f_Y}(\omega_v)} \right),$$

where $\omega_v = 2\pi v/n$, then large values of

$$\overline{D} \equiv \frac{1}{n/2 - 1} \sum_{v=1}^{n/2} |D_v|$$

suggest rejection of equivalent autocovariances. Using a central limit theorem argument, they noted that since $E|D_v| = \ln(4)$ and $\mathrm{Var}(|D_v|) = \pi^2/3 - (\ln(4))^2$, then an α-level test rejects equivalent autocovariances when

$$\overline{D} > \ln(4) + z_\alpha \sqrt{\frac{\pi^2/3 - (\ln(4))^2}{n/2 - 1}}.$$

Since our simulations use $\alpha = 0.05$ and $n = 1000$, the SD test in this section rejects equivalent autocovariances when $\overline{D} > 1.472$.

The TD test of Lund *et al.* [5] is based on a result from Bartlett (see [6, Chapter 7]) which states that

$$\begin{bmatrix} \widehat{\gamma}(0) \\ \widehat{\gamma}(1) \\ \widehat{\gamma}(2) \\ \vdots \\ \widehat{\gamma}(L) \end{bmatrix} \sim AN \left(\begin{bmatrix} \gamma(0) \\ \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(L) \end{bmatrix}, \frac{\mathbf{W}}{n} \right)$$

as $n \to \infty$, where $\mathbf{W}$ is an $(L+1) \times (L+1)$ matrix with (i, j)th entry

$$W_{ij} = (\eta - 3)\gamma(i)\gamma(j) + \sum_{k=-\infty}^{\infty} [\gamma(k)\gamma(k - i + j) + \gamma(k + j)\gamma(k - i)]$$

for $0 \leq i, j \leq L$. It is assumed that the series in question is invertible and that $E(Z_t^4) = \eta\sigma^4$ is finite. Under a null hypothesis of equal autocovariances,

$$V = \begin{bmatrix} \widehat{\gamma}_X(0) - \widehat{\gamma}_Y(0) \\ \widehat{\gamma}_X(1) - \widehat{\gamma}_Y(1) \\ \widehat{\gamma}_X(2) - \widehat{\gamma}_Y(2) \\ \vdots \\ \widehat{\gamma}_X(L) - \widehat{\gamma}_Y(L) \end{bmatrix} \sim AN\left(\mathbf{0}, \frac{2\mathbf{W}}{n}\right)$$

as $n \to \infty$, where $\hat{\gamma}_X(h)$ and $\hat{\gamma}_Y(h)$ estimate lag h autocovariances of $\{X_t\}$ and $\{Y_t\}$, respectively.

It follows that

$$\left(\frac{n}{2}\right) V^T \mathbf{W}^{-1} V \xrightarrow{D} \chi^2_{L+1},$$

where $\mathbf{W}$ is estimated by $\widehat{\mathbf{W}}$. This estimate replaces $\gamma(h)$ with $2^{-1}(\widehat{\gamma}_X(h) - \widehat{\gamma}_Y(h))$ and truncates the infinite sum with index values between $\pm\lfloor n^{1/3} \rfloor$. Thus, if $(n/2)V^T\widehat{\mathbf{W}}^{-1}V$ exceeds the $(1-\alpha)$th quantile of the χ^2_{L+1} distribution, the null hypothesis is rejected. For the simulations below, $L = 10$ with critical value $\chi^2_{11,0.05} = 19.68$.

4.3. *Error and power*

We now summarize the simulation results, where both stationary and non-stationary processes are considered. In general, the TD test outperforms the AL test in the stationary ARMA case, but as mentioned above, AL is somewhat robust against non-stationarity.

4.3.1. *Stationary processes*

We provide simulations for AR(1), AR(2), MA(1) and MA(2) models. These simulations are reflective of the general behaviour we expect for stationary ARMA processes. In stationary settings, the TD test is typically the most powerful and outperforms the AL test. However, the AL test tends to have higher power than the SD test. The type I errors of the TD test are close to the nominal level (0.05 in our simulations) as long as the AR parameters are not close to stationarity boundaries, but become inflated for nearly non-stationary processes.

