TESTING FOR STATIONARITY IN THE FREQUENCY DOMAIN USING A SPHERICITY STATISTIC

Ramon F. Brcich

Signal Processing Group
Institute for Telecommunications
Technische Universität Darmstadt
Merckstrasse 25, Darmstadt 64283, Germany
Email : `r.brcich@ieee.org`

D. Robert Iskander

Centre for Health Research
Queensland University of Technology
Kelvin Grove Q4059, Brisbane, Australia
Email: `d.iskander@qut.edu.au`

ABSTRACT

Non-stationary time series are common in practice and distinguishing them from stationary series has important implications for estimation, identification and forecasting as stationarity models often lead to simpler solutions. We develop a statistical test for stationarity based on the ratio of the arithmetic to geometric means of spectra obtained from non-overlapping segments of the time series. The power of this test is compared to one of the most well-known tests in time series analysis, the KPSS test. Results indicate that the proposed test can replace the KPSS test, particularly for time-varying AR models where it significantly outperforms the current standard.

1. INTRODUCTION

It is known that non-stationary time series are the rule, rather than the exception, in practice [1]. Methods for dealing directly with non-stationary series are not as well developed as for stationary series where theory for estimation, identification and forecasting is well established. Furthermore, most statistical signal processing techniques that are applied to linear and non-linear signal or system analysis require the assumption of stationarity. Since most non-stationary series can be made stationary by suitable transforms such as differencing, it is of value to test whether a series is stationary or non-stationary. Furthermore, long term forecasts based on a stationary model, when in fact a series is non-stationary, can be seriously in error. However, determining whether a signal is stationary is not a straightforward task.

A generic approach to testing for non-stationarity in the frequency domain is to split a series into non-overlapping segments and then to compare the spectra [2, 3, 4, 5].

Time domain approaches generally test using a non-stationary null of a unit root, or random walk, a popular model in econometrics. However, these tests are poorly regarded for their low power and the sense of a non-stationary null [2]. The KPSS test for a stationary null versus a non-stationarity alternative of a unit root overcomes this [6] and is currently one of the most referenced tests. However, in any case, a unit root is a limited form of non-stationarity.

Here, the frequency domain approach is taken to test whether an arbitrary number of spectra are equal at any frequency using a sphericity test. This combines certain aspects of [3, 4, 5].

In order to develop a test for stationarity it is necessary to first define what type of stationarity is being tested for as several exist. Recall that for a time series, X_t, $t \in \mathbb{Z}$, two types of stationarity are generally defined in the literature [7]. The first is simply referred to as stationarity in place of the terms weak, covariance, wide sense or second order stationarity which are in common usage. Stationarity implies the first and second order moments and the autocorrelation function, R_{XX}, exist and are time invariant.

Definition 1 *The time series X_t is stationary if*

$$\mathsf{E}[X_t] = \mu, \qquad \mathsf{E}\left[X_t^2\right] < \infty,$$
$$R_{XX}(\tau) = \mathsf{Cov}[(X_t - \mu)(X_{t+\tau} - \mu)], \qquad \forall\, t, \tau \in \mathbb{Z}.$$

Strict stationarity is the most restrictive definition in that it constrains all statistical properties to be time invariant.

Definition 2 *The time series X_t is strictly stationary if the joint distributions of $(X_{t_1}, \ldots, X_{t_K})^\mathsf{T}$ and $(X_{t_1+\tau}, \ldots, X_{t_K+\tau})^\mathsf{T}$ are equal for all $K \in \mathbb{Z}^+$ and $t_1, \ldots, t_K, \tau \in \mathbb{Z}$.*

In practice, strict stationarity is too difficult to test for and so either Def. (1) or a definition stronger than stationarity but weaker than strict stationarity is used.

