Comparison of time series with unequal length

Caiado, Jorge and Crato, Nuno and Peña, Daniel

December 2007

Comparison of time series with unequal length

Jorge Caiado[a] Nuno Crato[b] and Daniel Peña[c]

[a]Department of Economics and Management, School of Business Administration, Polytechnic Institute of Setubal, and CEMAPRE, Campus do IPS, Estefanilha, 2914-503 Setúbal, Portugal. Tel.: +351 265 709 438. Fax: +351 265 709 301. E-mail: jcaiado@esce.ips.pt

[b]Department of Mathematics, School of Economics and Business, Technical University of Lisbon, and CEMAPRE, Rua do Quelhas 6, 1200-781 Lisboa, Portugal. Tel.: +351 213 925 846. E-mail: ncrato@iseg.utl.pt

[c]Department of Statistics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Spain. Tel.: +34 916 249 806. Fax.: +34 916 249 849. E-mail: dpena@est-econ.uc3m.es

Abstract: The comparison and classification of time series is an important issue in practical time series analysis. For these purposes, various methods have been proposed in the literature, but all have shortcomings, especially when the observed time series have different sample sizes. In this paper, we propose spectral domain methods for handling time series of unequal length. The methods make the spectral estimates comparable, by producing statistics at the same frequency. A first sensible approach may consist on zero-padding the shorter time series in order to increase the corresponding number of periodogram ordinates. We show that this works well provided the sample sizes are not very different, but does not give good results in case the time series lengths are very unbalanced. For this latter case, we study some periodogram-based comparison methods and construct a test. Both the methods and the test display reasonable properties for series of any lengths. Additionally and for reference, we develop a parametric comparison method. The procedures are assessed by a Monte Carlo simulation study. As an illustrative example, a periodogram method is used to compare and cluster industrial production series of some developed countries.

Keywords: Cluster analysis; Interpolated periodogram; Reduced periodogram; Spectral analysis; Time series; Zero-padding.

1. Introduction

The classification analysis of time series has useful applications in several fields. In population studies, for instance, one may be interested in identifying similarities in the series of birth and death rates for different populations. In finance, one may interested in identifying dependences in financial market returns for classifying and grouping stocks. In economics, one may be interested in a cluster analysis of some countries by looking at the main macroeconomic time series indicators.

Methods for comparing time series have been studied for some time now, mainly by using autocorrelation analysis and by model fitting methods. More recently, building upon early work of Coates and Diggle (1986), Diggle and Fisher (1991), and Dargahi-Noubary (1992), spectral analysis methods have been considered. Specific frequency-domain methods for discrimination and clustering analysis of time series were proposed by Maharaj (2002), by Quinn (2006), and by Caiado, Crato and Peña (2006). As this last paper shows, spectral methods can work very well for comparing time series.

A problem that often arises in real applications is the difficulty of finding time series of equal length. For instance, in the business cycle study of some industrialized countries, Camacho, Pérez-Quiróz and Saiz (2004) found time series of unequal length and had to truncate data in order to compare the series. Only after the truncation, they were able to use spectral estimates to compute distances across countries.

In this paper, we build upon the work presented in Caiado, Crato and Peña (2006) and use periodogram based metrics for comparison of time series. We extend the method, proposing a process of adjusting the number of used periodogram ordinates of the studied series. The easiest way of accomplishing this is to zero-pad the shorter series by making it as long as the larger series. This way, both periodograms will have the same number of ordinates and we will be able to compute a derivative of the Euclidean distance between the periodogram ordinates of both series. This approach, however, proves to introduce distortions in the estimated periodogram ordinates and these distortions are serious when the zero-padded series is significantly shorter than the larger series. We then study other approaches. Firstly, we construct a reduced frequency periodogram for the larger series, by computing the periodogram ordinates only at the smaller series corresponding frequencies. This method has the further advantage of allowing the construction of an appropriate test. Secondly, we construct an interpolated periodogram for the longer series, obtaining for this larger series a shorter number of periodogram ordinates. This method seems to work particularly well for comparison purposes.

The remainder of the paper is organized as follows. In Section 2, we introduce the periodogram based metrics for handling series of unequal length. In Section 3, we present nonparametric and parametric tests of hypothesis to determine whether two series have been generated by stochastic processes with similar properties. In Section 4, we present the results of a Monte Carlo simulation study on the performance of the various metrics and tests. In Section 5, we apply a periodogram metric to analyze industrial time series of developed economies. In Section 6, we summarize the paper.

2. Periodogram-based metrics

Let $\{x_t, t = 1, ..., n_x\}$ and $\{y_t, t = 1, ..., n_y\}$ be two stationary processes with different sample sizes $n_x \neq n_y$. The periodogram of x_t is given by

$$P_x(\omega_j) = \frac{1}{n_x} \left| \sum_{t=1}^{n_x} x_t e^{-it\omega_j} \right|^2, \tag{1}$$

where $\omega_j = 2\pi j/n_x$, for $j = 1, ..., m_x$, with $m_x = [n_x/2]$, the largest integer less or equal to $n_x/2$. Similar expression is defined for $P_y(\omega_p)$, with $\omega_p = 2\pi p/n_y$, for $p = 1, ..., m_y$, with $m_y = [n_y/2]$. The Euclidean distance between the periodograms ordinates $P_x(\omega_j)$ and $P_y(\omega_p)$ is not adequate for comparison of series x_t and y_t, since $m_x \neq m_y$ and ω_j and ω_p do not coincide.

In the signal processing literature (Wang and Blostein, 2004), a common solution to the problem of unequal length consists of extending the shorter series y_t, by adding zeros and getting y_t',

$$y_t' = \begin{cases} y_t, & t = 1, ..., n_y \\ 0, & t = n_y + 1, ..., n_x, \end{cases}$$

and then computing the periodogram of y_t'

$$P_{y'}(\omega_j) = \frac{1}{n_x} \left| \sum_{t=1}^{n_x} y_t' e^{-it\omega_j} \right|^2 . \tag{2}$$

This approach, called "zero-padding", matches the frequencies of the series and produces a nicely smoothed periodogram. A *zero-padding* based metric for handling series of unequal length using the periodogram ordinates is then defined by

$$d_{ZP}(x, y) = \sqrt{\frac{1}{m_x} \sum_{j=1}^{m_x} [P_x(\omega_j) - P_{y'}(\omega_j)]^2}. \tag{3}$$

