TESTING EQUALITY OF STATIONARY AUTOCOVARIANCES

Robert Lund*
Department of Mathematical Sciences
Clemson University
Clemson, SC 29634-0975
Lund@clemson.edu

Hany Bassily
Department of Mechanical Engineering
Clemson University
Clemson, SC 29634-0921

Brani Vidakovic
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

August 9, 2007

* Corresponding Author

Abstract: This paper studies tests for assessing whether two stationary and independent time series have the same dynamics — specifically, whether the autocovariances of both series coincide at all lags. Several frequency domain statistics previously proposed for this purpose are reviewed. A time domain statistic is then developed and investigated. The performance of these statistics are compared. As the previous literature on this topic resides almost exclusively within the spectral domain, it is perhaps surprising that the time domain test outperforms the frequency domain tests. Multivariate versions of the results are then investigated. The methods are applied in the analysis of temperatures and precipitations from two towns in the state of Georgia. Our interest here is driven by the need to identify a good climatological reference series for a given station. Efforts are made to keep the exposition rudimentary and expository.

Key words and phrases: Autocovariance, Multivariate Series; Periodogram, Short-Memory, Spectral Density.

1 Introduction.

This paper overviews testing procedures for assessing whether two stationary time series have the same dynamics (equivalent autocovariances). Such problems were posed by Coates and Diggle (1986), who studied the homogeneity of a single wheat price series over time and compared the wall thicknesses of a gas pipe at two different locations. Kakizawa *et al.* (1998) studied whether seismological series were more likely earthquakes or a nuclear tests. Classification procedures for functional magnetic resonance image series can be based on these methods and are important in the diagnosis of diseases (see Shumway and Stoffer 2006). Our interests for pursuing this problem lie with the development of climate reference stations (stations that can serve as surrogates for one and other). With a good reference station, one can calibrate new gauges and check for erroneous observations (erroneous observations are commonly generated by today's automatic weather observing systems).

It is advantageous to have two series with known equal autocovariances. For example, a $(1-\alpha)\times 100\%$ large sample confidence interval for an unknown mean μ of the stationary series $\{X_t\}$ from a sample of size n is

$$\overline{X} \pm \frac{z_{\alpha/2}}{n}\left[\gamma_X(0) + 2\sum_{h=1}^{n-1}\left(1-\frac{h}{n}\right)\gamma_X(h)\right]^{1/2}, \tag{1.1}$$

where z_α is the upper $(1-\alpha)$th quantile of the standard normal distribution and $\gamma_X(h)$ is the autocovariance of $\{X_t\}$ at lag h. A 'surrogate' estimator of $\gamma_X(\cdot)$, based on a series $\{Y_t\}$ that has the same autocovariances as $\{X_t\}$ could be used to gauge the interval width in (1.1). In climatology, two nearby towns may experience similar weather patterns and hence have similar autocovariances. Other applications of these methods involve astronomical classification problems, where say a star under study must be classified into one of several types with understood autocovariances (white dwarf, red giant, pulsar, etc.) and the need to calibrate a new machine to produce items similar to an old machine.

In mathematical terms, we explore whether or not the autocovariances of the two stationary series $\{X_t\}$ and $\{Y_t\}$ satisfy

$$\gamma_X(h) = \gamma_Y(h) \tag{1.2}$$

for each and every lag $h \geq 0$, where $\gamma_Y(h) = \text{Cov}(Y_{t+h}, Y_t)$. We take $\{X_t\}$ and $\{Y_t\}$ as independent; extensions to more than two series and to cases where $\{X_t\}$ and $\{Y_t\}$ are correlated will be commented upon later, but not pursued in detail here. Some of our work will be conducted in the frequency domain. This is because of the convenient and well-known properties of Fourier transforms of stationary series. In fact, two short-memory stationary series have the same autocovariances if and only if their spectral densities coincide at all frequencies (technically, almost every frequency in the Lebesgue sense). Specifically, (1.2) holds if and only if

$$f_X(\omega) = f_Y(\omega), \qquad \omega \in [0, 2\pi). \tag{1.3}$$

In (1.3), $f_X(\omega)$ and $f_Y(\omega)$ denote the spectral densities of $\{X_t\}$ and $\{Y_t\}$ at frequency ω, which are computed from the autocovariances via

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \gamma_X(h), \qquad f_Y(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \gamma_Y(h), \tag{1.4}$$

where $i = \sqrt{-1}$. In this article, we restrict ourselves to the short-memory case where

$$\sum_{h=0}^{\infty} |\gamma_X(h)| < \infty \qquad \text{and} \qquad \sum_{h=0}^{\infty} |\gamma_Y(h)| < \infty. \tag{1.5}$$

Short memory guarantees the existence (finiteness) of a spectral density at all frequencies and holds for any causal autoregressive moving-average (ARMA) time series.

In some senses, this paper reviews spectral domain methods for the problem. For the first time, we also present time domain approaches to the problem. Recent authors have considered spectral-based extensions to such applications as locally stationary and nonstationary series: Shumway (2006), Huang, Ombao and Stoffer (2004), Choi, Ombao, and Ray (2006), and Bengtsson and Cavanaugh (2006). We will concentrate on the simple stationary case as there is much to be done even in this setting.

The rest of this paper proceeds as follows. The next section motivates several simple test statistics for our problem. Section 3 provides short simulations for feel. Section 4 applies the results to 50 years of monthly temperature data from Athens and Atlanta, Georgia. Section 5 moves to the case of multivariate data. Section 6 revisits similarity of Athens and Atlanta weather with a bivariate analysis of temperatures and precipitations. Section 7 concludes with several comments.

2 Test Statistics.

We begin in the frequency domain. The spectral density of $\{X_t\}$ at frequency $\omega \in [0, 2\pi)$ is typically estimated by the periodogram, denoted by $I_X(\omega)$:

$$\hat{f}_X(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} X_t e^{-it\omega} \right|^2 := I_X(\omega) \tag{2.1}$$

(see Chapter 10 of Brockwell and Davis 1991). As $I_X(\omega)$ uniquely determines $\{X_t\}_{t=0}^{n-1}$ from its values at the Fourier frequencies $\omega_j = 2\pi j/n$ only, we focus exclusively on these Fourier frequencies. The conjugate symmetry relationships $I(-\omega) = I(\omega)$ and $I(2\pi - \omega) = I(\omega)$ reduce issues to consideration of $\omega_j = 2\pi j/n$ for $0 \leq j \leq \lfloor n/2 \rfloor$, where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x. For simplicity of exposition, we take n as an even integer henceforth so as to render $n/2$ whole.

For a collection of m distinct Fourier frequencies $\omega_1, \omega_2, \ldots, \omega_m$ such that $0 < \omega_1 < \ldots < \omega_m < \pi$, the $I_X(\omega_i)$ are asymptotically independent exponentially distributed random variables with means $E[I_X(\omega_j)] = f_X(\omega_j)$ (Proposition 10.3.2 in Brockwell and Davis 1991). The technical assumptions needed here are a linear process with independent innovations with a finite fourth moment:

$$X_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k}, \tag{2.2}$$

where $\{Z_t\}$ is zero mean independent and identically distributed noise with $E[Z_t^4] < \infty$.