Figure 3 plots results when $\{X_t\}$ and $\{Y_t\}$ follow an MA(2) process. In this case, the AL test performs better than the SD test and nearly as well as the TD test. However, it is possible for the TD test to have significantly better power than AL, such as when $\{X_t\}$ and $\{Y_t\}$ are drawn from MA(1) models, as described in Figure 4.

In Figure 5, $\{X_t\}$ and $\{Y_t\}$ follow an AR(1) process. When ϕ gets close to unity, the type I errors for both the AL and SD tests remain close to 0.05, while the TD test errors become inflated. In Figure 6, $\{X_t\}$ and $\{Y_t\}$ follow an AR(2) process. Both the SD and AL tests have a steady Type I error around 5%, while the TD's Type I error blows up at $\phi_{X,2} = \phi_{Y,2} \approx 0.45$, which is

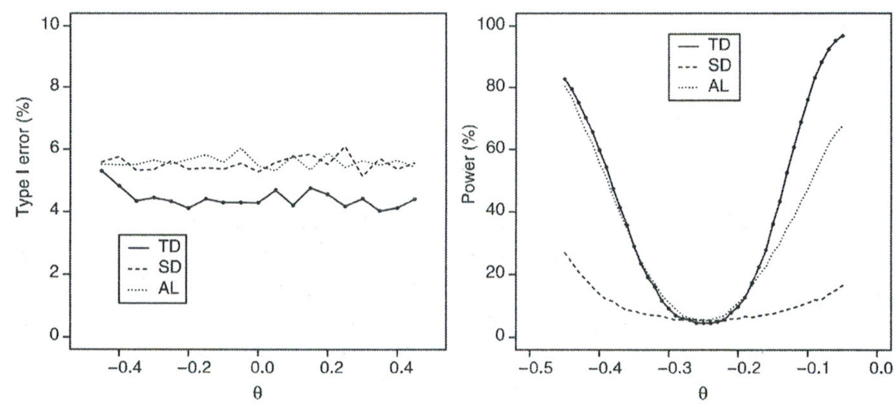

Figure 3. Type I error when $\{X_t\}$ and $\{Y_t\}$ follow the same MA(2) process with $\theta_{X,1} = \theta_{Y,1} = 0.5$ and $-0.45 \leq \theta_{X,2} = \theta_{Y,2} \leq 0.45$ (left). Power when $\{X_t\}$ and $\{Y_t\}$ follow different MA(2) processes with $\theta_{X,1} = \theta_{Y,1} = 0.5$, $\theta_{X,2} = -0.25$, and $-0.45 \leq \theta_{Y,2} \leq -0.05$ (right).

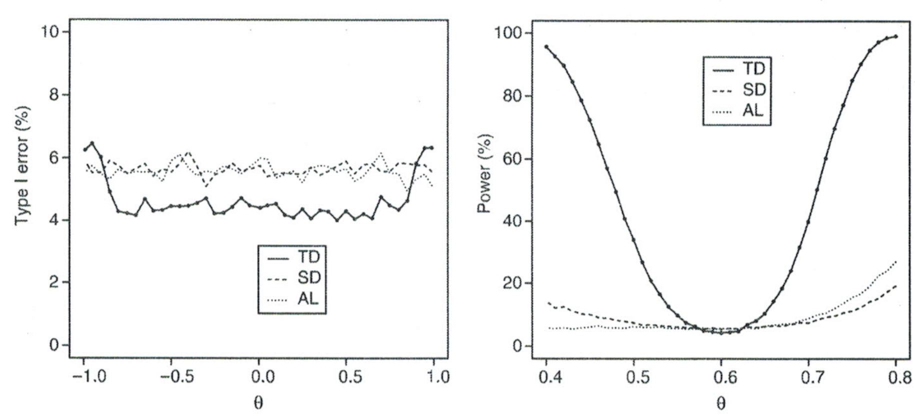

Figure 4. Type I error when $\{X_t\}$ and $\{Y_t\}$ follow the same MA(1) process with $-1 < \theta_X = \theta_Y < 1$ (left). Power when $\{X_t\}$ and $\{Y_t\}$ follow different MA(1) processes with $\theta_X = 0.6$ and $0.4 \leq \theta_Y \leq 0.8$ (right).