Here we concentrate on testing for stationarity with a proviso that the observations have been detrended to remove non-stationarity arising from a slowly changing mean so that $\mathsf{E}[X_t] = 0$. This type of non-stationarity is commonly due to deterministic trends, which it is standard practice to remove by linear least squares or low pass filtering [2, 7]. The hypothesis test is then generally formulated as

$$\begin{aligned} \mathsf{H}: &\quad X_t \text{ is stationary,} \\ \mathsf{K}: &\quad X_t \text{ is non-stationary.} \end{aligned}$$

2. STATIONARITY TESTS

2.1. Time Domain Tests

Most time domain approaches test for a non-stationary null of a unit root against a stationary alternative. Aside from low power, they have a basic conceptual problem since the non-stationary null will be retained unless there is significant evidence otherwise, which is often lacking. In [6] a test for a stationary null versus the non-stationarity alternative of a unit root was presented. Testing for a unit root is important in econometrics where it is a common model, however, it is a limited form of non-stationarity and more general tests are needed.

1-4244-0469-X/06/$20.00 ©2006 IEEE

2.2. Frequency Domain Tests

A general approach to testing for stationarity versus non-stationarity determines whether the spectrum of X_t is changing with time. This is the frequency domain equivalent of testing whether the autocorrelation function is time invariant. If the variance can be assumed finite and the time series has been suitably detrended as mentioned previously, then this test covers the stationarity conditions of Def. (1).

Let X_t, $t = 1, \ldots, N$, be N observations of a suitably detrended time series. Divide X_t into M non-overlapping segments of length T, X_t^m, $t = 1, \ldots, N/M$, $m = 1, \ldots, M$ and denote the spectrum of the mth segment by $C_{XX}^m(\omega)$. A test for the null hypothesis H, that X_t is stationarity, versus the alternative hypothesis K, that X_t is non-stationary, can be constructed as follows

$$\begin{array}{ll} \mathsf{H}: & C_{XX}^1(\omega) = \cdots = C_{XX}^M(\omega)\;\forall\,\omega \\ \mathsf{K}: & \text{not } H. \end{array} \tag{1}$$

This concept has been previously implemented in several ways:

1. Test directly if two spectra are proportional by testing whether the spectral values are proportional at chosen frequencies using a sphericity statistic [5]. This test is only valid for $M = 2$.
2. Test if the evolutionary spectrum is constant with respect to time by performing an ANOVA (analysis of variance) on a two dimensional least squares regression fit to a grid of values in the time-frequency plane [3].
3. The evolutionary spectrum based test above was simplified by taking the spectra of non-overlapping segments of the observations and then utilising ANOVA to test whether the spectrum is time invariant [4]. This test is valid for arbitrary M. Conceptually, this is equivalent to replacing the evolutionary spectrum in the previous test with a spectrogram using non-overlapping segments.

2.2.1. *Sphericity Based Test*

The frequency domain sphericity based test proposed here uses properties of the sphericity statistic in a similar way to the above test for proportionality of two spectra, we now consider this test in more detail.

In [8] a measure of distance between two spectra $C_{XX}(\omega)$ and $C_{YY}(\omega)$ was proposed as

$$D(X,Y) = \log \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{C_{XX}(\omega)}{C_{YY}(\omega)}\,d\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\frac{C_{XX}(\omega)}{C_{YY}(\omega)}\,d\omega$$

which is akin to an entropy ratio. This measure can, for instance, be used to test whether two spectra are proportional [9, 10, 11], or, whether $C_{XX}(\omega)$ is a white process by setting $C_{YY}(\omega) = 1$ [12]. It was shown that this is equivalent to the time domain sphericity statistic

$$S(X,Y) = \log \frac{\mathrm{tr}\left(\Gamma_X\Gamma_Y^{-1}\right)/T}{\left(\det\left(\Gamma_X\Gamma_Y^{-1}\right)\right)^{1/T}}$$

where Γ_X and Γ_Y are the covariance matrices of X and Y respectively in the sense that $\lim_{T\to\infty}|\hat{S} - \hat{D}| = 0$ with probability one. In practice, the biased sample autocorrelation may be used to estimate S, while a smoothed periodogram is used to estimate D.

