TESTING EQUALITY OF AUTOCOVARIANCE FUNCTIONS

ROBERT LUND
Department of Mathematical Sciences
Clemson University
O-106 Martin Hall, Box 340975
Clemson, SC 29634-0975

BRANI VIDAKOVIC
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332

ANAND N. VIDYASHANKAR
Department of Statistical Science
Cornell University
Ithaca, NY 14853

September 22, 2004

Abstract: This paper introduces a simple frequency domain test to discern whether two stationary time series have the same autocovariance function. The driving idea is that two stationary short-memory autocovariances coincide over all lags if and only if the corresponding spectral densities agree. As the spectral density is easily estimated via the periodogram, and the asymptotics of the periodogram are well known, a statistic based on the log-ratio of periodogram ordinates is proposed and explored. An application of the method is given. The exposition is made accessible to a general audience, although rudimentary familiarity with spectral densities and periodograms is assumed.

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

1 Introduction.

The modern time series analyst is savvy in both time and frequency domains. The viewpoints are complementary and provide different insights, with some issues being more transparent in the frequency domain (proving that the sum of two independent ARMA series is again ARMA for example), while other issues are more succinctly cast within the time domain (recursive ARMA forecasting for example). This short paper presents a frequency domain test for discerning whether (or not) two series have the same autocovariance function.

Suppose that $\{X_t\}$ and $\{Y_t\}$ are stationary series with finite second moments and autocovariances $\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h})$ and $\gamma_Y(h) = \mathrm{Cov}(Y_t, Y_{t+h})$ at lag h. We take $\{X_t\}$ and $\{Y_t\}$ as independent; extensions to more than two series and to correlated cases may merit consideration, but we keep the issue simple here. Our research objective is to develop a test for whether or not

$$\gamma_X(h) = \gamma_Y(h) \tag{1.1}$$

for all integral lags h. To ensure existence of a spectral density, it is assumed that $\{X_t\}$ and $\{Y_t\}$ have short-memory in that

$$\sum_{h=0}^{\infty} |\gamma_X(h)| < \infty \qquad \text{and} \qquad \sum_{h=0}^{\infty} |\gamma_Y(h)| < \infty; \tag{1.2}$$

this is the case for any causal autoregressive moving-average (ARMA) time series. Under short-memory, the spectral density of $\{X_t\}$, denoted by $f_X(\omega)$ at frequency $\omega \in [0, 2\pi)$, exists (is finite) and is given by

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-ih\omega} \gamma_X(h), \tag{1.3}$$

where $i = \sqrt{-1}$; a similar formula holds for $f_Y(\cdot)$.

Testing for equality of autocovariances arises in inference and quality control settings. For example, a $(1-\alpha) \times 100\%$ large sample confidence interval for an unknown mean $\mu \equiv E[X_t]$ from the sample $X_0, \ldots, X_{n-1}$ is

$$\overline{X} \pm z_{\alpha/2} \left[n^{-1} \left\{ \gamma_X(0) + 2 \sum_{h=1}^{n-1} (1 - h/n)\, \hat{\gamma}_X(h) \right\} \right]^{1/2}, \tag{1.4}$$

where z_α is the upper $(1-\alpha)$th quantile of the standard normal distribution, $\bar{X} = \frac{1}{n}\sum_{t=0}^{n-1} X_t$ is the sample series average, and $\hat{\gamma}_X(h)$ is a suitably good estimate of $\gamma_X(h)$. In many settings, such as application of (1.4), a surrogate estimate of $\gamma_X(\cdot)$ would be useful and is available upon suitable 'digging'. Such is the case in climate time series where two adjacent towns typically experience similar weather and hence enjoy similar autocovariances. The notion of a reference series in climatology is developed in Easterling and Peterson (1995) and the references therein. If $\{X_t\}$ and $\{Y_t\}$ have equivalent autocovariances, then one can of course substitute $\hat{\gamma}_Y(h)$ for $\hat{\gamma}_X(h)$. Besides confidence intervals for mean series values, autocovariances are key quantities for inferences involving linear trends (cf. Lee and Lund 2004) and in forecasting future series values (cf. Brockwell and Davis 1991). Another surrogate-driven example arises in astronomical classification problems. For example, if the sampled autocovariances of a star under study resemble those for a known white dwarf star, then one has evidence that the star under study may also be a white dwarf. Yet another use of surrogate autocovariances lies in quality control for correlated processes. Here, a new machine may need to be calibrated to produce items similar to an old machine.