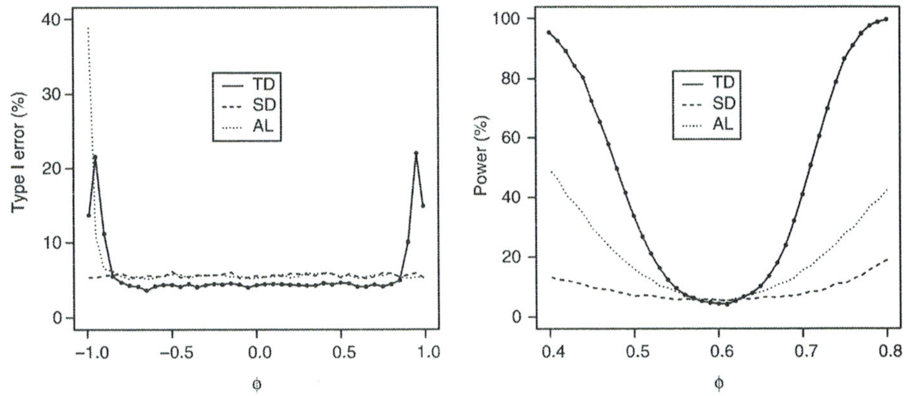

Figure 5. Type I error when $\{X_t\}$ and $\{Y_t\}$ follow the same AR(1) process with $-1 < \phi_X = \phi_Y < 1$ (left). Power when $\{X_t\}$ and $\{Y_t\}$ follow different AR(1) processes with $\phi_X = 0.6$ and $0.4 \le \phi_Y \le 0.8$ (right).

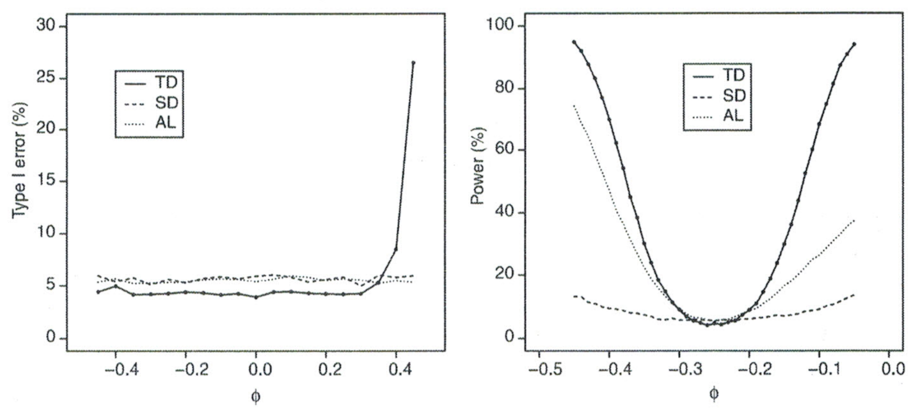

Figure 6. Type I error when $\{X_t\}$ and $\{Y_t\}$ follow the same AR(2) process with $\phi_{X,1} = \phi_{Y,1} = 0.5$ and $-0.45 \le \phi_{X,2} = \phi_{Y,2} \le 0.45$ (left). Power when $\{X_t\}$ and $\{Y_t\}$ follow different AR(2) processes with $\phi_{X,1} = \phi_{Y,1} = 0.5$, $\phi_{X,2} = -0.25$, and $-0.45 \le \phi_{Y,2} \le -0.05$ (right).

where this AR(2) process approaches non-stationarity. In both Figures 5 and 6, the SD test has the lowest power.