We propose another interpretation of this statistic which is obtained by constructing a sphericity statistic, i.e. the ratio of the arithmetic to geometric means, from values of the spectrum $C_{XX}(\omega_k)$ at discrete frequencies ω_k, $k = 1, \ldots, T$,

$$S(X) = \frac{\frac{1}{T}\sum_{k=1}^{T} C_{XX}(\omega_k)}{\left(\prod_{k=1}^{T} C_{XX}(\omega_k)\right)^{\frac{1}{T}}}.$$

Taking the logarithm of $S(X)$, $S_L(X)$, gives

$$S_L(X) = \log\frac{1}{2\pi}\frac{2\pi}{T}\sum_{k=1}^{T} C_{XX}(\omega_k) - \frac{1}{2\pi}\frac{2\pi}{T}\sum_{k=1}^{T}\log C_{XX}(\omega_k).$$

So that if

$$\lim_{T\to\infty}\log\frac{2\pi}{T}\sum_{k=1}^{T} C_{XX}(\omega_k) = \log\int_{-\pi}^{\pi} C_{XX}(\omega)\,d\omega$$
$$\lim_{T\to\infty}\frac{2\pi}{T}\sum_{k=1}^{T}\log C_{XX}(\omega_k) = \int_{-\pi}^{\pi}\log C_{XX}(\omega)\,d\omega$$

holds then $D(X,1)$ can be interpreted as the limit, $T\to\infty$, of a sphericity statistic taken at a set of frequencies ω_k in that it tests whether the spectral values $C_{XX}(\omega_k)$, $k = 1, \ldots, T$, are equal. Note that $S_L(X) = 0$ iff the spectral values are equal and $S_L(X) > 0$ if not. Replacing $C_{XX}(\omega_k)$ with a ratio of two spectra, it is straightforward to see how $S_L(X)$ is a test statistic for the hypothesis that two spectra are equally proportional at all frequencies.

2.2.2. *Proposed Sphericity Based Test*

The sphericity statistic described above can only be used to test whether $M = 2$ spectra are proportional. We wish to generalise this test to arbitrary M for the following reasons:

1. It enables us to have a variable resolution in time. This is useful for determining which segments of the observations are stationarity or non-stationarity and where the transition regions lie. Furthermore, since non-stationarity may be apparent only at certain time scales, it is important that it is possible to test for non-stationarity over various intervals of time.
2. The proposed test enables us to determine whether non-stationarity exists only for some frequencies. This may occur, for example, due to the presence of a sinusoidal signal whose frequency changes over time.

To develop the test, we make use of the following well known relationship between the geometric and arithmetic means [13].

Property 1 *For positive real numbers a_m, $m = 1, \ldots, M$, the arithmetic mean, $A(a_1, \ldots, a_M) = 1/M\sum_{m=1}^{M} a_m$, and the geometric mean, $G(a_1, \ldots, a_M) = \left(\prod_{m=1}^{M} a_m\right)^{1/M}$, satisfy*

$$\log\frac{A(a_1, \ldots, a_M)}{G(a_1, \ldots, a_M)} \geq 0$$

with equality iff $a_1 = \cdots = a_M$.

This motivates use of the statistic

$$\begin{aligned} S_M(\omega_k) &= \log\frac{A(C_{XX}^1(\omega_k), \ldots, C_{XX}^M(\omega_k))}{G(C_{XX}^1(\omega_k), \ldots, C_{XX}^M(\omega_k))} \\ &= \log\frac{1}{M}\sum_{m=1}^{M} C_{XX}^m(\omega_k) - \frac{1}{M}\sum_{m=1}^{M}\log C_{XX}^m(\omega_k) \end{aligned}$$

to be used in a test for (1), but only at frequency ω_k. Note that here we use the sphericity statistic to test for equality of M spectra at a specific frequency in comparison to its use in Section 2.2.1 where it was used to test whether two spectra are equally proportional across all frequencies. In comparison to the two frequency domain stationarity tests in Section 2.2, we are testing for time invariance of the spectra of non-overlapping segments of the observations at a specific frequency using a sphericity statistic instead of across all frequencies simultaneously using ANOVA. The proposed test is then a hybrid between the approaches in Section 2.2.1.

