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, 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, 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: This paper deals with classification and clustering analysis for independent time series with unequal length. A periodogram-based statistic is used to determine whether the time series at hand are generated by the same stochastic mechanism. To deal with the problem of different lengths and consequently that the periodograms compared are calculated at different Fourier frequencies, an interpolation method is proposed. This method consists of a linear interpolation of the individual periodogram ordinates at Fourier frequencies. Nonparametric and parametric test statistics are proposed to test the hypothesis that the two series are generated by the same stochastic mechanism and their random behavior under null are investigated. The performance of the methods is investigated by a Monte Carlo simulation study. As an illustrative example, the interpolated periodogram method is applied to classify industrial production indices series of European and some developed countries.

Keywords: Classification; Cluster analysis; Euclidean metric; Periodogram; Spectral analysis; Time series.

1. Introduction

The classification analysis of time series has useful applications in several fields. In Management, we may be interested in identifying similarities in financial assets for investment and risk management purposes. In Finance, we may interested in identifying dependences in financial market returns for classifying and grouping stocks. In Economics, an application would be the cluster analysis of some countries by looking at the main macroeconomic time series indicators.

The comparison of time series has been studied in literature using both time and frequency domain methods. The classification of time series using spectral analysis approaches were considered by Coates and Diggle (1986), Diggle and Fisher (1991), Dargahi-

Noubary (1992), Diggle and al Wasel (1997), Kakizawa, Shumway and Taniguchi (1998), Maharaj (2002), Caiado, Crato and Peña (2006), among others. However, existing spectral methods for discrimination and clustering analysis of time series cannot be applied directly to series with unequal length.

Our interest in this problem arose from the business cycle study of some industrialized countries done by Camacho, Pérez-Quiróz and Saiz (2004). They had time series of unequal length and had to truncate data in order to compare them. Therefore they used information about truncated time series spectra to compute the distances across countries. This is a common problem with real time series data. We then try and develop a method without this drawback.

Caiado, Crato and Peña (2006) proposed a periodogram based metric for classification of time series and used it to compare near nonstationary and nonstationary time series. We now extend this method for classifying times series with different lengths. For such cases, we know that the Euclidean distance between the periodogram ordinates cannot be used. One possible way to deal with this problem is to interpolate linearly one of the periodograms in order to estimate ordinates of the same Fourier frequencies.

The remainder of the paper is organized as follows. In Section 2, we introduce the interpolation procedure. In Section 3, we present nonparametric and parametric tests of hypothesis to determine whether two series have the same generating process. In Section 4, we present the results of a Monte Carlo simulation study on the performance of the methods. In Section 5, we apply the classification method for clustering some industrialized countries from the information about their industrial production. Section 6 summarizes the paper.

2. Interpolation of the periodogram

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 ordinates of x_t are given by

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

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 $n_x/2$, and the frequency ω is in the range $[-\pi, \pi]$. 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 periodogram ordinates $P_x(\omega_j)$ and $P_y(\omega_p)$ is not adequate for comparison of series x_t and y_t since $m_x \neq m_y$. One way to obtain an adequate distance measure is to interpolate the periodogram ordinates of the series with longer (shorter) length from the series with the shorter (longer) length. Without loss of generatility, 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 $m_y < m_x$. We estimate the periodogram ordinates of x_t as

$$\begin{aligned} P'_x(\omega_p) &= P_x(\omega_r) + (P_x(\omega_{r+1}) - P_x(\omega_r)) \times \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}} \\ &= P_x(\omega_r)\left(1 - \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right) + P_x(\omega_{r+1})\left(\frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right). \end{aligned} \tag{2}$$

Since, asymptotically, $P_x(\omega_j) \sim f_x(\omega_j)\chi^2_{(2)}/2$ where $f_x(\omega_j)$ is the spectral density for each $j = 1, ..., m_x$, it follows that

$$E\left[P'_x(\omega_p)\right] \approx \left(1 - \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right) f_x(\omega_r) + \left(\frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right) f_x(\omega_{r+1}). \tag{3}$$