The above arguments show that if $\{X_t\}$ and $\{Y_t\}$ have the same autocovariances, then R_ℓ defined by

$$R_\ell = \frac{I_X(\omega_\ell)/f_X(\omega_\ell)}{I_Y(\omega_\ell)/f_Y(\omega_\ell)} = \frac{I_X(\omega_\ell)}{I_Y(\omega_\ell)} \tag{2.3}$$

is distributed approximately as the ratio of two independent exponentially distributed random variables, with numerator and denominator both having a unit mean (henceforth referred to as standard). From the asymptotic independence of the periodogram at distinct frequencies, it follows that R_i and R_j are asymptotically independent when $1 \leq i \neq j \leq n/2$. In short, under a null hypothesis of equal autocovariances, $f_X(\omega) = f_Y(\omega)$ except on a subset of $(0, \pi)$ with Lebesgue measure zero (which we tacitly ignore) and $\{D_\ell\}_{\ell=1}^{n/2}$ is approximately an independent and identically distributed (IID) random sample from a distribution equivalent to that of E_1/E_2, where E_1 and E_2 are independent standard exponential variates. The ratio of two independent exponentially distributed random variables has the logistic probability density function

$$\frac{1}{(1+x)^2}, \qquad x \geq 0. \tag{2.4}$$

(see Whittwer 1984, Lawless 1982, Coates and Diggle 1986).

Several methods readily suggest themselves for a test of equal autocovariances. An empirical check merely constructs a probability plot of the ranked R_ℓ's against the distribution in (2.4). More definitive tests can, however, be devised.

The mean of the logistic distribution in (2.4) is infinite. However, one can construct a test statistic based on percentiles. For example, the number of R_ℓ's exceeding the fixed threshold M, over the indices $1 \leq \ell \leq n/2$, has approximately a binomial distribution with $n/2$ trials and success probability $(1+M)^{-1}$. A central limit approximation would reject equal autocovariances when the count $V = \#\{\ell : 1 \leq \ell \leq n/2 \text{ and } R_\ell > M\}$ exceeds

$$\frac{\ell}{2(M+1)} + z_\alpha \sqrt{\frac{\ell M}{2(M+1)^2}}.$$

Observe that tests based on $\{R_\ell\}$ are nonparametric in nature: a functional form for the spectral density is not specified.

Taking logarithms in (2.3) brings us to a method proposed by Coates and Diggle (1986). Let

$$D_\ell = \log(R_\ell) = \log(\hat{f}_X(\omega_\ell)) - \log(\hat{f}_Y(\omega_\ell)). \tag{2.5}$$

Then D_ℓ has the log-logistic probability density function

$$\frac{e^x}{(1+e^x)^2}, \qquad -\infty < x < \infty, \tag{2.6}$$

for each $1 \le \ell \le n/2$. Observe that $E[D_\ell] = 0$. The sample average of absolute deviations is hence

$$\overline{D} := \frac{2}{n} \sum_{\ell=1}^{n/2} |D_\ell|. \tag{2.7}$$

Large absolute values of $\overline{D}$ support rejection of equivalent time series dynamics. One can verify that $E[|D_\ell|] = \log(4)$ and $\mathrm{Var}(|D_\ell|) = \pi^2/3 - (\log 4)^2 \approx 1.368$. An αth level central limit theorem based hypothesis test hence rejects equal autocovariances when

$$\begin{aligned} |\overline{D}| &> \log 4 + z_\alpha \sqrt{\frac{\pi^2/3 - (\log 4)^2}{n/2}} \\ &\approx \log(4) + z_\alpha \sqrt{\frac{2.736}{n}}. \end{aligned} \tag{2.8}$$

The independence of $|D_i|$ and $|D_j|$ when $i \neq j$ assumed in the '$\overline{D}$ test' above holds exactly for each finite n when $\{X_t\}$ and $\{Y_t\}$ are Gaussian; Davis and Mikosch (1999) discuss extensions to non-Gaussian cases.

As the periodogram is an inconsistent estimator of the spectral density (Brockwell and Davis 1991), pooling the D_ℓ over all ℓ in some fashion seems natural. In fact, one could devise forms of the previous tests after smoothing the periodogram. Simulations were conducted to investigate such a practice with varying weighting schemes, but showed little practical improvements. Hence, we focus on tests involving the raw periodogram (unsmoothed) in the remainder of this article. It is also worth commenting that Kolmogorov-Smirnov type distances of the form

$$M = \max_{1 \leq \ell \leq n/2} |D_\ell|$$

exhibited very poor power in simulations and applications. This was also noted by Diggle and Coates (1986).

Another statistic can be devised from likelihood ratio principles. Under the null hypothesis, $\hat{f}_X(\omega_\ell)$ and $\hat{f}_Y(\omega_\ell)$ should be approximately exponentially distributed with mean $f_X(\omega_\ell) = f_Y(\omega_\ell)$. Using this along with the asymptotic independence of the periodogram at disjoint frequencies leads to the log likelihood ratio statistic, L_{Rat}, defined by

$$L_{\mathrm{Rat}} = \sum_{\ell=1}^{n/2} \log \left(\frac{\hat{f}_X(\omega_\ell)\hat{f}_Y(\omega_\ell)}{\left[\frac{\hat{f}_X(\omega_\ell)+\hat{f}_Y(\omega_\ell)}{2}\right]^2} \right) \tag{2.9}$$

One rejects equal autocovariances when L_{Rat} is too small. The large sample form of this test rejects equal autocovariances when

$$\frac{2}{n} L_{\mathrm{Rat}} < \mu - z_\alpha \sqrt{\frac{2\sigma^2}{n}}.$$

Here, μ and σ^2 are the mean and variance of

$$\log \left(\frac{4E_1E_2}{(E_1+E_2)^2} \right), \tag{2.10}$$

where E_1 and E_2 are independent unit mean exponential random variates. An involved computation provides $\mu = \ln(4) - 2 \approx -0.614$ and $\sigma^2 = 4 - \pi^2/3 + (4 - 8\gamma + 2\gamma^2) \approx 0.759$. Here, $\gamma \approx 0.577$ is Euler's constant.

For a finite n, the distribution of L_{Rat} is the n-fold convolution of IID summands, each summand distributed as the variate in (2.10). This form of this distribution is not overly important; in particular, it does not depend on the underlying spectral density of the series. One could in principle simulate percentiles if desired. As this test will not perform well in simulations, we will not explore this issue further here.

The L_{Rat} statistic can be modified to test for equality of autocovariances for M different independent stationary series; the statistic in (2.9) is merely modified to

$$L_{\text{Rat}} = \sum_{\ell=1}^{n/2} \log \left(\frac{\prod_{i=1}^{M} \hat{f}_i(\omega_\ell)}{\left(\frac{\sum_{i=1}^{M} \hat{f}_i(\omega_\ell)}{M} \right)^M} \right), \tag{2.11}$$

where $\hat{f}_i(\omega_\ell)$ is the estimated spectral density of the ith series, $1 \le i \le M$, at frequency ω_ℓ.