To test (1) at all frequencies we may perform either a multiple hypothesis test or find a suitable projection of the statistics $S_M(\omega_k)\ \forall\omega_k$ onto one dimension. Here we do this by taking the mean over all frequencies to obtain $\overline{S}_M$. Estimates of the spectrum are obtained by kernel smoothing of the periodogram using a rectangular window.

3. SIMULATIONS AND DISCUSSION

All simulations were run over 1000 Monte Carlo realisations. The number of observations was $N = 1024$ while the number of non-overlapping segments was varied over $M = \{2, 4, 8, 16\}$ so that the corresponding segment lengths were $T = \{512, 256, 128, 64\}$. A rectangular window was used for smoothing the periodogram to estimate the spectrum of the mth segment, $C^m_{XX}(\omega)$. The kernel bandwidth was varied so that K evenly spaced frequencies between 0 and π were evaluated for $K = \{4, 6, 8, 10\}$ with $\omega_1 = 0$ and $\omega_K = \pi$.

Rejection rates were evaluated under the null hypothesis for six stationary models, all ARMA(p, q) (Autoregressive Moving Average) processes:

1. iid Gaussian: $X_t = \varepsilon_t$.
2. AR(1), $\alpha_1 = 0.5$: $X_t = 0.5X_{t-1} + \varepsilon_t$.
3. AR(1), $\alpha_1 = -0.5$.
4. AR(1), $\alpha_1 = 0.95$.
5. MA(1) with $\beta_1 = 1$: $X_t = \varepsilon_t + \varepsilon_{t-1}$.
6. AR(5): $X_t = 0.5X_{t-1} - 0.6X_{t-2} + 0.3X_{t-3} - 0.4X_{t-4} + 0.2X_{t-5} + \varepsilon_t$.

For the alternative hypothesis, nine non-stationary models were used including ARIMA(p, d, q) (AR Integrated MA), TVAR(p) (time varying AR(p)) and GARCH (Generalised AR Conditionally Heteroscedastic) processes. Briefly, ARIMA processes generalise ARMA processes by allowing for unit roots of multiplicity d. TVAR processes allow for time varying AR parameters. GARCH processes allow the variance of the innovation sequence in an AR process to vary. These are the most commonly used non-stationary models in time series analysis [2, 1, 7]. The specific cases were:

1. Random walk: ARIMA$(0, 1, 0)$, $X_t = X_{t-1} + \varepsilon_t$.
2. Integrated random walk: ARIMA$(0, 2, 0)$.
3. ARI: ARIMA$(1, 1, 0)$ with $\alpha_1 = 0.5$.
4. IMA: ARIMA$(0, 1, 1)$ with $\beta_1 = 1$.
5. ARIMA: ARIMA$(1, 1, 1)$ with $\alpha_1 = 0.5, \beta_1 = 1$.
6. TVAR A: TVAR(1) with α_1 varying linearly between 0.2 and 0.8.
7. TVAR B: TVAR(1) with α_1 varying linearly between -0.5 and 0.5.
8. GARCH A: An AR(1) model with $\alpha_1 = 0.5$ where the variance of the innovations varies linearly from 0.5 to 2.
9. GARCH B: An AR(1) model with $\alpha_1 = 0.5$ where the variance of the innovations varies linearly from 0.1 to 1.

In each case the innovations, ε_t, were iid $\mathcal{N}(0, 1)$.