Alternatively, since periodograms can be calculated at any frequency, we may simply compute the periodogram of the longer series x_t at the frequencies of the shorter series y_t, that is

$$P_x^{RP}(\omega_p) = \frac{1}{n_x} \left| \sum_{t=1}^{n_x} x_t e^{-it\omega_p} \right|^2 , \tag{4}$$

where $\omega_p = 2\pi p/n_y$, for $p = 1, ..., m_y < m_x$, which we will call the *reduced periodogram.* The reduced periodogram based metric is then defined by

$$d_{RP}(x, y) = \sqrt{\frac{1}{m_y} \sum_{p=1}^{m_y} [P_x^{RP}(\omega_p) - P_y(\omega_p)]^2}. \tag{5}$$

Another way to solve the problem is to interpolate the periodogram ordinates of the series with longer (shorter) length from the series with the shorter (longer) length. Without loss of generality, let $r = [p\frac{m_x}{m_y}]$ be the largest integer less or equal to $p\frac{m_x}{m_y}$ for $p = 1, ..., m_y$, and $m_y < m_x$. The periodogram ordinates of x_t can be estimated as

$$\begin{aligned} P_x^{IP}(\omega_p) &= P_x(\omega_r) + (P_x(\omega_{r+1}) - P_x(\omega_r)) \times \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}} \\ &= P_x(\omega_r) \left(1 - \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}} \right) + P_x(\omega_{r+1}) \left(\frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}} \right) . \end{aligned} \tag{6}$$

This procedure will yield an interpolated periodogram with the same Fourier frequencies of the shorter periodogram $P_y(\omega_p)$. We then use the following distance measure based on the *interpolated periodogram* ordinates,

$$d_{IP}(x,y) = \sqrt{\frac{1}{m_y}\sum_{p=1}^{m_y}\left[P_x^{IP}(\omega_p) - P_y(\omega_p)\right]^2}. \tag{7}$$

If we are only interested in the dependence structure and not in the process scale, then we can standardize the periodograms, dividing them by the sample variances: $NP_x^{IP}(\omega_p) = P_x^{IP}(\omega_p)/\widehat{\sigma}_x$ and $NP_y(\omega_p) = P_y(\omega_p)/\widehat{\sigma}_y$. Since the variance of the periodogram is proportional to the spectrum at the same Fourier frequencies, we may use a distance measure between the logarithms of the normalized periodograms

$$d_{ILNP}(x,y) = \sqrt{\frac{1}{m_y}\sum_{p=1}^{m_y}\left[\log NP^{IP}(\omega_p) - \log NP_y(\omega_p)\right]^2}. \tag{8}$$

It is straightforward to show that the distance measures (7) and (8) fulfil the usual properties of a metric except the triangle inequality: (i) $d(x,y) = 0$ if $P_x^{IP}(\omega_p) = P_y(\omega_p)$ or $NP^{IP}(\omega_p) = NP_y(\omega_p)$; (ii) $d(x,y) \geq 0$; and (iii) $d(x,y) = d(y,x)$.

Another useful measure of distance discussed, for instance, in Caiado, Crato and Peña (2006) is the Kullback-Leibler (KL) form

$$d_{KL}(x,y) = \sum_{p=1}^{m_y}\left[\frac{NP_x^{IP}(\omega_p)}{NP_y(\omega_p)} - \log\frac{NP_x^{IP}(\omega_p)}{NP_y(\omega_p)} - 1\right]. \tag{9}$$

This measure is greater than or equal to zero, with equality if $NP_x^{IP}(\omega_p) = NP_y(\omega_p)$ for all p. The potential success of measure (9) should be related with measure (8), and it can also be applied to zero padding and reduced periodogram approaches.

The performance of the three periodogram-based metrics (padding, reduced and interpolated) for both mean Euclidean and Kullback-Leibler forms will be checked by Monte Carlo simulation.

3. Hypotheses testing procedures

3.1. Nonparametric approaches

Based on the distance measures described above, we now suggest a test of hypotheses to determine whether two independent time series are realizations of stochastic processes with identical second-order properties. Given two independent stationary series x_t and y_t, with $n_x = n_y$, the null hypothesis to be tested is $H_0 : f_x(\omega_j) = f_y(\omega_j)$, that is, there is no difference between the underlying spectra of the series $\{x_t\}$ and $\{y_t\}$ at all Fourier frequencies ω_j.

Since, asymptotically, $P_x(\omega_j) \sim f_x(\omega_j)\chi^2_{(2)}/2$, where $f_x(\omega)$ is the spectral density function, it follows that $E\left[P_x(\omega_j)/\widehat{\gamma}_{0,x}\right] = f_x(\omega_j)/\sigma_x^2$ and $Var\left[P_x(\omega_j)/\widehat{\gamma}_{0,x}\right] = f_x^2(\omega_j)/\sigma_x^4$.

Similar expressions apply to the periodogram of y_t. As the two series x_t and y_t are independent, $P_x(\omega_j)$ and $P_y(\omega_j)$ must be independently distributed as well.

Under some suitable conditions, the logarithmic transformation of the sample spectrum is closer to the normal distribution than to a chi-square distribution (Jenkins and Priestley, 1957). The following statistic provides a test of significance for comparing the log normalized periodograms of the two series,

$$D_{NP} = \frac{\frac{1}{m}\sum_{p=1}^{m}\left[\log NP_x(\omega_p) - \log NP_y(\omega_p)\right]}{\sqrt{\left(s^2_{LNP,x} + s^2_{LNP,y}\right)/m}}, \tag{10}$$

where $s^2_{LNP,x} = \frac{1}{m}\sum_{j=1}^{m}\left[\log NP_x(\omega_j) - \overline{x}_{LNP}\right]^2$ and $\overline{x}_{LNP} = \frac{1}{m}\sum_{j=1}^{m}\log NP_x(\omega_j)$ ($s^2_{LNP,y}$ and $\overline{y}_{LNP}$ are similarly defined). This statistic is approximately normally distributed with zero mean and unit variance.

For different lengths, $m_x \neq m_y$, regular periodograms cannot be used as the Fourier frequencies are different. One approach could be the use of the interpolated periodogram for the longer series. However, interpolation destroys the independence of periodogram values across frequencies. This does not happen with the reduced periodogram, which can be used in equation (10). In this case, we simply compute the reduced periodogram of the longer series x_t at the frequencies $\omega_p = 2\pi p/n_y$, for $p = 1, ..., m_y$ as defined in (4).