We now move to time domain tests. Time domain tests are based on Bartlett's asymptotic limit formula (see Chapter 7 of Brockwell and Davis 1991). Bartlett's result states that the first L sample autocovariances are asymptotically normal:

$$\begin{pmatrix} \hat{\gamma}(0) \\ \hat{\gamma}(1) \\ \hat{\gamma}(2) \\ \vdots \\ \hat{\gamma}(L) \end{pmatrix} \xrightarrow{D} N \left(\begin{pmatrix} \gamma(0) \\ \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(L) \end{pmatrix}, \frac{W}{n} \right) \tag{2.12}$$

as $n \to \infty$. Here, $\mathbf{W}$ is an $(L+1) \times (L+1)$ matrix with i,jth entry

$$W_{i,j} = \lim_{n \to \infty} n\text{Cov}(\hat{\gamma}(i), \hat{\gamma}(j)). \tag{2.13}$$

A tedious computation, explored in the multivariate case later, reveals

$$W_{i,j} = \sum_{r=-\infty}^{\infty} [E[X_t X_{t+i} X_{t+r} X_{t+r+j}] - \gamma(i)\gamma(j)] \tag{2.14}$$

for $0 \leq i, j \leq L$. Fourth moments arise in (2.14) since variances of sample covariances are involved. The term $E[X_t X_{t+i} X_{t+r} X_{t+r+j}]$ does not depend on time t since $\{X_t\}$ is strictly stationary by the assumptions in (2.2). Proposition 7.3.1 in Brockwell and Davis (1991) provides the equivalent form

$$W_{i,j} = (\eta - 3)\gamma(i)\gamma(j) + \sum_{k=-\infty}^{\infty} [\gamma(k)\gamma(k - i + j) + \gamma(k + j)\gamma(k - i)] \tag{2.15}$$

for $0 \leq i, j \leq L$. Here, $E[Z_t^4] = \eta\sigma^4$. When $\{Z_t\}$ is Gaussian, $\eta = 3$ and the first term in (2.15) is zero. The parameter L can be selected by the user. The technical conditions needed here are the same as in (2.2) (see Theorem 7.2.1 in Brockwell and Davis 1991 for discussion and variants).

The nonnegative definite estimator of $\gamma(h)$ is

$$\hat{\gamma}(h) = \sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X}), \tag{2.16}$$

where $\bar{X} = n^{-1} \sum_{t=1}^{n} X_t$. From (2.12), it follows that, under a null hypothesis of equal autocovariances,

$$\begin{pmatrix} \hat{\gamma}_X(0) - \hat{\gamma}_Y(0) \\ \hat{\gamma}_X(1) - \hat{\gamma}_Y(1) \\ \hat{\gamma}_X(2) - \hat{\gamma}_Y(2) \\ \vdots \\ \hat{\gamma}_X(L) - \hat{\gamma}_Y(L) \end{pmatrix} \xrightarrow{D} N\left(\mathbf{0}, \frac{2\mathbf{W}}{n}\right) \tag{2.17}$$

as $n \to \infty$. Hence,

$$\frac{n}{2}\widehat{\mathbf{\Delta}\gamma}^T \mathbf{W}^{-1}\widehat{\mathbf{\Delta}\gamma} \xrightarrow{D} \chi^2_{L+1}$$

as $n \to \infty$, where χ^2_{L+1} denotes a chi-squared random variable with $L+1$ degrees of freedom and $\widehat{\mathbf{\Delta}\gamma}$ is an $(L+1)\times 1$ vector with ith entry $\hat{\gamma}_X(i) - \hat{\gamma}_Y(i)$. Hence, we reject equal autocovariances at level α if the test statistic C defined by

$$C = \frac{n}{2}\widehat{\mathbf{\Delta}\gamma}^T \hat{\mathbf{W}}^{-1}\widehat{\mathbf{\Delta}\gamma} \tag{2.18}$$

exceeds the $(1-\alpha)$th quantile of the χ^2 distribution with $L+1$ degrees of freedom. Here, $\hat{\mathbf{W}}$ is an estimated version of $\mathbf{W}$ that employs the null hypothesis estimate $\hat{\gamma}(h) = 2^{-1}[\hat{\gamma}_X(h) + \hat{\gamma}_Y(h)]$ in place of $\gamma(h)$ in (2.15).

An important issue with this test lies with computation of the infinite sum in (2.15). One needs to estimate an infinite number of autocovariances to estimate the $W_{i,j}$'s. Moreover, $\hat{\gamma}(k)$ in (2.16) is a biased estimator of $\gamma(k)$, with the bias increasing for large k (alternatives are possible but usually problematic in other aspects). We have investigated several truncations of the infinite sum in (2.15) — particularly forms involving $n^{1/2}$ and $n^{1/3}$. These forms satisfy the hypotheses in Theorem A.1 in Berkes et al. (2006).

3 Simulation Performance.

To explore the methods in the last section, we have simulated two independent AR(1) series of length $n = 1024$. The series length is taken as a power of 2 so that the Fast Fourier Transform algorithm can be employed to rapidly compute the $I_X(\omega_\ell)$'s and $I_Y(\omega_\ell)$'s. Both series have the same autoregressive parameter ϕ, and this parameter is varied within the causal model range of $|\phi| < 1$. The innovations are chosen as normally distributed noise with a unit variance. Hence, the two series indeed have equivalent autocovariances. Table 1

Table 1: Empirical Type I Errors

ϕ	$\overline{D}$	L_{Rat}	C with $L = 5$	C with $L = 10$
-0.75	0.0496	0.0493	0.0194	0.0187
-0.50	0.0490	0.0497	0.0293	0.0259
-0.25	0.0492	0.0488	0.0506	0.0372
0.00	0.0503	0.0496	0.0693	0.0488
0.25	0.0504	0.0486	0.0518	0.0387
0.375	0.0503	0.0480	0.0401	0.0311
0.50	0.0513	0.0482	0.0326	0.0253
0.75	0.0500	0.0472	0.0218	0.0188

reports the proportion of times the $\overline{D}$, L_{Rat}, and C (with $L = 5$ and $L = 10$) statistics reject the hypothesis of equal autocovariances at level 5%. As the simulated series are Gaussian, we use $\eta = 3$ in (2.15). Each empirical Type I error is aggregated from 10,000 independent simulations; hence the empirical percentiles reported are quite accurate. In the computation of the C statistics, the infinite sum in (2.15) was truncated to the range $|k| \leq 5n^{1/2}$.

The empirical Type I errors in Table 1 show that the $\overline{D}$ and L_{Rat} tests reject approximately at rates close to 5% (perhaps slightly larger). The C tests are conservative for the most part, rejecting at a rate less than 5%, except for the pure white noise case ($\phi = 0$) when $L = 5$. The Type I errors for all statistics get closer to 5% for larger n; hence, any biases in the Type I errors are attributed to asymptotic approximations.