In Figure 7, the power of the tests is examined when $\{X_t\}$ follows an AR(1) process and $\{Y_t\}$ follows an MA(1) process. We select $\theta_Y = \sqrt{\phi_X^2/(1-\phi_X^2)}$ so that $\gamma_X(0) = \gamma_Y(0)$, but $\gamma_X(h) \neq \gamma_Y(h)$ for $h \ge 1$. All three tests have poor power when the magnitude of the AR parameter is close to zero, which is when both processes reduce to white noise.

4.3.2. *Non-stationary processes*

We have seen that the TD method, which requires stationarity, becomes unstable as we approach a unit root. Here, we consider non-stationary explosive autoregressions, unit root models, and ARIMA(p, d, q) models.

In Figure 8, $\{X_t\}$ and $\{Y_t\}$ both follow an ARIMA(1,1,0) process. Since the AL test is robust against this type of non-stationarity, we see that it performs significantly better than the other two tests presumably since these tests require stationarity. Both the TD and SD tests have type I errors significantly larger than 0.05, and yet have lower power for much of the parameter space.

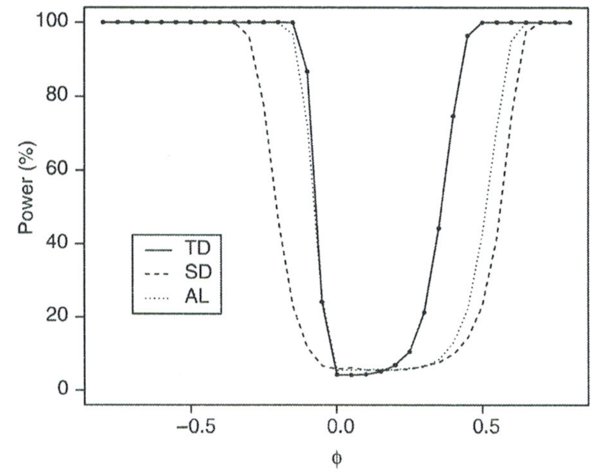

Figure 7. Power when $\{X_t\} \sim$ AR(1) and $\{Y_t\} \sim$ MA(1), where $\gamma_X(0) = \gamma_Y(0)$, but $\gamma_X(h) \neq \gamma_Y(h)$ for $h \geq 1$.

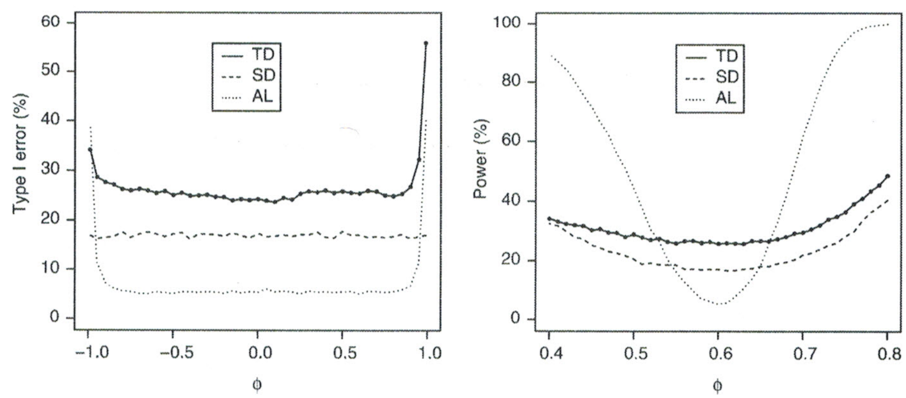

Figure 8. Type I error when $\{X_t\}$ and $\{Y_t\}$ follow the same ARIMA(1,1,0) process with $-1 < \phi_X = \phi_Y < 1$ (left). Power when $\{X_t\}$ and $\{Y_t\}$ follow different ARIMA(1, 1, 0) processes with $\phi_X = 0.6$ and $0.4 \leq \phi_Y \leq 0.8$ (right).