Thus, a fundamental statistical question tests whether two series are drawn from processes with the same autocovariances. The objective of this paper is to present a simple frequency domain test for this purpose. The rest of this paper proceeds as follows. The next section motivates a test statistic for equality of autocovariance functions. Section 3 provides a simulation example for general feel; Section 4 closes with an application to two temperature series from stations at Athens and Atlanta, GA, USA.

2 The Test Statistic.

The spectral density of $\{X_t\}$ at frequency $\omega \in [0, 2\pi)$ is typically estimated by the periodogram, denoted by $I_X(\omega)$ and defined as

$$I_X(\omega) = \frac{1}{2\pi n} \left| \sum_{t=0}^{n-1} X_t e^{-it\omega} \right|^2 \tag{2.1}$$

(cf. 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 $I(2\pi - \omega) = I(\omega)$ further reduce the problem 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 exponential random variables with means $E[I_X(\omega_j)] = f_X(\omega_j)$ (cf. Proposition 10.3.2 in Brockwell and Davis 1991). The simple crux of this paper can now be easily stated: if $\{X_t\}$ and $\{Y_t\}$ have the same autocovariances, then $\{R_\ell\}_{\ell=1}^{n/2}$ defined by

$$R_\ell = \frac{I_X(\omega_\ell)/f_X(\omega_\ell)}{I_Y(\omega_\ell)/f_Y(\omega_\ell)} = \frac{I_X(\omega_\ell)}{I_Y(\omega_\ell)} \tag{2.2}$$

is distributed approximately as the ratio of two independent exponentially distributed random variables, with numerator and denominator both having unit mean (henceforth referred to as standard). From asymptotic independence of the periodogram, it follows that R_i and R_j are asymptotically independent when $i \neq j$. Observe that a functional form for the spectral density is not required to compute the R_ℓ's. Equation (2.2) and its log transformation will serve as key elements in our analysis.

We reiterate the well-known fact that the 'raw' periodogram $I_X(\omega)$ is an inconsistent estimator of the spectral density $f_X(\omega)$; specifically, $\text{var}(I_X(\omega_j))$ does not tend to zero as $n \to \infty$. Wahba (1980), states "it will be hopelessly wiggly even when $f(\omega)$ is a smooth function" and $n \to \infty$. However, under vast generality (cf. Theorem 5.2.6 of Brillinger 1981; Section 10.4 of Brockwell and Davis; Chapter 3 of Shumway and Stoffer 2000),

$$I_X(\omega_\ell) \overset{iid}{\approx} f_X(\omega_\ell)E_{X,\ell}, \tag{2.3}$$

where $\overset{iid}{\approx}$ is interpreted as approximately independent and identically distributed over distinct frequencies and $E_{X,\ell}$ denotes a standard exponential random variable. For $\omega = 0$ and $\omega = \pi$, the right-hand side of (2.3) must be modified to $2f(\omega_\ell)E_{X,\ell}$. We ignore these two 'outer Fourier frequencies' and work only with the Fourier frequencies strictly in $(0, \pi)$; since this excludes only R_0, its effect on overall results will be negligible asymptotically.