Another useful test is based on the likelihood ratio approach. For instance, we may compute the pooled spectra $P(\omega_j) = \frac{1}{2}\left[P_x(\omega_j) + P_y(\omega_j)\right]$ (or $NP(\omega_j) = \frac{1}{2}\left[NP_x(\omega_j) + NP_y(\omega_j)\right]$) under the null hypothesis $H_0 : f_x(\omega_j) = f_y(\omega_j)$ and use the likelihood ratio test

$$-2\log\lambda \approx -\sum_{p=1}^{m}\log NP_x(\omega_j) - \sum_{p=1}^{m}\log NP_y(\omega_j) + 2\sum_{p=1}^{m}\log NP(\omega_j), \tag{11}$$

which is distributed proportionally to a chi-square random variable. We fix the size of the Type I error equal to the significance level α. Thus the test rejects H_0 if -2log $\lambda < k$, where k is a nonnegative constant. This test can be easily extended to time series of unequal lengths using the reduced periodogram approach.

3.2. Parametric approach

The problem of comparison of series of unequal length can also be analyzed by a parametric approach. Suppose we have two independent time series x_t and y_t generated by the same ARMA(p,q) process, but with different parameter values. Let the $k = p + q$ estimated parameters be grouped in the vectors $\widehat{\beta}_x$ and $\widehat{\beta}_y$ with estimated covariance matrices V_x and V_y, respectively. We want to check whether they are different realizations of the same stochastic process, so that $E[\widehat{\beta}_x] = E[\widehat{\beta}_y] = \beta$. Then $\delta = \widehat{\beta}_x - \widehat{\beta}_y$ for large samples will be an approximately normally distributed vector with zero mean and covariance matrix

$$V_\delta = V_x + V_y, \tag{12}$$

and therefore, we can use the statistic

$$D_P = \delta' V_\delta^{-1} \delta, \tag{13}$$

which is asymptotically a chi-square distribution with k degrees of freedom under the null $\beta_x = \beta_y$. Hamilton (1994, Section 14.3) suggested a similar statistic to test for structural stability of autoregressive and moving average (ARMA) models over different subperiods.

In order to test if two generating ARMA processes are equal, the model for each time series can be selected by Akaike's Information Criterion (AIC) or Bayesian's Information Criterion (BIC) selection criterions. If the obtained model is the same for the two time series, then the statistic D_P is computed by using the estimated parameters in each time series. If the selected models selected are different, the problem is more involved. Two methods have been proposed:

(i) We can start fitting a large ARMA model to both processes which encompass the two models to be compared, for instance the larger of two AIC or BIC selected models. This method has two main problems: (1) the estimated parameters will be highly correlated for the overparametrized estimated model (or models) and the corresponding covariance matrix (or matrices) may be close to singular; (2) we have to be very careful to avoid possible near cancellation of the AR and MA roots on both sides.

(ii) Alternatively, in order to avoid the serious problem of near cancellation of roots, we can use AR approximations and thus fit to both processes the larger selected AR model and then compare the estimated parameters (see Maharaj, 1996). This method also has the problem that (1) we may get very correlated estimated parameters, specially when we have MA generating processes; and (2) we may need very large AR models.

Given these problems, we propose an alternative approach to apply the parametric test when the selected models are different.

(iii) We fit both selected models, say $M1$ and $M2$, to both time series and compute the statistic D_P in these two situations, i.e., $D_P(M1)$ and $D_P(M2)$. If $D_P(M1) \leq \chi^2_{(k_1)}$ and $D_P(M2) \leq \chi^2_{(k_2)}$, where k_1 and k_2 are the degrees of freedom associated with the models $M1$ and $M2$, the null hypothesis is not rejected and we conclude that the processes are generated by the same model. If the null hypothesis is rejected in one of the models, or in both, then we conclude that the generating processes are different. Since we have two comparison statements to be made, the Bonferroni inequality suggests each test with a significance level $\alpha/2$ to ensure that the overall significance level is at most α. In our simulation study, we will use this alternative approach (iii).

4. Monte Carlo simulations

4.1. Performance of the periodogram-based metrics

To illustrate the performance of the periodogram-based metrics (zero-padding, reduced and interpolated), we performed a set of simulations. For each of the considered processes, we simulated pairs of series of different sample sizes, $(n_1, n_2) = \{(50, 100),\ (200, 100),\ (500, 250),\ (1000, 500)\}$. So four different series were simulated for each replication. For each case, we performed 1000 replications. We performed the following comparisons:

(a) AR(1), $\phi = 0.9$ versus AR(1), $\phi = 0.5$;
(b) AR(1), $\phi = 0.9$ versus ARIMA(0,1,0);
(c) AR(2), $\phi_1 = 0.6$, $\phi_2 = -0.3$ versus MA(2), $\theta_1 = -0.6$, $\theta_2 = 0.3$;
(d) ARFIMA(0,0.45,0) versus white noise;

(e) ARFIMA(0,0.45,0) versus AR(1), $\phi = 0.95$;
(f) ARFIMA(0,0.45,0) versus IMA(1,1), $\theta = 0.4$.

In case (a), we compare low-order models of similar type and similar autocorrelation functions. In case (b), we compare a nonstationary process and a near nonstationary AR process. In case (c), we compare selected second-order ARMA processes in order to deal with peak spectra. In case (d), we compare stationary processes with very different characteristics of persistence. In case (e), we compare near-nonstationary long memory and short memory processes. In case (f), we compare a long-memory process and a nonstationary process with a MA component. The rational for these choices was to generate processes with similar sample characteristics. Case (d) is an apparent exception to this rule. In this case, we were simply interested in knowing whether our methods could succeed in distinguishing long memory from no memory models. We performed additional simulations for other models, which are available upon request.

The fractional noise was simulated using the finite Fourier method of Davies and Harte (1987). The four generated series with zero mean and unit variance white noise were grouped into two clusters by the complete linkage algorithm (see, for instance, Johnson and Wichern, 2002) and using the padding, reduced and interpolated Euclidean distances between the log normalized periodogram ordinates.

Table 1 provides the percentages of success on the comparison in cases (a) to (f). The first rows of each cell show the results for the zero-padding periodogram approach. The second rows of each cell show the results for the reduced periodogram approach. The third rows of each cell show the results for the interpolated periodogram approach. For instance, the value 63.4 in the upper-left cell means that 63.4% of the times the two AR(1), $\phi = 0.9$, $n_1 = 50$ and $n_2 = 100$ processes were grouped into one cluster and the two AR(1), $\phi = 0.5$, $n_1 = 50$ and $n_2 = 100$ processes were grouped into another cluster using the zero-padding periodogram method.