Table 2 reports empirical powers for the three tests in Table 1. Here, the first series is a Gaussian AR(1) process with autocorrelation coefficient ϕ and white noise variance of unity and the second series is a Gaussian MA(1) series with moving average parameter

Table 2: Empirical Powers

ϕ	$\overline{D}$	L_{Rat}	C with $L = 5$	C with $L = 10$
-0.75	1.000	1.000	1.000	1.000
-0.50	1.000	1.000	1.000	1.000
-0.25	0.786	0.723	1.000	1.000
0.00	0.050	0.050	0.069	0.049
0.25	0.055	0.053	0.140	0.094
0.375	0.080	0.075	0.672	0.516
0.50	0.232	0.202	0.999	0.998
0.75	0.999	0.998	1.000	1.000

set to $\theta = \pm\sqrt{\phi^2/(1-\phi^2)}$, where the plus sign is chosen if $\phi > 0$ and the minus sign when $\phi < 0$. This selection makes the variances of the two series identical and the lag-one autocovariances have the same sign; hence, difference between the two processes lie with differences of spectral shapes.

Surprisingly, the time domain test statistic C has the greatest power. In the case where $\phi = 0.375$, the power was about 10 times larger than that for the frequency domain methods. Our reason for surprise is that 1) past literature has almost exclusively focused on frequency domain tests and 2) the time domain tests seem more conservative in rejecting the null, as can be seen in Table 1. The L_{Rat} and $\overline{D}$ statistics have roughly the same power, with $\overline{D}$ performing slightly better uniformly in ϕ. The power of the C statistic appears symmetric in ϕ. Overall, it seems that time domain tests deserve further examination.

4 An Application.

This section studies monthly temperatures at Athens and Atlanta, Georgia, USA. Athens and Atlanta both lie in the Piedmont region of North Georgia, approximately 75 miles apart. Figure 1 plots monthly averaged temperatures (averaged over all days in month) for these two stations during the period Jan 1950 — Dec 2003. There are $n = 648$ observations for each series.

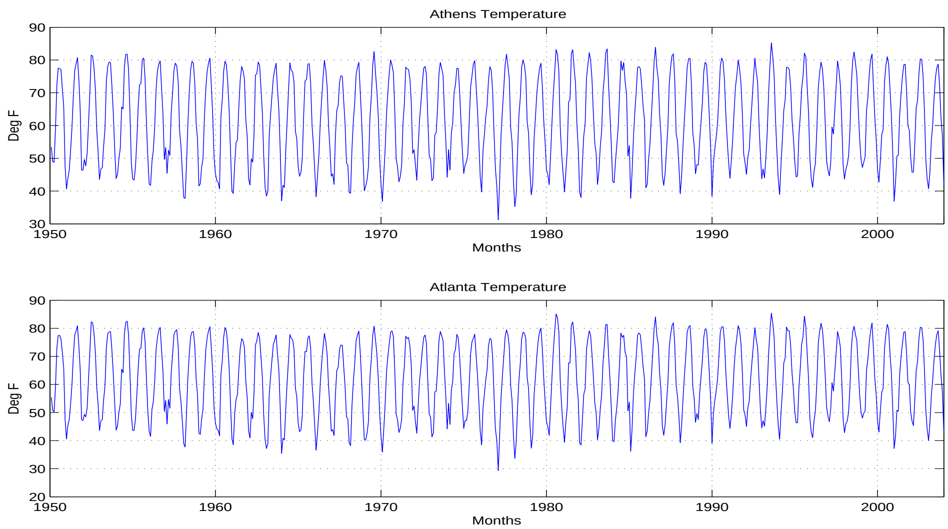

Figure 1: Athens and Atlanta monthly temperatures from Jan 1950 — Dec 2003.

As seasonality is clearly present in this data (winter temperatures are colder and more variable than summer temperatures), some preprocessing of the individual series is needed. To make zero mean unit variance series, we seasonally adjust these series by subtracting a monthly sample mean and then dividing by a monthly sample standard deviation. Lund *et al.* (1995) examines such seasonal adjustments for temperature series and demonstrates their stationarizing effects. Indeed, the seasonally standardized Athens and Atlanta temperatures pass checks for stationarity. The sample autocovariances of the Athens and Atlanta seasonally standardized series are displayed in Figure 2 and show strong positive coherence.

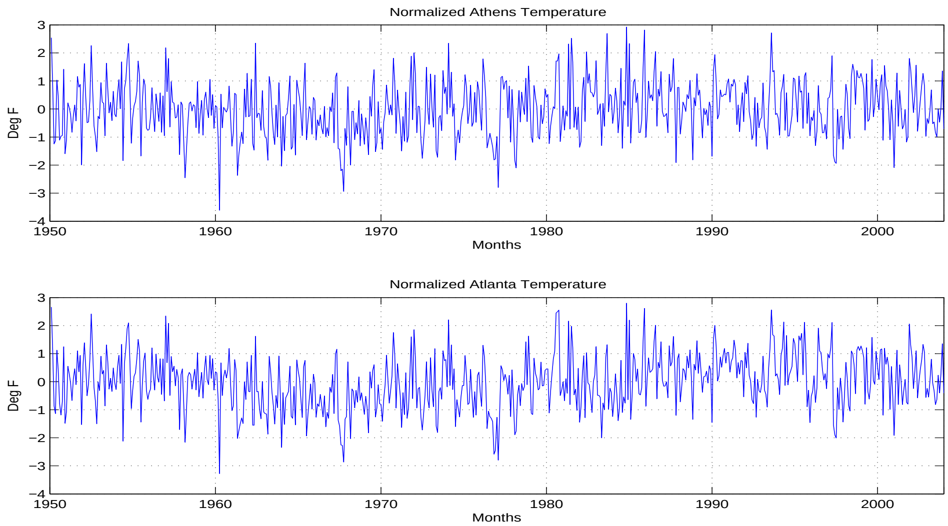

Figure 2: Seasonally standardized Athens and Atlanta temperatures.

The dashed lines are 95% confidence bounds (pointwise) for white noise.

Figures 1 and 2 support the local folklore that Athens and Atlanta enjoy similar weather. The three equality of autocovariance statistics and their p-values (in parentheses) are $\overline{D} = 0.5484$, (1.000), $L_{\mathrm{Rat}} = -48.6324$ (1.000), $C = 3.8915$ (0.140) with $L = 5$ and $C = 4.1283$ (0.260) with $L = 10$. As all tests give the same conclusion, Athens and Atlanta temperatures appear to have equal standardized autocovariance functions. Since the seasonal mean and standard deviations from the two sites are also very similar (we will not delve into tests for equality of means further here), the two towns indeed enjoy similar temperatures.