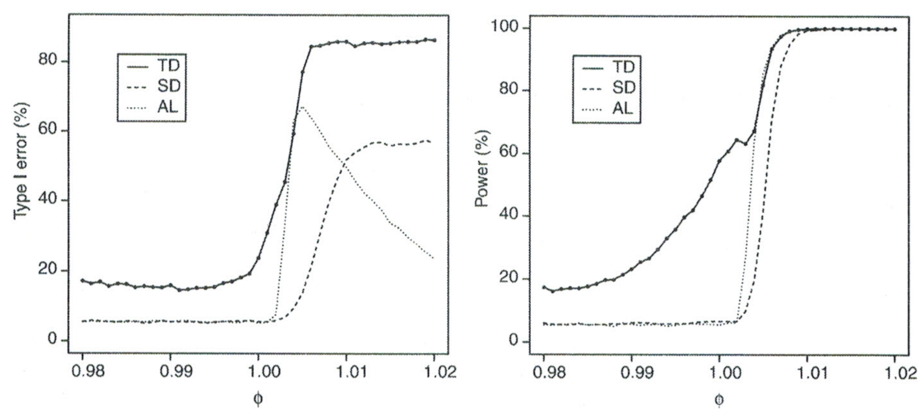

Figure 9. Type I error when $\{X_t\}$ and $\{Y_t\}$ follow the same AR(1) process with $0.98 \leq \phi_X = \phi_Y \leq 1.02$ (left). Power when $\{X_t\}$ and $\{Y_t\}$ follow different AR(1) processes with $\phi_X = 0.98$ and $0.98 \leq \phi_Y \leq 1.02$ (right).

In Figure 9, we compare two AR(1) processes that have parameters close to 1. Specifically, the $\{X_t\}$ process follows the possibly explosive autoregression

$$X_t = \phi_X X_{t-1} + Z_t, \quad t \geq 1,$$

where $\phi \in [0.98, 1.02]$ and $X_0 = 0$. The $\{Y_t\}$ process takes the same form. Once ϕ exceeds unity, the type I error for all three tests degrades, although the AL test's type I error decreases to nominal levels as ϕ moves away from unity. We note that the set of ϕ's for which the AL method provides a size 0.05 test includes the unit root case. The type I error for the TD and SD methods make these tests unusable in this situation.

5. An application

In the twenty-first century, the Internet has become an indispensable part of our lives. The two most popular web browsers are Internet Explorer and Mozilla Firefox, which are owned by Microsoft and Time Warner, respectively. While IE has been in existence longer than Mozilla and is used by more people, is it true that their parent companies have stocks that behave differently?

Figure 10 shows the daily stock prices for these two companies from 17 February 2006 to 17 February 2011, where $n = 1260$. When the AL test is applied to the two processes at the 5% significance level, the test statistic is $2.388 > 1.96 = z_{0.025}$. Thus, the null hypothesis of equal autocovariance structures is rejected. Are the two series actually different though?

It is important to note that around the first week of October 2008, the stock market began a serious decline which lasted for several months. Both Time Warner and Microsoft experienced a big dip in stock prices which most likely introduced non-stationarity.

When the two series are truncated to contain stock prices from 17 February 2006 to 30 September 2008 ($n = 659$), the AL test yields a test statistic of $1.789 < 1.96 = z_{0.025}$. So, in this case, the null hypothesis of equal autocovariances is retained. However, when the two series are truncated to contain stock prices from 1 July 2009[1] to 17 February 2011 ($n = 413$), the AL test yields a test statistic value of $2.980 > 1.96 = z_{0.025}$. Thus, the null hypothesis of equal autocovariance structures is rejected.