The interpolated-periodogram based metric shows a remarkable good performance on the comparisons among stationary processes with ARMA and ARFIMA formulations, and shows a performance that increases significantly with the sample size on the comparison between ARMA and ARIMA processes and between ARIMA and ARFIMA processes.

The zero padding method works well for classifying longer series of similar length. However, it is not able to separate well nearnonstationary processes with large samples from nonstationary processes with short samples, and, more importantly, it does not perform well on the comparison between longer stationary and shorter near-nonstationary ARMA processes. In fact, when sample sizes are very unbalanced, the shorter series periodogram is distorted by the zero-padding method. Zero padding is equivalent to add new ordinate values that are linear combinations of the periodogram ordinates of the original series. Naturally, the resulting statistics and tests suffer from this problem.

The reduced periodogram method is always dominated by the other methods. In particularly, it displays a very poor performance for distinguishing similar processes with small samples.

We also investigated the performance of Kullback-Leibler statistic, but the results are not included as it does not work well. Results are available from the authors upon request.

Table 1
Percentages of success on the comparison of pairs of simulated time series models: Zero-padding (ZP); Reduced periodogram (RP) and Interpolated periodogram (IP)

		(a) AR(1), 0.9 vs. AR(1), 0.5				(b) AR(1), 0.9 vs. ARIMA(0,1,0)			
n_1, n_2	Metric	50,100	200,100	500,250	1000,500	50,100	200,100	500,250	1000,500
50,100	ZP	63.4	74.0	78.1	80.3	20.8	45.6	88.4	97.6
	RP	48.1	52.0	54.2	51.9	12.5	31.2	74.2	92.7
	IP	61.2	73.4	98.4	100.0	16.4	42.4	88.0	99.7
200,100	ZP	87.6	91.4	92.4	93.5	25.6	52.0	88.4	97.2
	RP	71.3	76.9	81.4	82.9	11.0	31.0	67.6	89.2
	IP	84.8	87.9	95.4	99.9	22.8	36.0	76.6	96.4
500,250	ZP	97.6	99.2	99.1	99.3	31.6	61.6	88.4	96.4
	RP	83.6	93.3	97.2	98.7	11.9	25.4	68.7	88.7
	IP	99.1	98.6	99.2	99.9	82.4	58.2	74.8	92.0
1000,500	ZP	98.8	99.6	100.0	100.0	36.4	60.0	84.0	96.8
	RP	91.0	97.4	99.8	99.9	11.1	24.8	67.6	86.7
	IP	100.0	100.0	99.9	100.0	99.8	96.4	79.4	89.0

		(c) AR(2), 0.6, -0.3 vs. MA(2), -0.6, 0.3				(d) ARFIMA(0,0.45,0) vs. White noise			
n_1, n_2	Metric	50,100	200,100	500,250	1000,500	50,100	200,100	500,250	1000,500
50,100	ZP	32.5	44.6	62.8	71.0	39.2	41.6	49.6	57.2
	RP	25.7	31.7	42.4	41.7	28.5	31.7	32.8	34.6
	IP	34.9	49.5	94.7	100.0	45.5	54.6	95.3	100.0
200,100	ZP	40.9	47.8	71.2	83.3	54.8	60.1	70.4	79.3
	RP	30.8	39.9	48.9	54.3	34.8	41.9	51.3	49.6
	IP	55.2	58.8	80.7	98.7	63.8	66.7	82.8	99.4
500,250	ZP	53.0	67.5	77.6	92.3	82.4	89.2	90.1	94.4
	RP	33.2	47.4	63.6	73.6	43.7	57.1	77.0	83.7
	IP	93.4	81.3	88.4	91.3	95.5	87.0	93.7	95.9
1000,500	ZP	57.9	76.1	91.2	95.7	90.8	96.4	98.8	99.1
	RP	32.2	46.3	69.3	84.9	49.1	63.5	88.2	94.1
	IP	100.0	98.5	93.3	98.8	100.0	99.1	98.2	99.5

Table 1
(Continued)

n_1, n_2	Metric	(e) ARFIMA(0,0.45,0) vs. AR(1), 0.95 50,100	200,100	500,250	1000,500	(f) ARFIMA(0,0.45,0) vs. IMA(1,1), 0.4 50,100	200,100	500,250	1000,500
50,100	ZP	72.4	91.5	99.2	99.5	53.3	77.5	92.9	98.2
	RP	53.2	80.7	94.6	97.8	30.2	58.7	83.0	94.7
	IP	63.5	86.3	98.2	100.0	35.6	66.1	94.6	99.9
200,100	ZP	74.1	89.8	99.5	99.8	54.4	73.8	93.9	98.2
	RP	54.8	82.4	97.0	98.9	25.9	52.9	81.9	92.9
	IP	74.9	85.2	95.2	99.7	49.5	63.9	85.9	97.8
500,250	ZP	75.6	90.4	98.8	99.9	54.9	73.7	91.1	96.9
	RP	53.4	80.7	97.5	99.6	25.5	52.7	83.1	92.4
	IP	98.7	93.4	93.9	97.6	94.6	83.7	83.3	93.6
1000,500	ZP	71.3	90.7	99.1	100.0	55.3	68.6	87.3	96.7
	RP	51.6	79.9	97.2	99.8	25.9	52.7	79.9	91.9
	IP	100.0	99.9	96.5	97.8	100.0	99.5	92.0	93.7

4.2. Power and size of the tests

We obtained the estimates of the power and size of the proposed tests for simulated series from the following processes:

(a) AR(1), $\phi = 0.5$ versus AR(1), $\phi = 0.1$, 0.3, 0.5, 0.7, 0.9;

(b) White noise versus AR(1), $\phi = 0$, 0.2, 0.4, 0.6, 0.8;

(c) AR(2), $\phi_1 = 0.5$, $\phi_2 = -0.5$ versus AR(2), $\phi_1 = 0.5$, $\phi_2 = -0.1$, -0.3, -0.5, -0.7, -0.9;

(d) ARMA(1,1), $\phi = 0.2$, $\theta = -0.5$ versus ARMA(1,1), $\phi = 0.2$, $\theta = -0.1$, -0.3, -0.5, -0.7, -0.9;

(e) AR(1), $\phi = 0.5$ versus ARFIMA(1,d,0), $\phi = 0.5$, $d = 0$, 0.1, 0.2, 0.3, 0.4;

(f) AR(1), $\phi = 0.7$ versus AR(1), $\phi = 0.7$, 0.8, 0.9, 1.0.