5 Multivariate Versions.

This section examines tests for equality of autocovariances when $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ are d-dimensional, where $d \geq 2$, zero mean stationary series. This problem was considered in Kakizawa et al. (1998), where issues of clustering many series into similar groups was pursued. There, computing the probability that two series were misclassified as being

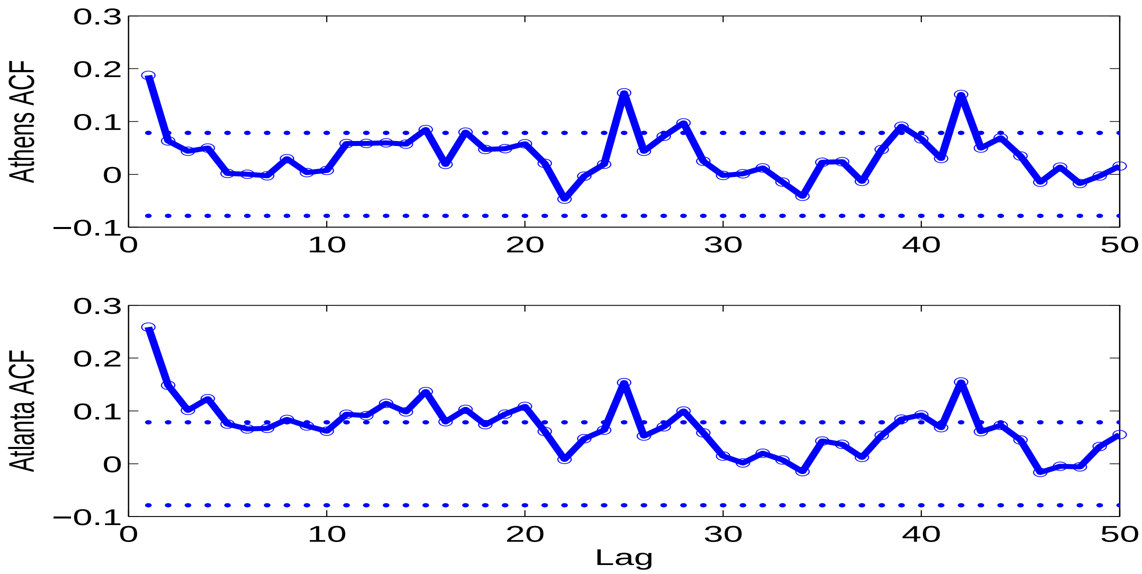

Figure 3: Athens and Atlanta sample autocovariances.

different was involved. Because of this, our goal is to develop simple level α tests for whether (or not) two multivariate time series have the same autocovariances. We will consider both spectral and time domain tests for this purpose.

For a spectral test, we now develop a multivariate version of the $\overline{D}$ test of Section 2. Because of its poor performance in the univariate case, likelihood ratio statistics are not pursued in multivariate settings. The periodogram of $\{\mathbf{X}_t\}$ at frequency $\omega_j = 2\pi j/n$ is

$$\mathbf{I}_X(\omega_j) = \mathbf{J}_X(\omega_j)\mathbf{J}_X^*(\omega_j),$$

where * denotes complex conjugation and

$$\mathbf{J}_X^*(\omega_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \mathbf{X}_t e^{-it\omega_j} \tag{5.1}$$

(similar definitions apply to $\{\mathbf{Y}_t\}$).

One multivariate version of the $\overline{D}$ statistic in (2.7) merely involves log-determinants:

$$\overline{D} = \frac{2}{n}\sum_{\ell=1}^{n/2} |\log(\det(\hat{\mathbf{f}}_X^s(\omega_j))) - \log(\det(\hat{\mathbf{f}}_Y^s(\omega_j)))|, \tag{5.2}$$

where $\hat{\mathbf{f}}_X^s(\omega_j)$ and $\hat{\mathbf{f}}_Y^s(\omega_j)$ denote the uniformly smoothed spectral estimates

$$\hat{\mathbf{f}}_X^s(\omega_j) = \frac{\sum_{k=-M}^{M} \mathbf{I}_X(\omega_{j+k})}{2M+1} \quad \text{and} \quad \hat{\mathbf{f}}_Y^s(\omega_j) = \frac{\sum_{k=-M}^{M} \mathbf{I}_Y(\omega_{j+k})}{2M+1}. \tag{5.3}$$

In the multivariate setting, one must smooth the raw periodogram before comparing estimates of the spectral densities. Mathematically, this is evident from $\det(\mathbf{I}_X(\omega_j)) = 0$. In fact, we impose the condition $2M+1 \geq d$ upon the smoothing window width M; if this is not satisfied, then $\hat{\mathbf{f}}_X^s(\omega_j)$ and $\hat{\mathbf{f}}_Y^s(\omega_j)$ will have an infinite variance and 'regularity issues' will arise.

Under the null hypothesis that the autocovariance functions of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ agree, the summands in (5.2) are independent with zero mean and variances that are functions of M only (in fact, $\hat{\mathbf{f}}_X^s(\omega_\ell)$ has a Wishart type distribution). The null hypothesis of equivalent autocovariances is hence rejected if $|\overline{D}|$ is too large:

$$|\overline{D}| > \mu_M + z_\alpha \frac{\sigma_M}{\sqrt{n/2}}$$

Table 3 lists values of μ_M and σ_M^2 simulated from 100,000 independent simulations. It is perhaps possible to get explicit expressions for these quantities akin to (2.8), but we will not pursue such a computation here as the spectral tests do not seem to perform as well as the time domain tests constructed below.

We now consider time domain tests. Such tests are based on Bartlett's asymptotic formula for multivariate stationary series. Specifically, let

$$\mathbf{\Gamma}(h) = E[\mathbf{X}_{t+h}\mathbf{X}_t'] = \{\gamma_{i,j}(h)\}_{i,j=1}^d \tag{5.4}$$

Table 3: Constants for multivariate $\overline{D}$ test

M	μ_M	σ^2_M
1	1.1400	1.3806
2	0.8013	1.1190
3	0.6553	1.0473
4	0.5687	0.9822
5	0.5102	0.9708
7	0.4342	0.9593
10	0.3671	0.9554

and

$$\hat{\mathbf{\Gamma}}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(\mathbf{X}_{t+h} - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})' = \{\hat{\gamma}_{i,j}(h)\}_{i,j=1}^{d} \tag{5.5}$$

denote theoretical and sample autocovariances, respectively, of a multivariate series $\{\mathbf{X}_t\}$. Here, $\bar{\mathbf{X}} = n^{-1}\sum_{t=1}^{n}\mathbf{X}_t$. Bartlett's result states that the collection of sample autocovariances $\hat{\gamma}_{i,j}(h)$ are jointly asymptotically normal with

$$\lim_{n\to\infty} n\mathrm{Cov}(\hat{\gamma}_{i,j}(p), \hat{\gamma}_{k,\ell}(q)) = \sum_{r=-\infty}^{\infty}\{E[X_{t,i}X_{t+p,j}X_{t+r,k}X_{t+r+q,\ell}] - \gamma_{i,j}(p)\gamma_{k,\ell}(q)\}\,. \tag{5.6}$$

We derive (5.6) below as it is not in Anderson (1971), Brockwell and Davis (1991), or Fuller (1996). Moreover, extracting it from Hannan (1970) takes some effort. The technical assumptions needed here are merely that $\{\mathbf{X}_t\}$ has the linear process representation

$$\mathbf{X}_t = \sum_{k=-\infty}^{\infty}\mathbf{\Psi}_k\mathbf{Z}_{t-k}, \tag{5.7}$$

where $\{\mathbf{Z}_t\}$ is independent and identically distributed zero mean noise with a finite fourth moment and that $\sum_{k=-\infty}^{\infty} |\mathbf{\Psi}_k| < \infty$ (in a component by component sense). These assumptions ensure that $\{\mathbf{X}_t\}$ is fourth order stationary (in fact $\{\mathbf{X}_t\}$ is strictly stationary), which implies that $E[X_{t,i}X_{t+p,j}X_{t+r,k}X_{t+r+q,\ell}]$ does not depend on t.