When ARIMA models are fit to both of the original, untruncated series (using both parameter parsimony and AIC criteria), Time Warner follows an ARIMA(2, 1, 3) process, while Microsoft

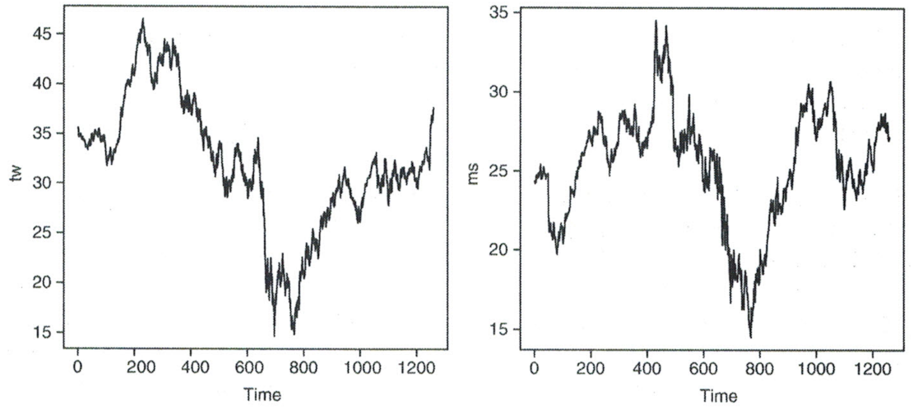

Figure 10. Daily stock prices for Time Warner (left) and Microsoft (right) from 17 February 2006 to 17 February 2011.

follows an ARIMA(4, 1, 4) process. The difference in models appears to confirm the conclusion of different autocovariance structures, and the two additional hypothesis tests for the truncated series suggest that the change in structures probably occurred sometime after the 2008 dip in stock prices.

6. Closing remarks

The simulations in Section 4 show the proposed AL test to be a reasonable, albeit imperfect, one. The AL method is intuitive and performs well for a variety of stationary and non-stationary processes. Overall, it competes respectably with both the TD and SD methods, although the TD test is generally more powerful when the data come from a stationary ARMA process.

One way to increase power is to increase the sample size. Figure 11 shows how the AL power increases as the sample size increases from 1000 to 20,000 for both AR(1) and MA(1) processes, with a fixed parameter difference of 0.1 in both cases. Specifically, $\phi_X = 0.6$ and $\phi_Y = 0.7$, while $\theta_X = 0.6$ and $\theta_Y = 0.7$. The AR(1) power increases more rapidly than the MA(1) power.

The ideas in this paper can also be extended to multivariate settings. If there are m observations per time unit, that is, $\mathbf{X}_t = (X_{1t}, X_{2t}, \ldots, X_{mt})^T$ and $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, \ldots, Y_{mt})^T$, define

$$S_t^{\mathbf{X}} = \sqrt{1 + \sum_{j=1}^{m} (X_{jt} - X_{j,t-1})^2} \quad \text{and} \quad S_t^{\mathbf{Y}} = \sqrt{1 + \sum_{j=1}^{m} (Y_{jt} - Y_{j,t-1})^2},$$

to get ALs $\sum_{t=2}^{n} S_t^{\mathbf{X}}$ and $\sum_{t=2}^{n} S_t^{\mathbf{Y}}$. If $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ are independent causal and invertible vector ARMA processes, then testing whether the two autocovariance functions agree involves a test statistic of the form

$$\frac{\sum_{t=2}^{n} S_t^{\mathbf{X}} - \sum_{t=2}^{n} S_t^{\mathbf{Y}}}{\sqrt{\mathrm{Var}\left(\sum_{t=2}^{n} S_t^{\mathbf{X}}\right) + \mathrm{Var}\left(\sum_{t=2}^{n} S_t^{\mathbf{Y}}\right)}}.$$

The ideas extend as before, regardless of the dimension m. This may be advantageous when m is very large and explicit vector ARMA modelling becomes problematic due to parsimony issues involved with large m.