Noting that $P_x(\omega_r)$ and $P_x(\omega_s)$, $r \neq s$ are asymptotically independently distributed (Priestley, 1981), we have

$$Var\left[P'_x(\omega_p)\right] \approx \left(1 - \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right)^2 f_x^2(\omega_r) + \left(\frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right)^2 f_x^2(\omega_{r+1}). \tag{4}$$

Since the two series $\{x_t\}$ and $\{y_t\}$ are independent realizations, $P_x(\omega_j)$ and $P_y(\omega_p)$ must be independently distributed as well. Thus, we get

$$\begin{aligned} &E\left[P'_x(\omega_p) - P_y(\omega_p)\right] \approx \\ \approx\; &\left(1 - \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right) f_x(\omega_r) + \left(\frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right) f_x(\omega_{r+1}) - f_y(\omega_p), \end{aligned} \tag{5}$$

and, since $Cov\left[P'_x(\omega_p), P_y(\omega_p)\right] = 0$,

$$\begin{aligned} &Var\left[P'_x(\omega_p) - P_y(\omega_p)\right] \approx \\ \approx\; &\left(1 - \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right)^2 f_x^2(\omega_r) + \left(\frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right)^2 f_x^2(\omega_{r+1}) + f_y^2(\omega_p). \end{aligned} \tag{6}$$

Assuming the processes are purely nondeterministic, $|f_x(\omega_{r+1}) - f_x(\omega_r)| \to 0$ as $m_x \to \infty$, for fixed m_y. Then, following Eq. (3),

$$\left|\left(1 - \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right) f_x(\omega_r) + \left(\frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right) f_x(\omega_{r+1})\right| \approx f_x(\omega_r) \tag{7}$$

and hence

$$E\left[P'_x(\omega_p) - P_y(\omega_p)\right] \approx f_x(\omega_p) - f_y(\omega_p). \tag{8}$$

Following Eq. (4), $\frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}} \to 1$ as $m_x \to \infty$, for fixed m_y, and we have

$$\left|\left(1 - \frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right)^2 f_x^2(\omega_r) + \left(\frac{\omega_{p,y} - \omega_{r,x}}{\omega_{r+1,x} - \omega_{r,x}}\right)^2 f_x^2(\omega_{r+1})\right| \approx f_x^2(\omega_r), \tag{9}$$

and hence

$$Var\left[P'_x(\omega_p) - P_y(\omega_p)\right] \approx f_x^2(\omega_p) + f_y^2(\omega_p). \tag{10}$$

We then propose the following distance between the periodogram ordinates of the two series,

$$d_P(x, y) = \sqrt{\frac{1}{m} \sum_{p=1}^{m} \left(P'_x(\omega_p) - P_y(\omega_p)\right)^2}, \tag{11}$$

where m is the number of periodogram ordinates of the series with shorter length (in this case, $m = m_y$). This distance seems to be computationally competitive with other spectrum based distances such as the Cramér-von Mises statistic (Diggle and Fisher, 1991) and the smoothed spectrum distance discussed in Maharaj (2002). If we are not interested in the process scale, before the interpolation procedure we can normalize the periodograms, dividing by the sample variances so that they capture only the autocorrelation structures, $NP'_x(\omega_p) = P'_x(\omega_p)/\widehat{\gamma}_{0,x}$ and $NP_y(\omega_p) = P_y(\omega_p)/\widehat{\gamma}_{0,y}$. Then, 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, as recommended by Caiado, Crato and Peña (2006),

$$d_{LNP}(x,y) = \sqrt{\frac{1}{m}\sum_{p=1}^{m}\left(\log NP'_x(\omega_p) - \log NP_y(\omega_p)\right)^2}. \tag{12}$$

3. Hypotheses testing procedures

3.1. Nonparametric approach

A test of hypotheses to determine whether two time series are realizations of the same stochastic mechanism is suggested. Given two independent stationary series $\{x_t\}$ and $\{y_t\}$, let $P_x(\omega_j), j = 1, ..., m$ and $P_y(\omega_j), j = 1, ..., m$ denote the underlying periodograms, 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 $\{y_t\}$.