From these comparisons we are able to see how the tests work for distinguishing similar models with different parameters. From the considerable set of values for the parameters, we can verify whether an increasing difference leads to better test power. As before, the results were based on 1000 replications of each pair of processes. For the parametric approach, we fitted ARMA(m,n) models to the series, with the orders $m = 0, 1, 2, 3$ and $n = 0, 1, 2, 3$ selected by BIC (the AIC does not work in selecting models for hypothesis testing, as noted by Peña and Rodriguez, 2005). Table 2 gives the results for cases (a) to (e) for 10% level of significance and for case (f) for 5% level of significance using the two nonparametric tests (periodogram-distance based test and periodogram-likelihood ratio test) and the parametric test.

For distinguishing series generated by different processes, the parametric test always performs better than both nonparametric tests. However, for comparing series generated by the same processes, the parametric test display a size larger than the one derived from the asymptotic distribution of the statistic.

Both nonparametric tests perform as well as the parametric approach in the large sample cases, specially when the two series were simulated from very different processes. The periodogram-distance based test performs almost always better than the likelihood ratio test. We did not record the size distortions obtained with the parametric test.

5. Application

As an illustration of the possibilities of these techniques, we compared the industrial time series of a set of developed countries. We used monthly data of seasonally adjusted industrial production indices for a large set of European and other industrialized economies. Available data are summarized on Table 3 (source data: Camacho, Pérez-Quiróz and Saiz, 2004). For such large data set, it is unavoidable that sample periods do not coincide. In order to use all available data, it is necessary to apply techniques such as the ones we have described.

In our application, we started by computing the log normalized interpolated periodograms for each of the $k = 30$ production series. The corresponding graphs are shown on Figure 1. We then computed all the corresponding $k(k-1)/2$ pairwise mean-Euclidean distances (Equation 8). In order to be able to interpret resulting data, we used two well-known clustering techniques (see sections 12.3 and 12.5, respectively, of Johnson and Wichern, 2002, for example).

Firstly, we used the multidimensional scaling approach, which creates a configuration of k points in a lower dimensional map (usually two or three). Figure 2 represents the resulting map of distances in two dimensions The first dimension seems to be almost directly related to the countries' development. The second dimension is not easy to interpret. However, looking at the 2-dimensional plot and comparing the relative positions with the periodograms plots, we can make sense of some of the results. Looking at the opposite positions of Cyprus and Ireland, for instance, we realize that this distance comes from very different spectral peaks at different frequencies—the interpolated LNP of Ireland series reaches the minimum value at frequencies $\omega_{29} = 2\pi(29)/85 = 2.14367$ and $\omega_{38} = 2\pi(38)/85 = 2.80895$, whereas the interpolated LNP of Cyprus series is dominated by large peaks at the same frequencies. It can also be seen that the old European Union countries (except Ireland) and the USA, Canada, Japan and Norway are close to each other and far from the new European Union countries and from the then candidate countries (Estonia, Turkey, Slovak Republic, Romania, Lithuania, Slovenia, Czech Republic and Latvia). More developed Poland and Hungary are in an intermediate position.

Secondly, we consider the method of clustering the series by a hierarchical clustering tree (or dendrogram). Figure 3 shows the dendrogram for the industrial production indices series by complete linkage method from which the clusters of countries can be identified. It can be seen at the tree that the interpolated-periodogram based method can group the series into three very reasonable clusters: Cluster 1 = {CN, US, NL, IT, ES, FR, SD, BG, BD, LX, UK, DK, OE, FN, GR, IR, PT, JP, NW}, Cluster 2 = {CY, CZ, SL, LI} and Cluster 3 = {ET, SK, RO, TK, HN, PO, LA}. Cluster 1 includes all the old European Union countries and the USA, Canada, Japan and Norway. Cluster 2 grouped four new European Union countries (Cyprus, Czech Republic, Slovenia and Lithuania). Cluster 3

Table 2
Estimates of power and size of tests of significance: Periodogram-distance based test (DT); Periodogram-likelihood ratio test (LRT) and Parametric test (PT)

(a) AR(1), $\phi = 0.5$ vs. AR(1), $\phi = 0.1, 0.3, 0.5, 0.7, 0.9$ ($\alpha = 10\%$)							
ϕ	Test	[50,50]	[150,75]	[200,200]	[500,250]	[1000,1000]	[2000,1000]
0.1	DT	0.15	0.29	0.56	0.66	0.99	1.00
	LRT	0.15	0.20	0.29	0.34	0.71	0.73
	PT	0.67	0.85	0.99	1.00	1.00	1.00
0.3	DT	0.11	0.18	0.31	0.41	0.90	0.88
	LRT	0.11	0.12	0.15	0.16	0.24	0.24
	PT	0.26	0.34	0.54	0.80	0.99	1.00
0.5	DT	0.08	0.10	0.08	0.09	0.09	0.10
	LRT	0.10	0.10	0.10	0.10	0.10	0.10
	PT	0.19	0.08	0.10	0.06	0.04	0.06
0.7	DT	0.23	0.29	0.62	0.70	1.00	1.00
	LRT	0.15	0.13	0.23	0.22	0.45	0.44
	PT	0.27	0.48	0.68	0.88	1.00	1.00
0.9	DT	0.67	0.86	1.00	1.00	1.00	1.00
	LRT	0.42	0.58	0.93	0.98	1.00	1.00
	PT	0.75	0.98	1.00	1.00	1.00	1.00

(b) White noise vs. AR(1), $\phi = 0, 0.2, 0.4, 0.6, 0.8$ ($\alpha = 10\%$)							
ϕ	Test	[50,50]	[150,75]	[200,200]	[500,250]	[1000,1000]	[2000,1000]
0.0	DT	0.03	0.09	0.03	0.09	0.05	0.08
	LRT	0.10	0.10	0.10	0.10	0.10	0.10
	PT	0.15	0.15	0.17	0.11	0.10	0.04
0.2	DT	0.05	0.10	0.05	0.12	0.10	0.17
	LRT	0.11	0.10	0.13	’.12	0.17	0.18
	PT	0.32	0.44	0.56	0.80	0.99	1.00
0.4	DT	0.10	0.14	0.26	0.36	0.87	0.85
	LRT	0.14	0.16	0.29	0.31	0.60	0.60
	PT	0.60	0.84	0.98	1.00	1.00	1.00
0.6	DT	0.33	0.44	0.87	0.92	1.00	1.00
	LRT	0.28	0.38	0.65	0.76	1.00	1.00
	PT	0.89	1.00	1.00	1.00	1.00	1.00
0.8	DT	0.73	0.86	1.00	1.00	1.00	1.00
	LRT	0.61	0.78	0.99	1.00	1.00	1.00
	PT	0.97	1.00	1.00	1.00	1.00	1.00

Note: Pairs of numbers within square brackets are sample sizes $[n_x, n_y]$.