Thresholds for the test were based on a 5% level of significance estimated empirically under a null distribution of iid Gaussian observations.

The frequencies $\omega = 0, \pi$ were excluded from $\overline{S}_M$. This was motivated by the observation that under the null, the distribution of $S_M(\omega_k)$ appears constant over frequency, excluding the cases $\omega = 0, \pi$. The apparent similarity of the distributions over frequency can be explained by the fact that the smoothed periodogram is asymptotically χ^2 with the same degrees of freedom for $\omega \neq 0, \pi$, while the degrees of freedom are reduced for $\omega = 0, \pi$.

Some representative results showing rejection rates for the proposed test over all frequency bins and number of segments are shown in Table 1. It is difficult to give optimal values for K and M, although to trade-off maintenance of the set level under the null and maximising the power, $K = 10$ frequency bins and $M = 4$ segments seems reasonable. Note that the AR(1) model with $\alpha_1 = 0.95$ is near to a non-stationary random walk model and so is often rejected, especially over shorter time intervals with smaller M.

The results of the proposed test were compared to one of the most well-known, the KPSS test [6]. In Table 2 the KPSS test and $\overline{S}_M$ are compared for $K = 10$ and $M = 4$. Both have similar performance under the null although an exception occurs for the near to non-stationary AR(1) model with $\alpha_1 = 0.95$ where the rejection rate is 37.4% for the KPSS test compared to 11.8% for $\overline{S}_M$. Under the alternative $\overline{S}_M$ has a much greater power for the TVAR and GARCH models. The only model where the KPSS test has a non-negligible higher power is for the random walk which it is specifically designed to detect.

Further simulations summarised in Table 3 for an AR(1) process, $\alpha \in [-1, 1]$, show that the behaviour of $\overline{S}_M$ depends only on $|\alpha_1|$. For the KPSS test, the rejection rate constantly decreases with α_1 to 0% for the non-stationary case of $\alpha_1 = -1$. This appears to be a consequence of the KPSS test implicitly testing for a random walk, i.e. an AR(1) process with $\alpha_1 = 1$.

4. CONCLUSION

A test for stationarity was proposed which exploited the properties of the sphericity statistic. First, it was shown that the sphericity statistic in the frequency domain can be interpreted as a ratio of the arithmetic to geometric means of spectra. Further, this test was generalised to an arbitrary number of spectra. Extensive simulations were performed to evaluate the power of the proposed test for different numbers of segments and frequency bins. Several stationary processes as well as non-stationary processes were used. The proposed test was compared to the well-known KPSS test as the current standard in stationarity testing. Results indicate that the proposed test can replace the KPSS test and that it shows significant improvements in performance in the case of alternatives modelled by a time-varying AR process.

Table 1. Rejection rates (%) for some of the models using $\overline{S}_M$.

Model			No. frequency bins K			
	M	T	4	6	8	10
AR(1),	16	64	7.2	5.2	3.9	5.5
$\alpha_1 = 0.5$	8	128	7.0	8.9	5.7	5.9
	4	256	6.7	5.8	6.0	6.5
	2	512	5.6	5.9	5.8	4.9
AR(1),	16	64	59.3	46.0	31.3	23.8
$\alpha_1 = 0.95$	8	128	34.8	31.6	26.0	21.6
	4	256	18.8	17.1	11.5	11.8
	2	512	9.3	9.7	7.0	6.7
MA(1),	16	64	9.3	8.3	7.1	6.7
$\beta_1 = 1$	8	128	8.7	10.6	8.5	8.8
	4	256	6.9	8.0	7.3	6.0
	2	512	5.7	7.7	5.6	6.2
AR(5)	16	64	11.7	19.4	9.7	6.2
	8	128	12.4	20.8	12.4	7.8
	4	256	9.6	17.4	8.2	7.0
	2	512	7.5	9.9	7.0	7.0
Random	16	64	98.3	96.9	91.3	80.7
walk	8	128	95.8	95.5	94.3	93.0
	4	256	91.4	88.7	86.9	85.4
	2	512	67.6	69.8	67.2	64.8
ARIMA	16	64	100.0	100.0	99.9	100.0
	8	128	100.0	100.0	100.0	99.8
	4	256	99.0	99.5	99.1	98.9
	2	512	88.3	88.4	86.4	86.3
TVAR B	16	64	69.8	85.3	77.3	76.3
	8	128	90.0	98.1	99.1	98.2
	4	256	95.2	99.8	99.6	99.9
	2	512	97.2	100.0	100.0	99.9
GARCH A	16	64	98.6	94.9	81.2	74.0
	8	128	99.7	99.6	98.9	98.4
	4	256	99.9	100.0	100.0	99.9
	2	512	100.0	99.9	100.0	99.8