We know that if $\widehat{f}_x(\omega_j), j = 1, ..., m$ is the sample spectrum of $\{x_t\}$, then $E\left[\widehat{f}_x(\omega_j)/\widehat{\gamma}_{0,x}\right] = f_x(\omega_j)/\sigma_x^2$ and $Var\left[\widehat{f}_x(\omega_j)/\widehat{\gamma}_{0,x}\right] = f_x^2(\omega_j)/\sigma_x^4$. A good approximation to $E\left[\log\left(\widehat{f}_x(\omega_j)/\widehat{\gamma}_{0,x}\right)\right]$ and $Var\left[\log\left(\widehat{f}_x(\omega_j)/\widehat{\gamma}_{0,x}\right)\right]$ is the sample mean and sample variance of the log normalized periodogram, that is,

$$E\left[\log\left(\widehat{f}_x(\omega_j)/\widehat{\gamma}_{0,x}\right)\right] \approx \frac{1}{m}\sum_{p=1}^{m}\log NP_x(\omega_p) = \overline{x}_{LNP} \tag{13}$$

and

$$Var\left[\log\left(\widehat{f}_x(\omega_j)/\widehat{\gamma}_{0,x}\right)\right] \approx \frac{1}{m}\sum_{p=1}^{m}\left(\log NP_x(\omega_p) - \overline{x}_{LNP}\right)^2 = s^2_{LNP,x}. \tag{14}$$

Similar expressions are given for $E\left[\log\left(\widehat{f}_y(\omega_j)/\widehat{\gamma}_{0,y}\right)\right]$ and $Var\left[\log\left(\widehat{f}_y(\omega_j)/\widehat{\gamma}_{0,y}\right)\right]$. Therefore, we have

$$E\left[\log\left(\widehat{f}_x(\omega_j)/\widehat{\gamma}_{0,x}\right) - \log\left(\widehat{f}_y(\omega_j)/\widehat{\gamma}_{0,y}\right)\right] \approx \overline{x}_{LNP} - \overline{y}_{LNP}, \tag{15}$$

and

$$Var\left[\log\left(\widehat{f}_x(\omega_j)/\widehat{\gamma}_{0,x}\right) - \log\left(\widehat{f}_y(\omega_j)/\widehat{\gamma}_{0,y}\right)\right] \approx s^2_{LNP,x} + s^2_{LNP,y}. \tag{16}$$

Under some suitable conditions, the logarithmic transformation of the sample spectrum will be closer to the normal distribution than to a chi-square distribution (Jenkins and Priestley, 1957). A test of significance for comparison of the log normalized periodograms of the two series is based on the statistic,

$$D_{NP} = \frac{\frac{1}{m}\sum_{p=1}^{m}\left(\log NP_x(\omega_p) - \log NP_y(\omega_p)\right)}{\sqrt{\left(s^2_{LNP,x} + s^2_{LNP,y}\right)/m}}, \tag{17}$$

with asymptotically normal distribution with zero mean and unit variance. For different lengths, $m_x \neq m_y$, the proposed statistic is similarly defined using the interpolated periodogram approach. Jenkins (1961) suggested a similar test. However, he assumes that spectral estimates are based on the same number of autocorrelation lags, even for series with unequal sample data.

3.2. Parametric approach

The problem of comparison of series of unequal length can also be analyzed by a parametric approach in the time domain. 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 mechanism, so that $E[\widehat{\beta}_x] = E[\widehat{\beta}_y] = \beta$. Then $\delta = \widehat{\beta}_x - \widehat{\beta}_y$ for large samples will be a normally distributed vector with zero mean and covariance matrix

$$V_\delta = V_x + V_y, \tag{18}$$

and therefore, we can use the statistic

$$D_P = \delta' V_\delta^{-1} \delta, \tag{19}$$

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 is selected by Akaike's Information Criterion (AIC) or Bayesian's Information Criterion (BIC) selection criterions. If the model obtained is the same for the two time series, then the statistic D_P is computed by using the parameters estimated in each time series. However, if the models selected are different two methods have been proposed:

(i) Fit 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 in general 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) In order to avoid the serious problem of near cancellation of roots, use AR approximations and thus fit to both processes the larger AR model selected and compare the estimated parameters. This method has the problem that we may need a large AR model and very correlated estimated parameters when we have MA generating processes.