Taking a logarithm in (2.3) yields a regression equation (called Wahba's formulation)

$$\log(I_X(\omega_\ell)) = \log(f_X(\omega_\ell)) + \log(E_{X,\ell}). \tag{2.4}$$

Subtracting versions of (2.4) for $\{X_t\}$ and $\{Y_t\}$ yields

$$\log(I_X(\omega_\ell)) - \log(I_Y(\omega_\ell)) = \log(f_X(\omega_\ell)) - \log(f_Y(\omega_\ell)) + \log(E_{X,\ell}/E_{Y,\ell}). \tag{2.5}$$

Thus, under a null hypothesis of equal autocovariance functions, $f_X(\omega) = f_Y(\omega)$ except on a subset of $(0, \pi)$ with Lebesgue measure zero (which we tacitly ignore) and $D_\ell := \log(R_\ell)$ has a distribution equivalent to that of $\log(E_1/E_2)$, where E_1 and E_2 are independent standard exponential variates. The next result explicitly identifies this probabilistic structure, perhaps familiar from Whittwer (1984), Lawless (1982), and Johnson *et al.* (1995) amongst others.

Lemma 1. Under the 'null hypothesis' that $\gamma_X(h) = \gamma_Y(h)$ for every integral $h > 0$, $D_\ell = \log(I_X(\omega_\ell)/I_Y(\omega_\ell))$ has the log-logistic probability density, given by

$$f(x) = \frac{e^x}{(1 + e^x)^2}, \qquad x \in R, \tag{2.6}$$

for each $1 \le \ell \le n/2$.

Proof. Let E_1 and E_2 be two independent standard exponential random variables and set $U = E_1/E_2$ and $V = E_2$. The joint distribution of U and V can be verified to have probability density function

$$f_{U,V}(u, v) = v \exp\{-v(1 + u)\}, \qquad u, v \ge 0. \tag{2.7}$$

Integrating out V yields the distribution of U:

$$f_U(u) = \int_0^\infty f_{U,V}(u, v)dv = \frac{1}{(1 + u)^2}, \qquad u \ge 0, \tag{2.8}$$

where the fact that $\Gamma(2) = 1$ has been used to carry out the above integration. It now follows that $D = \log(U)$ has density as claimed in (2.6). ♣

Several methods now suggest themselves for a test of equal autocovariances. An empirical check merely constructs a probability plot of the ranked D_ℓ's against the distribution in Lemma 1.

For more quantifiable inferences, a simple statistic is merely the sample average of absolute deviations, viz.

$$\overline{AD} = \frac{2}{n}\sum_{\ell=1}^{n/2} |D_\ell\,|. \tag{2.9}$$

Large values of $\overline{AD}$ are critical. From Lemma 1, one can verify that $E|D_\ell\,| = \log 4$ and $\mathrm{Var}|D_\ell\,| = \pi^2/3 - (\log 4)^2$. Hence, a central limit theorem based αth level hypothesis test rejects equal autocovariances when $\overline{AD}$ exceeds $E|D_\ell| + z_\alpha\{\mathrm{Var}|D_\ell\,|/(n/2)\}^{1/2}$, which is $\log 4 + z_\alpha\sqrt{\frac{\pi^2/3-(\log 4)^2}{n/2}}$.

It is worth commenting that Kolmogorov-Smirnov type distances involving

$$M = \max_{1\le\ell\le n/2} |\log(I_{X,\ell}) - \log(I_{Y,\ell})|$$

were experimented with, but demonstrated poor power in simulations. In view of the inconsistency of the periodogram noted earlier, this is not surprising. This also supports using a statistic based on all values of the periodogram. Although the hypothesis test above has a rejection region structured on the central limit assumption that D_i and D_j are independent when $i \neq j$, this is not overly crucial. Indeed, this independence holds approximately for each finite sample size n; moreover, we comment that D_i and D_j are indeed exactly independent when $i \neq j$ for all n when $\{X_t\}$ is Gaussian; Davis and Mikosch (1999) provide extensions to the non-Gaussian case.