In the case where $\{\mathbf{X}_t\}$ is Gaussian, (5.6) simplifies to

$$\lim_{n\to\infty} n\mathrm{Cov}(\hat{\gamma}_{i,j}(p), \hat{\gamma}_{k,\ell}(q)) = \sum_{r=-\infty}^{\infty} \left[\gamma_{i,k}(r)\gamma_{j,\ell}(r-p+q) + \gamma_{i,\ell}(r+q)\gamma_{j,k}(r-p)\right]. \quad (5.8)$$

The statistic analogous to (2.18) is simply

$$C = \frac{n}{2}\widehat{\mathbf{\Delta\Gamma}}^T \hat{\mathbf{W}}^{-1}\widehat{\mathbf{\Delta\Gamma}}. \quad (5.9)$$

Here, $\widehat{\mathbf{\Delta\Gamma}}$ is an $Ld^2 + d(d-1)/2$ dimensional vector whose elements are aggregated from sample autocovariances at lags $0, 1, \ldots, L$ as follows. More specifically, define

$$\eta(h)_{i,j} = \hat{\gamma}_X(h)_{i,j} - \hat{\gamma}_Y(h)_{i,j}$$

for $h \geq 0$ and $1 \leq i, j \leq d$. The first $d(d-1)/2$ elements of $\widehat{\mathbf{\Delta\Gamma}}$ are obtained by stacking the lag $h = 0$ components of η in the order $\eta(0)_{1,1}, \ldots, \eta(0)_{1,d}; \eta(0)_{2,2}, \ldots, \eta(0)_{2,d}; \ldots; \eta(0)_{d,d}$. The next d^2 elements of $\widehat{\mathbf{\Delta\Gamma}}$ are simply the lag $h = 1$ components of η stacked in the usual 'row first column second order' of $\eta(1)_{1,1}, \ldots, \eta(1)_{1,d}; \ldots; \eta_{d,1}(1), \ldots, \eta_{d,d}(1)$. The remaining components of $\widehat{\mathbf{\Delta\Gamma}}$ are the lag $2, \ldots, L$ components of η stacked in a row first column second order (there are d^2 components for each lag). As

$$\hat{\gamma}_{i,j}(0) = \hat{\gamma}_{j,i}(0),$$

we cannot use entries from 'both above and below the main diagonal' for lag $h = 0$ — this would make the covariance matrix of $\widehat{\mathbf{\Delta\Gamma}}$ singular. The quantity $\hat{\mathbf{W}}$ is simply the covariance matrix of $\widehat{\mathbf{\Delta\Gamma}}$, computed via estimating quantities in (5.6). Fourth moments are estimated as

$$\widehat{E}[X_{0,i}X_{p,j}X_{r,k}X_{r+q,\ell}] = n^{-1}\sum_{t\in S} X_{t,i}X_{t+p,j}X_{t+r,k}X_{t+r+q,\ell},$$

where the set S contains all indices t such that $t, t+p, t+r$, and $t+r+q$ all lie in $\{1,\ldots,n\}$. The infinite sum in (5.6) or (5.8) is truncated at $\pm n^{1/3}$. Specifically, we use

$$\widehat{\mathrm{Cov}}(\hat{\gamma}_{i,j}(p), \hat{\gamma}_{k,\ell}(q)) = n^{-1}\sum_{|r|\leq n^{1/3}} [\hat{\gamma}_{i,k}(r)\hat{\gamma}_{j,\ell}(r-p+q) + \hat{\gamma}_{i,\ell}(r+q)\hat{\gamma}_{j,k}(r-p)] \qquad (5.10)$$

as an estimator of the components in $\mathbf{W}$. In (5.10), we use

$$\hat{\gamma}_{i,j}(h) = \frac{1}{2}[\hat{\gamma}_{X;i,j}(h) + \hat{\gamma}_{Y;i,j}(h)]$$

as the estimate of $\gamma_{i,j}(h)$ in (5.10) under the null hypothesis that the autocovariances of $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ agree. Here, the extra subscripts of X and Y merely refer to sample covariances computed from $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ respectively.

The derivation of (5.6) and (5.8) proceeds as follows for fixed i, j, k, ℓ. Brockwell and Davis (1991, Section 7.3) prove the joint asymptotic normality under the hypothesized conditions. They also show that

$$\lim_{n\to\infty} n\mathrm{Cov}(\hat{\gamma}_{i,j}(p), \hat{\gamma}_{k,\ell}(q)) = \lim_{n\to\infty} n\mathrm{Cov}(\hat{\gamma}^*_{i,j}(p), \hat{\gamma}^*_{k,\ell}(q)),$$

where $\hat{\gamma}^*_{i,j}(p)$ is the unbiased edge-effect corrected estimator

$$\hat{\gamma}^*_{i,j}(p) = n^{-1}\sum_{t=1}^{n} X_{t,i}X_{t+p,j}. \qquad (5.11)$$

Equation (5.11) now provides

$$\text{Cov}(\hat{\gamma}^*_{i,j}(p), \hat{\gamma}^*_{k,\ell}(q)) = n^{-2} \sum_{t=1}^{n} \sum_{s=1}^{n} M(t-s), \tag{5.12}$$

where

$$M(t-s) = E[X_{t,i} X_{t+p,j} X_{s,k} X_{s+q,\ell}] - E[X_{t,i} X_{t+p,j}] E[X_{s,k} X_{s+q,\ell}]. \tag{5.13}$$

The fourth order stationarity of $\{\mathbf{X}_t\}$ implies that the right hand side of (5.13) is indeed a function of $t-s$ only. From (5.7) and the assumed fourth order moments on $\{\mathbf{Z}_t\}$, one can verify that $\sum_{h=-\infty}^{\infty} |M(h)| < \infty$. Combining the above relations and applying the dominated convergence theorem now gives

$$\begin{aligned} \lim_{n\to\infty} n\text{Cov}(\hat{\gamma}_{i,j}(p), \hat{\gamma}_{k,\ell}(q)) &= \lim_{n\to\infty} \sum_{|h|<n} \left(1 - \frac{h}{n}\right) M(h) \\ &= \sum_{h=-\infty}^{\infty} M(h). \end{aligned} \tag{5.14}$$

Combining (5.13) and (5.14) establishes (5.6). The arguments on page 209 of Hannan (1970) provide

$$\begin{aligned} E[X_{t,i} X_{t+p,j} X_{s,k} X_{s+q+\ell}] &= \gamma_{i,j}(p)\gamma_{k,\ell}(q) + \gamma_{i,k}(t-s)\gamma_{j,\ell}(t-s-p+q) \\ &+ \gamma_{i,\ell}(t-s+q)\gamma_{j,k}(t-s-p) \end{aligned}$$

for Gaussian processes. Using this in (5.6) establishes (5.8).