Given these problems we propose an alternative approach to apply the parametric test when the selected models are different. We fit both models to both time series and compute the statistic D_P in these two situations. If with the two models the null hypothesis is not rejected, we accept that the processes are generated by the same model. If the 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 least α. In our simulation study, we will use this alternative.

4. Monte Carlo simulations

4.1. Performance study

To illustrate the performance of the interpolated periodogram based metric, two series of different sample sizes, $(n_1, n_2) = \{(50, 100), (100, 100), (200, 100), (500, 250), (1000, 500), (2000, 1000)\}$, were simulated from each of the following processes. So four different series were simulated for each replication on each of the following (a) through (i) comparisons:

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

(b) MA(1), $\theta = -0.9$ versus MA(1), $\theta = -0.5$;

(c) ARMA(1,1), $\phi = 0.5$, $\theta = -0.2$ versus ARMA(1,1), $\phi = 0.2$, $\theta = -0.8$;

(d) AR(1), $\phi = 0.9$ versus ARIMA(0,1,0);

(e) IMA(1,1), $\theta = 0.8$ versus ARMA(1,1), $\phi = 0.95$, $\theta = 0.74$;

(f) ARFIMA(0,0.45,0) versus white noise;

(g) ARFIMA(0,0.45,0) versus ARMA(1,0), $\phi = 0.95$;

(h) ARFIMA(1,0.45,0), $\phi = 0.3$ versus ARIMA(1,1,0), $\phi = 0.3$;

(i) Determinist trend, $x_t = 1 + 0.02t + \varepsilon_t$ versus stochastic trend, $x_t = 0.02 + x_{t-1} + (1 - 0.9B)\varepsilon_t$.

In cases (a), (b) and (c), we compare models of similar type, but with low order parameters and similar autocorrelation functions. In case (d), we compare a nonstationary process and a stationary process with AR parameter value very close to the random walk model. In case (e), we compare nonstationary and near nonstationary processes of Whichern (1974). In cases (f), we compare stationary processes with different characteristics of persistence. In case (g), we compare long-memory and short-memory near nonstationary processes. In case (h), we compare a near nonstationary process with long memory and a nonstationary process. In case (i), we compare the trend-stationary and difference-stationary processes of Enders (1995, p. 252), but with a near unit root in the MA component of the stochastic formulation in order to have the two processes with similar properties. The rational for these choices was to generate processes with similar sample characteristics. Case (f) 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.

The fractional noise were simulated using the finite Fourier method of Davies and Harte (1987). The four generated series with zero mean and unit variance white noise were

Table 1
Percentages of success on the comparison between AR(1) and AR(1)

	AR(1): $\phi = 0.5$					
AR(1): $\phi = 0.9$	(50,100)	(100,100)	(200,100)	(500,250)	(1000,500)	(2000,1000)
(50,100)	61.2	58.6	73.4	98.4	100.0	100.0
(100,100)	73.2	72.1	71.9	95.6	100.0	100.0
(200,100)	84.8	81.1	87.9	95.4	99.9	100.0
(500,250)	99.1	98.0	98.6	99.2	99.9	100.0
(1000,500)	100.0	100.0	100.0	99.9	100.0	100.0
(2000,1000)	100.0	100.0	100.0	100.0	100.0	100.0

Table 2
Percentages of success on the comparison between MA(1) and MA(1)

	MA(1): $\theta = -0.5$					
MA(1): $\theta = -0.9$	(50,100)	(100,100)	(200,100)	(500,250)	(1000,500)	(2000,1000)
(50,100)	43.3	39.2	58.6	95.5	100.0	100.0
(100,100)	36.0	43.9	42.0	88.1	100.0	100.0
(200,100)	58.0	43.1	64.2	85.8	99.6	100.0
(500,250)	93.3	86.6	82.5	92.0	95.3	99.2
(1000,500)	100.0	99.8	99.4	96.3	99.4	99.6
(2000,1000)	100.0	100.0	100.0	99.7	100.0	100.0

grouped into two clusters by hierarchical method of complete linkage using the Euclidean mean distance between the log normalized periodogram ordinates defined in (8). This was repeated 1000 times. The mean percentages of success on the comparison in cases (a) to (i) are provided in Tables 1 to 9, respectively. For instance, in Table 1, the value 61.2 in the upper-left cell means that 61.2% 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.