3 Simulation Performance.

To gain some practical feel for the methods in the last section in a large sample setting, we simulate a series of length $n = 1000$ of two independent first-order autoregressive (AR(1)) series $\{X_t\}$ and $\{Y_t\}$. Both have autoregressive parameter 0.5 and a white noise variance of unity; hence $\{X_t\}$ and $\{Y_t\}$ indeed have equivalent autocovariances. A histogram aggregated from $50,000$ independent draws of $\overline{AD}$ is plotted in the top graphic in Figure 1. At level 5% ($z_{0.95} = 1.645$), 2571 simulations reject equality of autocovariances; hence, the empirical type I error of the test $2571/50000 = 0.0514$ is close to its theoretical value of 0.05. The bars on the x-axis indicate the theoretical mean (dashed) and 95%th quantile (solid).

To gain some feel for power aspects, we simulate a third series $\{Z_t\}$, a sample of length $n = 1000$ from an AR(1) model with autoregressive parameter 0.95 and white noise variance of unity. In this case, $\{X_t\}$ and $\{Z_t\}$ do not have equal autocovariances. Figure 2 reports a histogram of $50,000$ independent draws of $\overline{AD}$, computed from log-periodogram ratios of $\{X_t\}$ and $\{Z_t\}$, and reveals reasonably good separation from the null distribution. At level

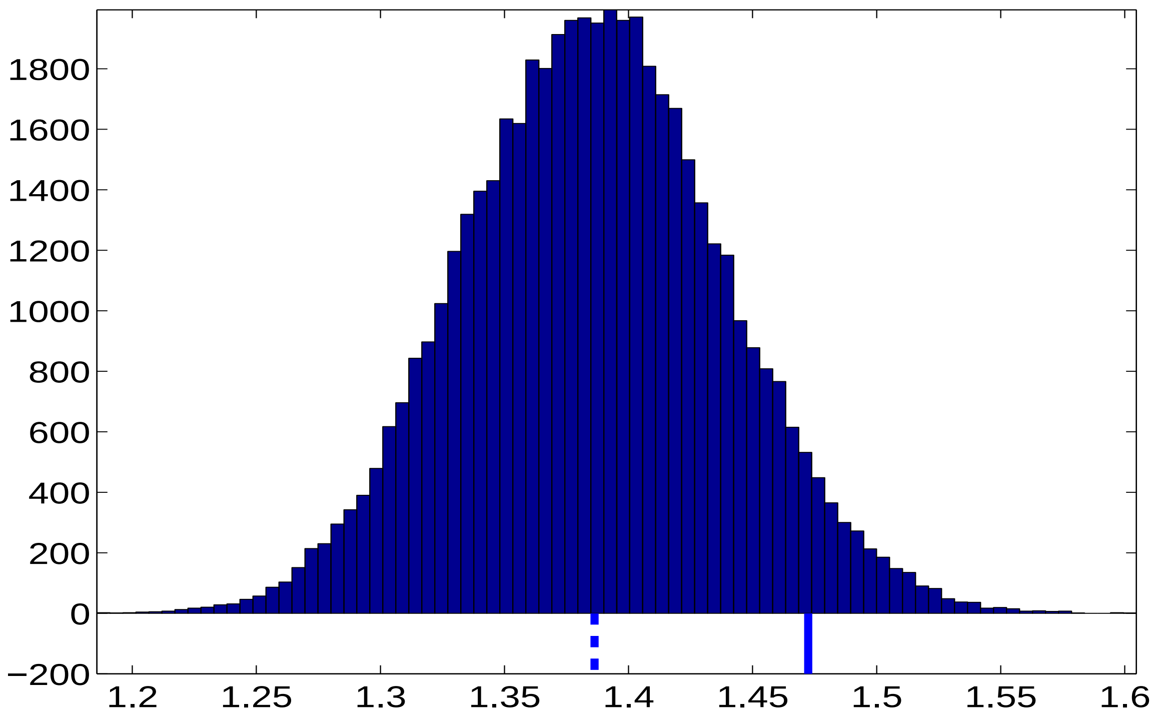

Figure 1: Histogram of 50,000 $\overline{AD}$ draws from AR(1) processes with equal autocovariances. The bars below the x-axis demarcate the theoretical mean (dashed) and 95%th quantile (solid).