Table 2
(Continued)

(c) AR(2), ϕ_1= 0.5, ϕ_2= -0.5 vs. AR(2), ϕ_1= 0.5, ϕ_2 ($\alpha = 10\%$)							
ϕ_2	Test	[50,50]	[150,75]	[200,200]	[500,250]	[1000,1000]	[2000,1000]
-0.1	DT	0.10	0.18	0.22	0.28	0.71	0.72
	LRT	0.18	0.19	0.33	0.35	0.73	0.73
	PT	0.59	0.94	1.00	1.00	1.00	1.00
-0.3	DT	0.09	0.12	0.21	0.24	0.62	0.63
	LRT	0.12	0.12	0.16	0.16	0.25	0.24
	PT	0.38	0.38	0.72	0.85	1.00	1.00
-0.5	DT	0.08	0.08	0.07	0.09	0.10	0.10
	LRT	0.10	0.10	0.10	0.10	0.10	0.10
	PT	0.28	0.19	0.13	0.12	0.11	0.10
-0.7	DT	0.20	0.26	0.57	0.64	0.99	1.00
	LRT	0.15	0.14	0.24	0.23	0.44	0.48
	PT	0.46	0.50	0.65	0.92	1.00	1.00
-0.9	DT	0.67	0.85	1.00	1.00	1.00	1.00
	LRT	0.46	0.64	0.95	0.97	1.00	1.00
	PT	0.82	0.99	1.00	1.00	1.00	1.00

(d) ARMA(1,1), $\phi = 0.2$, $\theta = -0.5$ versus ARMA(1,1), $\phi = 0.2$, θ ($\alpha = 10\%$)							
θ	Test	[50,50]	[150,75]	[200,200]	[500,250]	[1000,1000]	[2000,1000]
-0.1	DT	0.18	0.32	0.63	0.73	1.00	1.00
	LRT	0.16	0.22	0.33	0.36	0.73	0.74
	PT	0.64	0.84	0.98	1.00	1.00	1.00
-0.3	DT	0.09	0.14	0.24	0.32	0.79	0.78
	LRT	0.11	0.12	0.16	0.16	0.23	0.23
	PT	0.40	0.30	0.47	0.70	1.00	1.00
-0.5	DT	0.05	0.07	0.06	0.07	0.06	0.07
	LRT	0.10	0.10	0.10	0.10	0.10	0.10
	PT	0.18	0.13	0.11	0.10	0.05	0.06
-0.7	DT	0.09	0.09	0.21	0.28	0.80	0.77
	LRT	0.12	0.14	0.17	0.16	0.29	0.27
	PT	0.28	0.55	0.62	1.00	1.00	1.00
-0.9	DT	0.16	0.21	0.60	0.68	1.00	1.00
	LRT	0.22	0.25	0.48	0.55	0.96	0.96
	PT	0.80	0.98	1.00	1.00	1.00	1.00

Note: Pairs of numbers within square brackets are sample sizes $[n_x, n_y]$.

Table 2
(Continued)

(e) AR(1), $\phi = 0.5$ vs. ARFIMA$(1, d, 0)$, $\phi = 0.5, d$ ($\alpha = 10\%$)							
d	Test	[50,50]	[150,75]	[200,200]	[500,250]	[1000,1000]	[2000,1000]
0.0	DT	0.08	0.11	0.10	0.11	0.10	0.10
	LRT	0.10	0.10	0.10	0.10	0.10	0.10
	PT	0.20	0.16	0.11	0.08	0.10	0.04
0.1	DT	0.13	0.13	0.24	0.29	0.76	0.71
	LRT	0.11	0.11	0.14	0.12	0.15	0.14
	PT	0.22	0.20	0.29	0.39	0.72	0.87
0.2	DT	0.22	0.25	0.59	0.70	1.00	1.00
	LRT	0.13	0.15	0.22	0.24	0.49	0.48
	PT	0.33	0.32	0.69	0.81	1.00	1.00
0.3	DT	0.38	0.46	0.90	0.95	1.00	1.00
	LRT	0.20	0.24	0.46	0.50	0.95	0.94
	PT	0.38	0.60	0.91	0.99	1.00	1.00
0.4	DT	0.53	0.71	1.00	1.00	1.00	1.00
	LRT	0.29	0.40	0.78	0.86	1.00	1.00
	PT	0.60	0.89	1.00	1.00	1.00	1.00
(f) AR(1), $\phi = 0.7$ vs. AR(1), $\phi = 0.7, 0.8, 0.9, 1.0$ ($\alpha = 5\%$)							
ϕ	Test	[50,50]	[150,75]	[200,200]	[500,250]	[1000,1000]	[2000,1000]
0.7	DT	0.05	0.08	0.08	0.07	0.08	0.07
	LRT	0.05	0.05	0.05	0.05	0.05	0.05
	PT	0.06	0.07	0.05	0.04	0.05	0.05
0.8	DT	0.13	0.14	0.36	0.43	1.00	0.92
	LRT	0.06	0.06	0.10	0.10	0.19	0.18
	PT	0.15	0.13	0.29	0.38	0.84	0.98
0.9	DT	0.30	0.47	0.93	0.97	1.00	1.00
	LRT	0.17	0.22	0.49	0.58	0.98	0.98
	PT	0.32	0.54	0.88	0.99	1.00	1.00
1.0	DT	0.50	0.76	1.00	1.00	1.00	1.00
	LRT	0.22	0.47	0.97	0.99	1.00	1.00
	PT	0.59	0.95	1.00	1.00	1.00	1.00

Note: Pairs of numbers within square brackets are sample sizes $[n_x, n_y]$.