In the comparisons among stationary processes with ARMA and ARFIMA formulations, the interpolated periodogram based metric shows a remarkable good performance. The simulations results on the comparison between ARMA versus ARIMA processes show a performance that increases significantly with the sample size. The exception to this is case (i), in which the metric is unable to distinguish successfully between trend-stationary and difference-stationary processes of similar length, in particularly for large data samples. This can be easily explained by noting that periodogram of both processes are dominated by a divergence at low frequencies that conceals differences when the sample size is large. For unequal length, the discrimination between the two models works well.

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), $\phi = 0.1, 0.3, 0.5, 0.7, 0.9$;

Table 3
Percentages of success on the comparison between ARMA(1,1) and ARMA(1,1)

ARMA(1,1): $\phi = 0.5,\ \theta = -0.2$	ARMA(1,1): $\phi = 0.2,\ \theta = -0.8$ (50,100)	(100,100)	(200,100)	(500,250)	(1000,500)	(2000,1000)
(50,100)	36.6	26.3	48.7	92.8	100.0	100.0
(100,100)	30.1	31.4	33.8	84.4	100.0	100.0
(200,100)	50.7	32.4	54.7	77.5	98.2	100.0
(500,250)	93.7	82.7	74.5	84.9	92.1	98.7
(1000,500)	100.0	99.8	98.1	91.5	97.1	98.3
(2000,1000)	100.0	100.0	100.0	99.0	98.5	99.8

Table 4
Percentages of success on the comparison between AR(1) and ARIMA(0,1,0)

AR(1): $\phi = 0.9$	ARIMA(0,1,0) (50,100)	(100,100)	(200,100)	(500,250)	(1000,500)	(2000,1000)
(50,100)	16.4	26.2	42.4	88.0	99.7	100.0
(100,100)	11.6	22.7	30.8	78.9	98.3	100.0
(200,100)	22.8	19.4	36.0	76.6	96.4	100.0
(500,250)	82.4	59.8	58.2	74.8	92.0	97.7
(1000,500)	99.8	100.0	96.4	79.4	89.0	96.8
(2000,1000)	100.0	100.0	100.0	99.5	95.3	95.0

Table 5
Percentages of success on the comparison between IMA(1,1) and ARMA(1,1)

IMA(1,1): $\theta = 0.8$	ARMA(1,1): $\phi = 0.95,\ \theta = 0.74$ (50,100)	(100,100)	(200,100)	(500,250)	(1000,500)	(2000,1000)
(50,100)	14.6	11.2	26.6	84.8	100.0	100.0
(100,100)	11.6	11.1	8.2	60.7	100.0	100.0
(200,100)	26.9	10.2	20.6	46.2	92.7	100.0
(500,250)	81.8	60.1	48.7	41.1	54.4	90.3
(1000,500)	99.6	97.3	90.4	62.6	60.4	77.6
(2000,1000)	100.0	100.0	100.0	88.5	76.1	74.1

Table 6
Percentages of success on the comparison between ARFIMA(0,0.45,0) and white noise

ARFIMA(0,0.45,0)	White noise (50,100)	(100,100)	(200,100)	(500,250)	(1000,500)	(2000,1000)
(50,100)	45.5	35.1	54.6	95.3	100.0	100.0
(100,100)	34.8	41.0	40.1	88.1	100.0	100.0
(200,100)	63.8	44.5	66.7	82.8	99.4	100.0
(500,250)	95.5	87.4	87.0	93.7	95.9	99.4
(1000,500)	100.0	100.0	99.1	98.2	99.5	99.5
(2000,1000)	100.0	100.0	100.0	99.9	100.0	100.0

Table 7
Percentages of success on the comparison between ARFIMA(0,0.45,0) and ARMA(1,0)

	ARMA(1,0): $\phi = 0.95$					
ARFIMA(0,0.45,0)	(50,100)	(100,100)	(200,100)	(500,250)	(1000,500)	(2000,1000)
(50,100)	63.5	82.9	86.3	98.2	100.0	100.0
(100,100)	57.7	83.1	86.0	96.0	99.8	100.0
(200,100)	74.9	82.5	85.2	95.2	99.7	100.0
(500,250)	98.7	95.1	93.4	93.9	97.6	99.9
(1000,500)	100.0	100.0	99.9	96.5	97.8	99.6
(2000,1000)	100.0	100.0	100.0	100.0	99.4	99.5