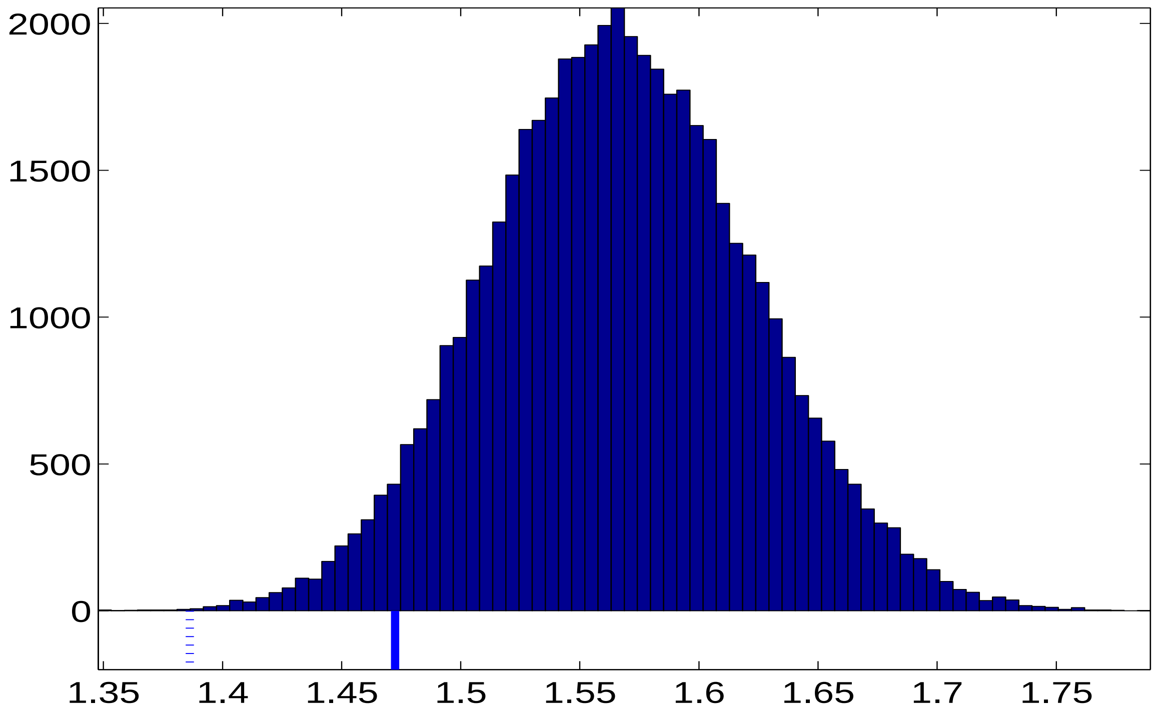

Figure 2: Histogram of 50,000 $\overline{AD}$ draws from two AR(1) processes with non-equal autocovariances. The bars below the x-axis demarcate the theoretical mean and 95%th quantile of the null distribution.

5%, one rejects equal autocovariances in 47877 of the 50000 (95.75%) simulations. This is an excellent empirical power for a test of such nonparametric nature.

Moving to small sample size performance, we consider series lengths of $n = 100$ and compare 50000 draws of $\overline{AD}$ from a Gaussian AR(1) $\{X_t\}$ with autoregressive parameter 0.50 and white noise variance of unity against a first order moving-average (MA(1)) $\{Y_t\}$ with moving-average coefficient of 2.0 and a unit white noise variance. In this case, autocovariances are not equal. At level 5%, the test rejects the null hypothesis of equal autocovariances 78.78% of the time. This is reasonable power for a time series test with such a small sample size.

4 An Application.

We close this work with a study of temperatures from Athens and Atlanta, Georgia, USA. Athens and Atlanta both lie in the Piedmont region of north Georgia and are approximately 75 miles apart. Figure 3 plots monthly averaged temperatures (averaged over all days in month) for these two stations during the period Jan 1950 – Dec 2003. There are 648 observations in each series.