Table 2. Rejection rates (%) for the KPSS test and $\overline{S}_M$ for $K = 10$ and $M = 4$.

Model	KPSS	$\overline{S}_M$
iid Gaussian	4.6	5.0
AR(1), $\alpha_1 = 0.5$	5.1	6.5
AR(1), $\alpha_1 = -0.5$	4.1	4.9
AR(1), $\alpha_1 = 0.95$	37.4	11.8
MA(1), $\beta_1 = 1$	5.9	6.0
AR(5)	3.0	7.0
Random walk	98.1	85.4
Integrated random walk	99.8	99.5
ARI	92.4	97.7
IMA	93.6	98.0
ARIMA	92.2	98.9
TVAR A	12.5	77.5
TVAR B	6.6	99.9
GARCH A	6.2	99.9
GARCH B	4.1	100.0

Table 3. Rejection rates (%) for an AR(1) process with $\alpha_1 \in [-1, 1]$ for the KPSS test and $\overline{S}_M$ statistic with $K = 10$, $M = 4$.

α_1	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
$\overline{S}_M$	84.4	6.0	5.4	4.4	4.8	4.7	5.9	5.3	5.7	8.1	85.9
KPSS	0.0	1.8	3.6	4.7	4.4	4.6	5.8	4.4	7.0	9.1	98.9

5. REFERENCES

[1] R. H. Schumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, Springer-Verlag, 2000.

[2] C. Chatfield, *The Analysis of Time Series: An Introduction*, Chapman and Hall, 2004.

[3] M. B. Priestley and T. Subba Rao, "A test for non-stationarity of time-series," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 31, no. 1, pp. 140–149, 1969.

[4] M. Grasse, M. R. Frater, and J. F. Arnold, "Testing VBR video traffic for stationarity," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 3, pp. 448–457, April 2000.

[5] S. Vaton, "A new test of stationarity and its application to teletraffic data," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1998, vol. 6, pp. 3449–3452.

[6] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics*, vol. 54, no. 1, pp. 159–178, 1992.

[7] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, 1987.

[8] A. Mokkadem, "Sur quelques nouveaux test pour les séries chronologiques," *Comptes Rendus de l'Academie des Sciences Serie I - Mathematique*, vol. 318, no. 8, pp. 755–758, April 1994.

[9] K. Drouiche, "Testing proportionality for autoregressive processes," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 672–681, March 2003.

[10] R. Martinez, P. Gomez, and K. Drouiche, "A test of homogeneity for autoregressive processes," *International Journal of Adaptive Control and Signal Processing*, vol. 16, pp. 231–242, 2002.

[11] A. Mokkadem, "A measure of information and its applications to test for randomness against arma alternatives and to goodness-of-fit test," *Stochastic Processes and their Applications*, vol. 72, no. 2, pp. 145–159, 1997.

[12] K. Drouiche, "A spectral density test for whiteness," *International Journal of Adaptive Control and Signal Processing*, vol. 13, pp. 633–650, 1999.

[13] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, Dover, 1965.