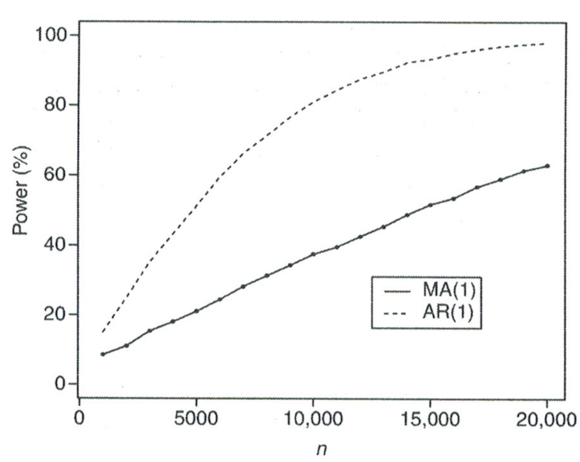

Figure 11. Power as a function of sample size for both AR(1) and MA(1) processes. For the former, $\phi_X = 0.6$ and $\phi_Y = 0.7$, while for the latter, $\theta_X = 0.6$ and $\theta_Y = 0.7$.

We can also relax the equal sample sizes assumption by examining

$$\frac{(1/n_1)\sum_{t=2}^{n_1} S_t^{\mathbf{X}} - (1/n_2)\sum_{t=2}^{n_2} S_t^{\mathbf{Y}}}{\sqrt{(1/n_1^2)\,\mathrm{Var}\left(\sum_{t=2}^{n_1} S_t^{\mathbf{X}}\right) + (1/n_2^2)\,\mathrm{Var}\left(\sum_{t=2}^{n_2} S_t^{\mathbf{Y}}\right)}}.$$

Here n_1 and n_2 denote the respective sample sizes of $\{X_t\}$ and $\{Y_t\}$.

Acknowledgement

The authors would like to give special thanks to Hany Bassily for first suggesting the AL idea.

Note

1. Most economists pegged June 2009 as the end of the recession.

References

[1] J. Caiado, N. Crato, and D. Peña, *A periodogram-based metric for time series classification*, Comput. Statist. Data Anal. 50 (2006), pp. 2668–2684.
[2] D.S. Coates and P.J. Diggle, *Tests for comparing two estimated spectral densities*, J. Time Ser. Anal. 7 (1986), pp. 7–20.
[3] Y. Kakizawa, R. Shumway, and M. Taniguchi, *Discrimination and clustering for multivariate time series*, J. Amer. Statist. Assoc. 93 (1998), pp. 328–340.
[4] P. Galeano and D. Peña, *Multivariate analysis in vector time series*, Statistics and Econometrics Series 15, Working Papers 1–24, Departamento de Estadística y Econometría, Madrid, 2000. Available at http://e-archivo.uc3m.es/bitstream/10016/162/1/ws012415.pdf.
[5] R. Lund, H. Bassily, and B. Vidakovic, *Testing equality of stationary autocovariances*, J. Time Ser. Anal. 30 (2009), pp. 332–348.
[6] P. J. Brockwell and R.A. Davis, *Time Series: Theory and Methods*, 2nd ed., Springer, New York, 2006.
[7] B.G. Quinn, *Statistical methods of spectrum change detection*, Digit. Signal Process. 16 (2006), pp. 588–596.
[8] H.Y. Huang, H. Ombao, and D.S. Stoffer, *Discrimination and classification of nonstationary time series using the SLEX model*, J. Amer. Statist. Assoc. 99 (2004), pp. 763–774.
[9] H. Choi, H. Ombao, and B. Ray, *Sequential change-point detection methods for nonstationary time series*, Technometrics 50 (2008), pp. 40–52.
[10] M. Corduas and D. Piccolo, *Time series clustering and classification by the autoregressive metric*, Comput. Statist. Data Anal. 52 (2008), pp. 1860–1872.
[11] E. Berkes, E. Gombay, and L. Horváth, *Testing for changes in the covariance structure of linear processes*, J. Statist. Plann. Inference 139 (2009), pp. 2044–2063.
[12] H.C. Ho and T. Hsing, *Limit theorems for functionals of moving averages*, Ann. Probab. 25 (1997), pp. 1636–1669.