The multivariate time domain test simply rejects equality of autocovariances when C is too large to be explained by the chance variation in a chi-squared distribution with $d(d-1)/2 + d^2 L$ degrees of freedom.

A short simulation study was conducted to gain some feel for the above methods. We first simulated two independent AR(1) series $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ of dimension two governed by the same propagation matrix $\mathbf{\Phi}$ and white noise variance matrix $\mathbf{\Sigma}$. We examined four values of $\mathbf{\Phi}$, namely

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0.8 & -0.3 \\ 0.9 & -0.4 \end{bmatrix}, \quad \mathbf{\Phi}_2 = \begin{bmatrix} 0.55 & -0.15 \\ 0.15 & 0.45 \end{bmatrix}, \mathbf{\Phi}_3 = \begin{bmatrix} 0.2 & 0.1 \\ -0.1 & 0.3 \end{bmatrix}, \quad \text{and} \quad \mathbf{\Phi}_4 = \begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{bmatrix},$$

and two values of $\mathbf{\Sigma}$:

$$\mathbf{\Sigma}_1 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}, \quad \text{and} \quad \mathbf{\Sigma}_2 = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}.$$

These choices all induce a causal model and are meant to provide a range of different autocorrelation structures; $\mathbf{\Phi}_4$ entails the case of white noise.

Table 4 reports empirical Type I errors for the multivariate $\overline{D}$ and C tests for several values of the parameters M and L. The empirical Type I errors are aggregated from 10,000 independent simulations in a level 95% test. The Type I errors of the $\overline{D}$ test are much closer to 5% with $M = 5$ than for $M = 1$. The Type I errors of all tests seem conservative (less than 0.05) for the most part.

Table 5 reports empirical powers for multivariate $\overline{D}$ and C tests for several values of the parameters M and L. Here, the $\{\mathbf{X}_t\}$ series is taken as AR(1) as above and $\{\mathbf{Y}_t\}$ is a $d = 2$ dimensional MA(1) series. We have selected the first order moving-average matrix and white noise coefficients so that the variances of both series — in particular, $\mathbf{\Gamma}(0)$ — are identical; hence, true differences in autocovariances must occur at lags 1 and greater. The powers were aggregated from 10,000 independent simulations in a level 95% test.

The powers in Table 5 again favor the time domain tests, drastically so in some cases. The lag parameter $L = 5$ is performing better than $L = 10$; the smoothing parameter $M = 5$ gives greater powers than $M = 1$. Again, the overall conclusion is evident: time domain versions of these tests merit more attention.

Table 4: Empirical Type I Errors

Case	$\overline{D}$ with $M=1$	$\overline{D}$ with $M=5$	C with $L=5$	C with $L=10$
$\mathbf{\Phi}_1, \mathbf{\Sigma}_1$	0.0389	0.0465	0.0427	0.0454
$\mathbf{\Phi}_2, \mathbf{\Sigma}_1$	0.0171	0.0435	0.0446	0.0410
$\mathbf{\Phi}_3, \mathbf{\Sigma}_1$	0.0285	0.0485	0.0464	0.0438
$\mathbf{\Phi}_4, \mathbf{\Sigma}_1$	0.0504	0.0437	0.0451	0.0390
$\mathbf{\Phi}_1, \mathbf{\Sigma}_2$	0.0391	0.0455	0.0456	0.0422
$\mathbf{\Phi}_2, \mathbf{\Sigma}_2$	0.0163	0.0455	0.0428	0.0411
$\mathbf{\Phi}_3, \mathbf{\Sigma}_2$	0.0284	0.0499	0.0426	0.0388
$\mathbf{\Phi}_4, \mathbf{\Sigma}_2$	0.0522	0.0492	0.0443	0.0451

6 Application Rejoinder.

We return to the Athens/Atlanta climate data comparison with rainfall data. Figure 4 displays the total monthly precipitation at these two stations over the same period of record as the temperatures. Although the mean rainfall cycle has a weaker seasonal cycle than that of mean temperatures, a seasonal mean is still present with Fall months being driest and Spring wettest. Figure 5 shows seasonally standardized versions of these series; they pass the stationarity tests of Lund *et al.* (1995). In short, we now have a 54 year monthly segment of a $d = 2$-variate stationary series.

We first compare the univariate seasonally standardized precipitation series from Athens and Atlanta. Proceeding as in Section 4, we obtain $\overline{D} = 1.0135$ (p-value is approximately 1.000); $L_{\text{Rat}} = -122.05$ (p-value is approximately 1.000); $C = 2.6257$ with $L = 5$ (p-value $\approx$ 0.8541); and $C = 4.7369$ with $L = 10$ (p-value ≈ 0.9432). Hence, marginally, the conclusion is that the seasonally adjusted precipitations have the same autocovariance structures.

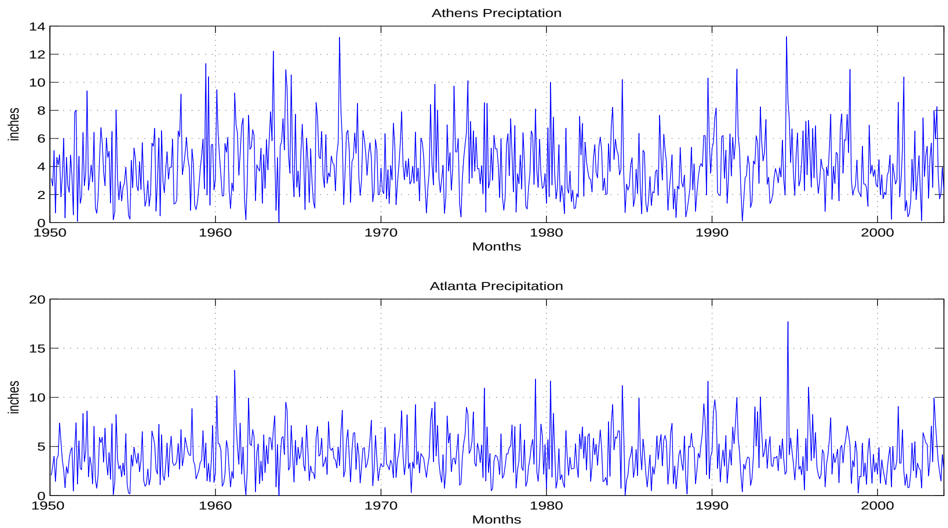

Figure 4: Athens and Atlanta raw monthly precipitations.