As seasonality arises in temperature series taken from temperate zone latitudes (winter

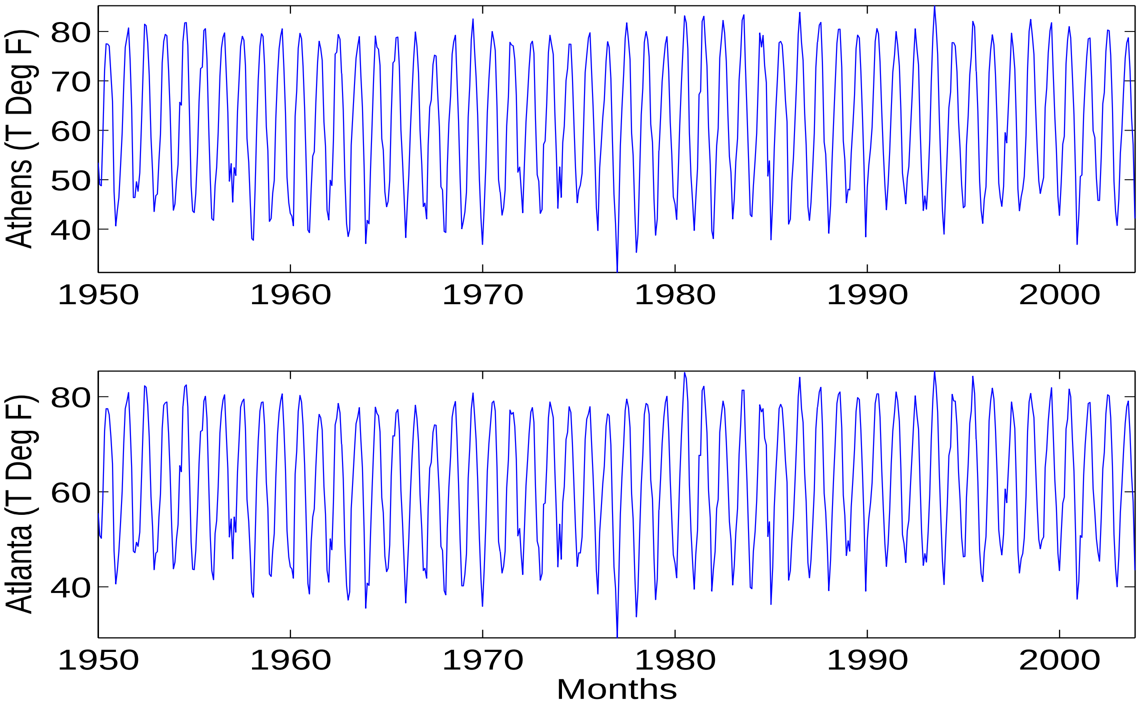

Figure 3: Athens and Atlanta temperatures in the period Jan 1950 – Dec 2003.

temperatures are colder and more variable than summer temperatures), we first standardize each series by month via subtracting a monthly sample mean and then dividing by a monthly sample standard deviation. Lund *et al.* (1995) explains more on the stationarizing effects of seasonal standardizations. The sample autocovariance functions for the Athens and Atlanta seasonally standardized series are displayed in Figure 4. The dashed lines here are 95% confidence bounds (pointwise) for white noise.

Figures 3 and 4 support the local folklore that Athens and Atlanta enjoy similar weather. More formally, the $\overline{AD}$ statistic here is 0.5220, which strongly supports (empirical p-value 0.5164) the hypothesis of equal autocovariances.

Hence, the measurements from Athens and Atlanta appear to have equal autocovariance functions. As the seasonal mean and standard deviations from the two sites are also very similar, the two towns are indeed similar climatologically. Implications of this are that one site could serve as a reference station for the other. This is very useful should a new gauge need to be calibrated, a forecast of future series values need to be made, or the quality/legitimacy of future values at one location be questioned.