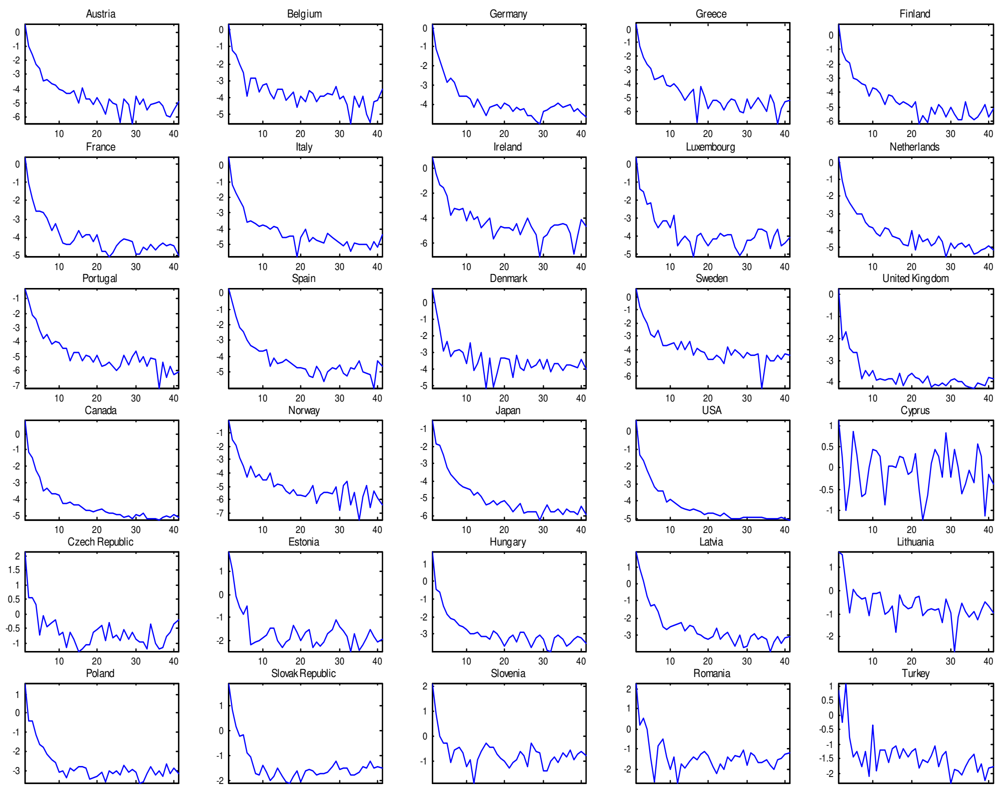

Figure 1. Log normalized interpolated periodograms of 30 European and some developed countries

Table 3
Industrial production indices series (countries and data avaibility)

Country	Code	Sample	n	Country	Code	Sample	n
Austria	OE	62:01-02:12	492	Canada	CN	62:01-03:01	493
Belgium	BG	62:01-03:01	493	Norway	NW	62:01-03:01	493
Germany	BD	62:01-03:01	493	Japan	JP	62:01-03:01	493
Greece	GR	62:01-03:01	493	USA	US	62:01-03:01	493
Finland	FN	62:01-03:01	493	Cyprus	CY	90:01-03:01	142
France	FR	62:01-03:01	493	Czech Republic	CZ	90:01-03:01	142
Italy	IT	62:01-03:01	493	Estonia	ET	95:01-03:01	97
Ireland	IR	75:07-03:01	331	Hungary	HN	90:01-03:01	142
Luxembourg	LX	62:01-03:01	493	Latvia	LA	90:01-03:01	142
Netherlands	NL	62:01-03:01	493	Lithuania	LI	96:01-03:01	85
Portugal	PT	62:01-03:01	493	Poland	PO	90:01-03:01	142
Spain	ES	65:01-03:01	457	Slovak Republic	SK	93:01-03:01	121
Denmark	DK	74:01-03:01	349	Slovenia	SL	90:01-03:01	142
Sweden	SD	62:01-03:01	493	Romania	RO	90:01-03:01	142
United Kingdom	UK	62:01-03:01	493	Turkey	TK	90:01-03:01	142

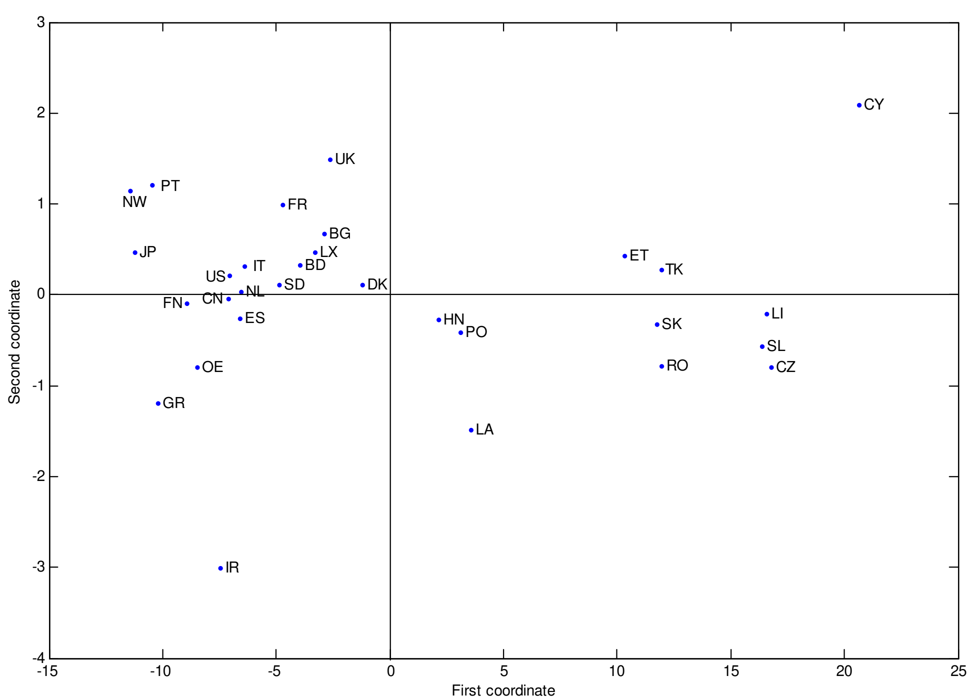

Figure 2. Map of 30 European and some developed countries using the interpolated-periodogram based metric (Multidimensional scaling)