Table 8
Percentages of success on the comparison between ARFIMA(1,0.45,0) and ARIMA(1,1,0)

ARFIMA(1,0.45,0):	ARIMA(1,1,0): $\phi = 0.3$					
$\phi = 0.3$	(50,100)	(100,100)	(200,100)	(500,250)	(1000,500)	(2000,1000)
(50,100)	49.9	70.4	79.0	95.4	99.8	99.9
(100,100)	43.1	67.6	75.1	94.2	98.7	100.0
(200,100)	56.1	64.0	73.0	92.3	97.6	99.9
(500,250)	95.3	85.4	84.0	91.1	95.1	99.4
(1000,500)	100.0	99.8	98.8	93.0	96.0	98.9
(2000,1000)	100.0	100.0	100.0	99.8	97.9	98.8

Table 9
Percentages of success on the comparison between deterministic trend and stochastic trend

	$x_t = 0.02 + x_{t-1} + (1 - 0.9B)\varepsilon_t$					
$x_t = 1 + 0.02t + \varepsilon_t$	(50,100)	(100,100)	(200,100)	(500,250)	(1000,500)	(2000,1000)
(50,100)	16.6	10.0	26.7	84.7	99.9	100.0
(100,100)	9.3	9.8	14.4	58.4	97.8	100.0
(200,100)	24.8	7.0	14.4	36.9	93.7	100.0
(500,250)	80.1	46.2	18.3	2.3	18.3	78.5
(1000,500)	99.9	99.9	88.5	7.8	0.2	6.2
(2000,1000)	100.0	100.0	100.0	57.6	3.8	0.0

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

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

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

(e) 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. 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). Tables 10, 11, 12 and 13 give the results for cases (a), (b), (c) and (d) (stationary versus stationary) for 10% level of significance using the nonparametric test D_{NP} and the parametric test D_P. Table 14 gives the estimated powers and sizes in the case (e) (stationary versus near nonstationary) for 5% level of significance using tests D_{NP} and D_P.

As expected, the tests for large samples are more powerful than the tests for small samples. The power of the parametric test for small samples is larger than the nonparametric test when the processes are distinct. Since the hypothesis testing procedures have asymptotic distribution, the poor performance of the nonparametric test for series of short length is not surprising. For the parametric test, overall estimates of the size for small samples exceed slightly the significance levels when the two series were simulated from the same process, whereas for the nonparametric test the estimated sizes were very close to the significance levels and do not change significantly with the increasing sample. For all the cases, the tests for equal and unequal length give similar results.

5. Application

Monthly data (seasonally adjusted) of industrial production indices series of European economies and some of the most industrialized countries are reported in Table 15 (source: Camacho, Pérez-Quiróz and Saiz, 2004).

5.1. Multidimensional scaling

The technique of multidimensional scaling, also known as principal coordinates analysis, creates a configuration of k points in a map of p dimensions (ideally, two or three dimensions) which gives the Euclidean distances among objects using the information about the similarity (or dissimilarity) matrix. In our application, we have the Euclidean distances between the log normalized interpolated periodogram ordinates of $k = 30$ production series shown in Figure **??**.

Figure 2 represents the map of distances using the 2-dimensional metric scaling. The first dimension seems to be related to development of the countries. The interpretation of the second dimension is not straightforward. However, looking at the interpolated periodograms of the distinct countries Cyprus and Ireland, we see that they have some peaks at different frequencies from each other. Moreover, 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

Table 10
Estimates of power and size of 0.10 test of significance for AR(1), ϕ=0.5 versus AR(1), ϕ=0.1, 0.3, 0.5, 0.7, 0.9