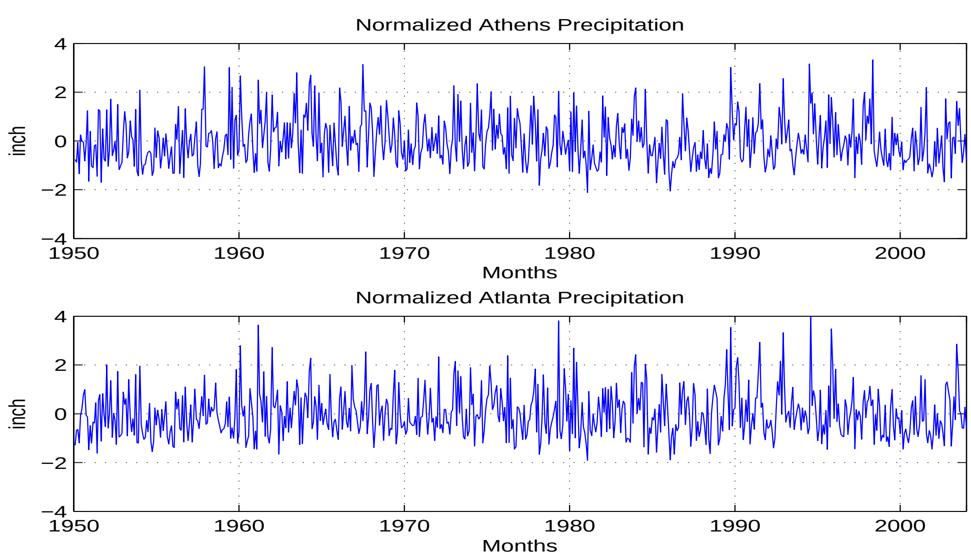

Figure 5: Monthly normalized Athens and Atlanta precipitations.

Table 5: Empirical Powers

Case	$\overline{D}$ with $M = 1$	$\overline{D}$ with $M = 5$	C with $L = 5$	C with $L = 10$
$\mathbf{\Phi}_1, \mathbf{\Sigma}_1$	0.0854	1.000	1.000	1.000
$\mathbf{\Phi}_2, \mathbf{\Sigma}_1$	0.361	1.000	1.000	1.000
$\mathbf{\Phi}_3, \mathbf{\Sigma}_1$	0.0411	0.154	0.653	0.465
$\mathbf{\Phi}_4, \mathbf{\Sigma}_1$	0.0523	0.0477	0.0451	0.0402
$\mathbf{\Phi}_1, \mathbf{\Sigma}_2$	0.0420	1.000	1.000	1.000
$\mathbf{\Phi}_2, \mathbf{\Sigma}_2$	0.357	1.000	1.000	1.000
$\mathbf{\Phi}_3, \mathbf{\Sigma}_2$	0.0390	0.162	0.826	0.664
$\mathbf{\Phi}_4, \mathbf{\Sigma}_2$	0.0534	0.0486	0.0464	0.0426

Moving to a $d = 2$ dimensional analysis of the standardized series, we obtain multivariate statistics of $\overline{D} = 0.7138$ with $M = 1$ (p-value of approximately 1.000); $\overline{D} = 0.2955$ with $M = 5$ (p-value of approximately 1.000); $C = 14.773$ with $L = 5$ (p-value ≈ 0.9206); and $C = 22.5796$ with $L = 10$ (p-value ≈ 0.9956). All results support the hypothesis that the two 2-dimensional series have the same autocovariance structures. Athens and Atlanta indeed enjoy similar weather.

7 Concluding Comments and Future Work.

The superior performance of time-domain tests for equality of autocovariances was unexpected. In hindsight, some rough performance of frequency domain statistics should have been expected as the periodogram is an inconsistent estimator of the spectral density. The time domain tests need further exploration. In particular, issues related to truncating the infinite sum in (2.15) should be explored further. Fuller (1996, Theorem 6.2.2) provides an expression for the asymptotic bias of $E[\hat{\gamma}(h)]$. One could in principle use this result to derive a class of reasonable truncation schemes to use in (2.15).

Extensions of the general methods are also needed. For example, one may have multiple realizations of one of the series, or the length of the two series being compared might be different. Also, one may be interested in comparing two correlated series. Here, time-domain methods seem promising. Specifically, it seems possible to derive a form of (5.6) or (5.8) under the case where $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_t\}$ are correlated; it is not clear to us how to proceed in frequency domain settings. Finally, one may wish to incorporate some information about the series if it is known. For example, if both series have an AR(1) structure, how should one proceed? (Azzalini 1984 considers this problem for AR(1) series). For time domain tests, we would simply derive explicit formulas for the $W_{i,j}$ imposing the AR(1) autocovariance structure (see Example 7.2.3 of Brockwell and Davis 1991).

Acknowledgements. Support from NSF Grant DMS 030447, NSA Grant E-24-60R, and NSF Grant DMS 0307057 are acknowledged. The authors thank Carol Krautmann for securing the temperature data studied. We thank Anand Vidyashankar for many insightful conversations on the topic.

References

[1] Anderson, T. W. (1971). *The Statistical Analysis of Time Series*, John Wiley and Sons, New York City.

[2] Azzalini, A. (1984). Estimation and hypothesis testing for collections of autoregressive time series, *Biometrika*, **71**, 85-90.

[3] Brockwell. P. J., and Davis, R. A. (1991). *Time Series: Theory and Methods*, Second Edition, Springer-Verlag, New York.

[4] Berkes, I., Horváth, L., Kokoszka, P., and Q-M. Shao (2006). On discriminating between long-range dependence and changes in the mean, *Annals of Statistics*, **34**, 1140-1165.

[5] Choi, H., Ombao, H., and Ray, B. (2006). Online change-point detection methods for non-stationary time series, preprint.

[6] Coates, D. S., and Diggle, P. (1986). Tests for comparing two estimated spectral densities, *Journal of Time Series Analysis*, **7**, 7-20.

[7] Davis, R. A., and Mikosch, T. (1999). The maximum of a periodogram of a non-Gaussian sequence, *Annals of Probability*, **27**, 522–536.

[8] Diggle, P., and Fisher, N. I. (1991). Nonparametric comparison of cumulative periodograms, Journal of the Royal Statistical Society, Series C, **40**, 423-434.

[9] Diggle P. J., and Wasel, I. (1997). Spectral analysis of replicated biomedical time series, *Journal of the Royal Statistical Society, Series C*, **46**, 31-71.

[10] Fuller, W. A. (1996). *Introduction to Statistical Time Series*, Second Edition, John Wiley & Sons, New York City.

[11] Johnson, N.L, Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Second Edition, Wiley, New York.

[12] Hannan, E. J. (1970). *Multiple Time Series*, John Wiley and Sons, New York City.

[13] Kakizawa, Y., Shumway, R., and Taniguchi, M. (1998). Discrimination and clustering for multivariate time series, *Journal of the American Statistical Association*, **93**, 328-340.

[14] Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*, Wiley, New York.

[15] Lee, J., and Lund, R. B. (2004). Revisiting simple linear regression with autocorrelated errors, *Biometrika*, **91**, 240-245.

[16] Maharaj, E. A. (2002). Comparison of non-stationary time series in the frequency domain, *Computational Statistics and Data Analysis*, **40**, 131-141.

[17] Lund, R. B., Hurd, H., Bloomfield, P., and Smith, R. L. (1995). Climatological time series with periodic correlation, *Journal of Climate*, **11**, 2787-2809.

[18] Shumway, R. H., and Stoffer, D. L. (2006). *Time Series Analysis and its Applications*, Second Edition, Springer, New York.

[19] Shumway, R. H. (2006). Time-frequency clustering and discrimination analysis, preprint.

[20] Whittwer, G. (1986). On the distribution of the periodogram for the stationary random sequences (With discussion). *Math. Opera. und Stat.,* **17**, 201–219.