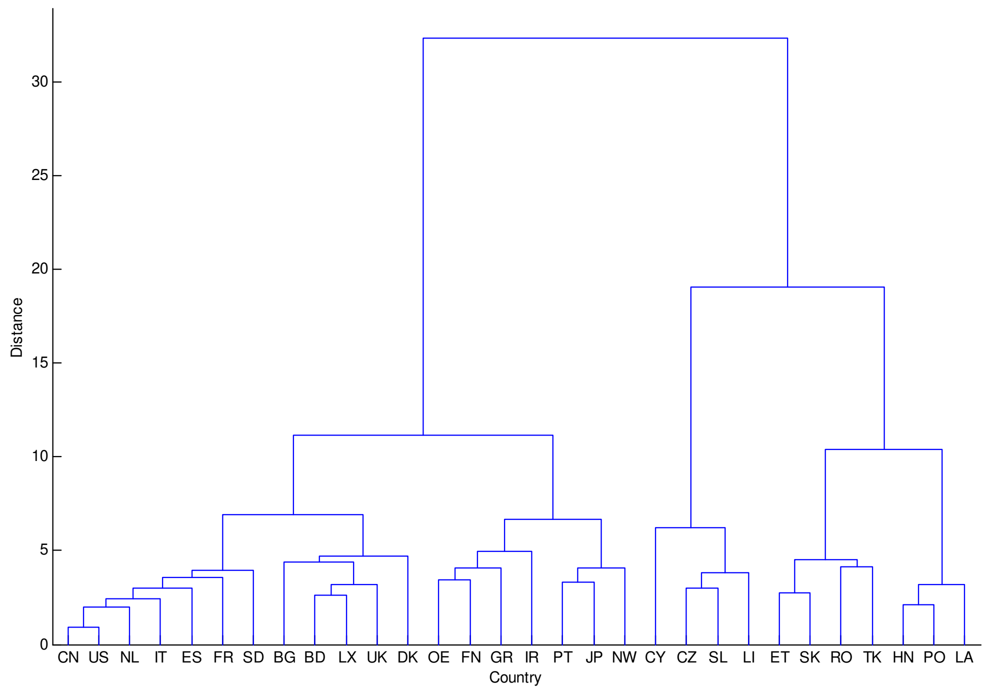

Figure 3. Dendrogram of industrial production series of 30 European and some developed countries using the interpolated-periodogram based metric (Complete linkage algorithm)

includes the other new European Union countries (Estonia, Slovak Republic, Hungary, Poland, Latvia) and the then candidate countries (Romania and Turkey).

These results seem to be very reasonable. Interestingly, they group together the more developed countries. They differ slightly from the ones of Camacho, Pérez-Quiróz and Saiz (2004). They found three clusters. The first includes most of the old European countries and the new European countries Cyprus, Lithuania, Slovenia and Hungary together; the second includes the industrialized countries USA, Canada, United Kingdom, Japan; and the third includes the other new European countries (Latvia, Estonia, Czech Republic and Poland), the candidates countries (Romania and Turkey), and the industrialized country Norway.

6. Concluding remarks

This paper focused on development of spectral-based methods for classification and clustering analysis for time series with unequal length. It proposed zero-padding, reduced periodogram, and interpolated periodogram metrics to deal with the problem of different lengths and, consequently, different periodogram Fourier frequencies.

From a simulation study, it can be concluded that the interpolated periodogram approach performs very well for a wide type of comparisons: (i) stationary processes with similar sample properties, (ii) nonstationary and near nonstationary processes, and (iii) short-memory and long-memory processes. Moreover, in the comparison of time series of very different length, the interpolated periodogram method is preferred to the zero-padding and the reduced periodogram methods. One application to industrial production series also demonstrates the merits of the method.

To formally test whether two time series are generated by stochastic processes with similar properties, we proposed periodogram-distance and spectral likelihood ratio statistics. We found that the power of the periodogram-distance based test is in general greater than the likelihood ratio test. We also considered a parametric approach based on the distance between parameter estimates of the same model. For small samples, we found that the parametric test is more powerful than the periodogram-distance based test, but the former tends to overestimate the size of the test when the two series were simulated from the same process. For large samples, both methods had very high power to distinguish between distinct processes and the estimated sizes were close to the significance levels of the tests. However, contrarily to the periodogram-distance based test, which is easy to implement and computationally fast, the parametric approach needs ad-hoc and computationally heavy ARMA modelling of several time series.

Acknowledgment: This research was supported by a grant from the Fundação para a Ciência e a Tecnologia (FEDER/POCI 2010) and by MEC project SEJ2004-03303, Spain.

REFERENCES

1. Caiado, J., Crato, N. and Peña, D. (2006). "A periodogram-based metric for time series classification", *Computational Statistics & Data Analysis*, 50, 2668-2684.
2. Camacho, M., Pérez-Quiróz, G. and Saiz, L. (2004). "Are European business cycles close enough to be just one?", Bank of Spain, Working Paper n.º 0408.
3. Coates, D. S. and Diggle, P. J. (1986). "Tests for comparing two estimated spectral densities", *Journal of Time Series Analysis,* 7, 7-20.
4. Dargahi-Noubary, G. R. (1992). "Discrimination between Gaussian time series based on their spectral differences", *Communications in Statistics: Theory and Methods*, 21, 2439-2458.
5. Davies, R. B. and Harte, D. S. (1987). "Tests for Hurst effect", *Biometrika*, 74, 95-102.
6. Diggle, P. J. and Fisher, N. I. (1991). "Nonparametric comparison of cumulative periodograms", *Applied Statistics*, 40, 423-434.
7. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton NJ.
8. Jenkins, G. M. and Priestley, M. B. (1957). "The spectral analysis of time series", *Journal of the Royal Statistical Society Series B* , 19, 1-12.
9. Johnson, R. and Wichern, D. (2002). *Applied Multivariate Statistical Analysis*. 5th Ed., Prentice-Hall.
10. Maharaj, E. A. (1996). "A significance test for classifying ARMA models", *Journal of Statistical Computing and Simulation*, 54, 305-331.
11. Maharaj, E. A. (2002). "Comparison of non-stationary time series in the frequency

domain", *Computational Statistics* & *Data Analysis*, 40, 131-141.

12. Peña, D. and Rodriguez, J. (2005). "Detecting non linearity in time series by model selection criteria", *International Journal of Forecasting*, 21, 731-748.
13. Quinn, B. G. (2006). "Statistical methods of spectrum change detection", *Digital Signal Processing*, 16, 588-596.
14. Wang, N. and Blostein, S. (2004). "Adaptive zero-padding OFDM over frequency-selective multipath channels", *Journal on Applied Signal Processing*, 10, 1478-1488.