Nonparametric test D_{NP} (equal and unequal lengths)						
ϕ	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
0.1	0.15	0.29	0.56	0.90	0.99	1.00
0.3	0.11	0.17	0.31	0.63	0.90	1.00
0.5	0.08	0.09	0.08	0.09	0.09	0.09
0.7	0.23	0.41	0.62	0.95	1.00	1.00
0.9	0.67	0.96	1.00	1.00	1.00	1.00
ϕ	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
0.1	0.18	0.22	0.30	0.71	0.93	1.00
0.3	0.09	0.17	0.16	0.37	0.63	0.87
0.5	0.08	0.06	0.07	0.08	0.05	0.08
0.7	0.22	0.27	0.38	0.77	0.96	1.00
0.9	0.72	0.85	0.96	1.00	1.00	1.00
Parametric test D_P (equal and unequal lengths)						
ϕ	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
0.1	0.67	0.84	0.99	1.00	1.00	1.00
0.3	0.26	0.39	0.54	0.90	0.99	1.00
0.5	0.19	0.12	0.10	0.06	0.04	0.06
0.7	0.27	0.37	0.68	0.98	1.00	1.00
0.9	0.75	0.96	1.00	1.00	1.00	1.00
ϕ	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
0.1	0.72	0.85	0.89	1.00	1.00	1.00
0.3	0.34	0.34	0.34	0.80	0.97	1.00
0.5	0.12	0.08	0.10	0.06	0.04	0.06
0.7	0.33	0.48	0.50	0.88	1.00	1.00
0.9	0.93	0.98	0.98	1.00	1.00	1.00

Table 11
Estimates of power and size of 0.10 test of significance for ARMA(1,1), ϕ=0.2 and θ=-0.5 versus ARMA(1,1), ϕ=0.2 and θ=-0.1, -0.3, -0.5, -0.7, -0.9

Nonparametric test D_{NP} (equal and unequal lengths)						
θ	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
-0.1	0.21	0.34	0.59	0.96	1.00	1.00
-0.3	0.08	0.15	0.23	0.50	0.79	0.96
-0.5	0.05	0.05	0.07	0.05	0.06	0.06
-0.7	0.08	0.14	0.25	0.48	0.80	0.97
-0.9	0.18	0.30	0.58	0.95	1.00	1.00
θ	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
-0.1	0.19	0.30	0.34	0.71	0.95	1.00
-0.3	0.08	0.11	0.12	0.29	0.50	0.85
-0.5	0.04	0.04	0.03	0.03	0.03	0.05
-0.7	0.06	0.06	0.10	0.22	0.46	0.76
-0.9	0.12	0.12	0.28	0.70	0.97	1.00
Parametric test D_P (equal and unequal lengths)						
θ	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
-0.1	0.64	0.80	0.98	1.00	1.00	1.00
-0.3	0.40	0.26	0.47	0.70	1.00	1.00
-0.5	0.18	0.10	0.11	0.10	0.05	0.06
-0.7	0.28	0.46	0.62	0.91	1.00	1.00
-0.9	0.80	0.97	1.00	1.00	1.00	1.00
θ	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
-0.1	0.59	0.84	0.87	1.00	1.00	1.00
-0.3	0.28	0.30	0.37	0.70	0.98	1.00
-0.5	0.15	0.13	0.12	0.10	0.04	0.06
-0.7	0.49	0.55	0.59	1.00	1.00	1.00
-0.9	0.95	0.98	1.00	1.00	1.00	1.00

Table 12
Estimates of power and size of 0.10 test of significance for white noise versus AR(1), ϕ=0, 0.2, 0.4, 0.6, 0.8

Nonparametric test D_{NP} (equal and unequal lengths)						
ϕ	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
0.0	0.03	0.04	0.03	0.05	0.05	0.04
0.2	0.05	0.05	0.05	0.08	0.10	0.20
0.4	0.10	0.15	0.26	0.59	0.87	0.99
0.6	0.33	0.57	0.87	1.00	1.00	1.00
0.8	0.73	0.97	1.00	1.00	1.00	1.00
ϕ	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
0.0	0.04	0.03	0.04	0.03	0.04	0.03
0.2	0.05	0.05	0.05	0.06	0.07	0.12
0.4	0.10	0.09	0.16	0.35	0.60	0.88
0.6	0.33	0.37	0.60	0.93	1.00	1.00
0.8	0.76	0.89	0.98	1.00	1.00	1.00
Parametric test D_P(equal and unequal lengths)						
ϕ	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
0.0	0.15	0.19	0.17	0.15	0.10	0.05
0.2	0.32	0.31	0.56	0.89	0.99	1.00
0.4	0.60	0.84	0.98	1.00	1.00	1.00
0.6	0.89	1.00	1.00	1.00	1.00	1.00
0.8	0.97	1.00	1.00	1.00	1.00	1.00
ϕ	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
0.0	0.16	0.15	0.16	0.11	0.10	0.04
0.2	0.30	0.44	0.49	0.80	0.96	1.00
0.4	0.68	0.84	0.96	1.00	1.00	1.00
0.6	0.96	1.00	1.00	1.00	1.00	1.00
0.8	1.00	1.00	1.00	1.00	1.00	1.00