We close by remarking that Athens and Atlanta temperatures are indeed correlated. Future work will consider extension of the above methods to cases where $\{X_t\}$ and $\{Y_t\}$ are correlated (not easy), and to settings with three or more series.

Acknowledgements. Support from NSF Grant DMS 030447, NSA Grant E-24-60R, and

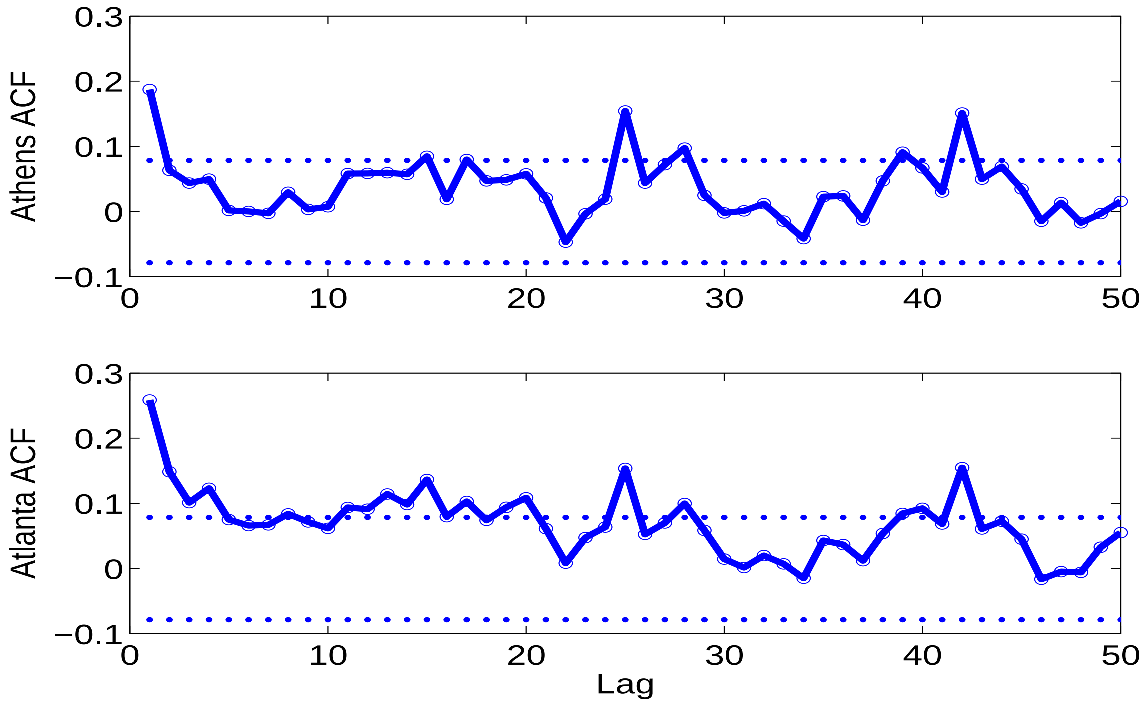

Figure 4: Athens and Atlanta sample autocovariances.

NSF Grant DMS 0307057 are acknowledged. The authors thank Carol Krautmann for securing the temperature data used.

References

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

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

[3] Easterling, D. R., and Peterson, T. C. (1995) A new method for detecting undocumented discontinuities in climatological time series, *International Journal of Climatology*, **15**, 369–377.

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

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

[6] Lee, J., and Lund, R. B. (2004). Revisiting Simple Linear Regression with Autocorrelated Errors, *Biometrika*, **91**, 240-245.

[7] Lund, R. B., Hurd, H., Bloomfield, P., and Smith, R. L. (1995). Climatological Time Series with Periodic Correlation, *Journal of Climate*, **11**, 2787-2809.

[8] Shumway, R. H., and Stoffer, D. L. (2000). *Time Series Analysis and its Applications*, Springer, New York.

[9] Whaba, G. (1980). Automatic smoothing of the log periodogram, *Journal of the American Statistical Association*, **75**, 122-132.

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