Table 13
Estimates of power and size of 0.10 test of significance for ARMA(1,0), ϕ=0.5 versus ARFIMA(1,d,0), ϕ=0.5, d=0, 0.1, 0.2, 0.3, 0.4

Nonparametric test D_{NP} (equal and unequal lengths)						
d	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
0.0	0.08	0.09	0.10	0.10	0.10	0.09
0.1	0.13	0.16	0.24	0.45	0.76	0.89
0.2	0.22	0.36	0.59	0.92	1.00	1.00
0.3	0.38	0.64	0.90	1.00	1.00	1.00
0.4	0.53	0.86	1.00	1.00	1.00	1.00
d	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
0.0	0.07	0.07	0.08	0.08	0.08	0.08
0.1	0.11	0.10	0.16	0.25	0.43	0.71
0.2	0.20	0.22	0.35	0.71	0.96	1.00
0.3	0.33	0.42	0.62	0.97	1.00	1.00
0.4	0.50	0.72	0.88	1.00	1.00	1.00
Parametric test D_P (equal and unequal lengths)						
d	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
0.0	0.20	0.18	0.11	0.08	0.10	0.10
0.1	0.22	0.23	0.29	0.44	0.72	0.95
0.2	0.33	0.39	0.69	0.91	1.00	1.00
0.3	0.38	0.74	0.91	1.00	1.00	1.00
0.4	0.60	0.85	1.00	1.00	1.00	1.00
d	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
0.0	0.11	0.16	0.12	0.08	0.06	0.04
0.1	0.22	0.20	0.23	0.39	0.60	0.87
0.2	0.24	0.32	0.46	0.81	0.99	1.00
0.3	0.49	0.60	0.68	0.99	1.00	1.00
0.4	0.62	0.89	0.91	1.00	1.00	1.00

Table 14
Estimates of power and size of 0.05 test of significance for AR(1), ϕ= 0.7 versus AR(1), ϕ= 0.7, 0.8, 0.9, 1.0

Nonparametric test D_{NP} (equal and unequal lengths)						
ϕ	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
0.7	0.05	0.08	0.08	0.07	0.08	0.08
0.8	0.13	0.21	0.36	0.67	1.00	1.00
0.9	0.30	0.64	0.93	1.00	1.00	1.00
1.0	0.50	0.91	1.00	1.00	1.00	1.00
ϕ	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
0.7	0.05	0.06	0.05	0.05	0.05	0.06
0.8	0.08	0.11	0.16	0.41	0.73	0.95
0.9	0.26	0.43	0.63	0.96	1.00	1.00
1.0	0.46	0.74	0.91	1.00	1.00	1.00
Parametric test D_P (equal and unequal lengths)						
ϕ	(50,50)	(100,100)	(200,200)	(500,500)	(1000,1000)	(2000,2000)
0.7	0.06	0.07	0.05	0.03	0.05	0.04
0.8	0.15	0.10	0.29	0.57	0.84	0.97
0.9	0.32	0.52	0.88	0.99	1.00	1.00
1.0	0.59	0.93	1.00	1.00	1.00	1.00
ϕ	(100,50)	(150,75)	(200,100)	(500,250)	(1000,500)	(2000,1000)
0.7	0.13	0.07	0.06	0.04	0.03	0.05
0.8	0.15	0.13	0.24	0.38	0.73	0.98
0.9	0.45	0.54	0.71	0.99	1.00	1.00
1.0	0.87	0.95	1.00	1.00	1.00	1.00

Table 15
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

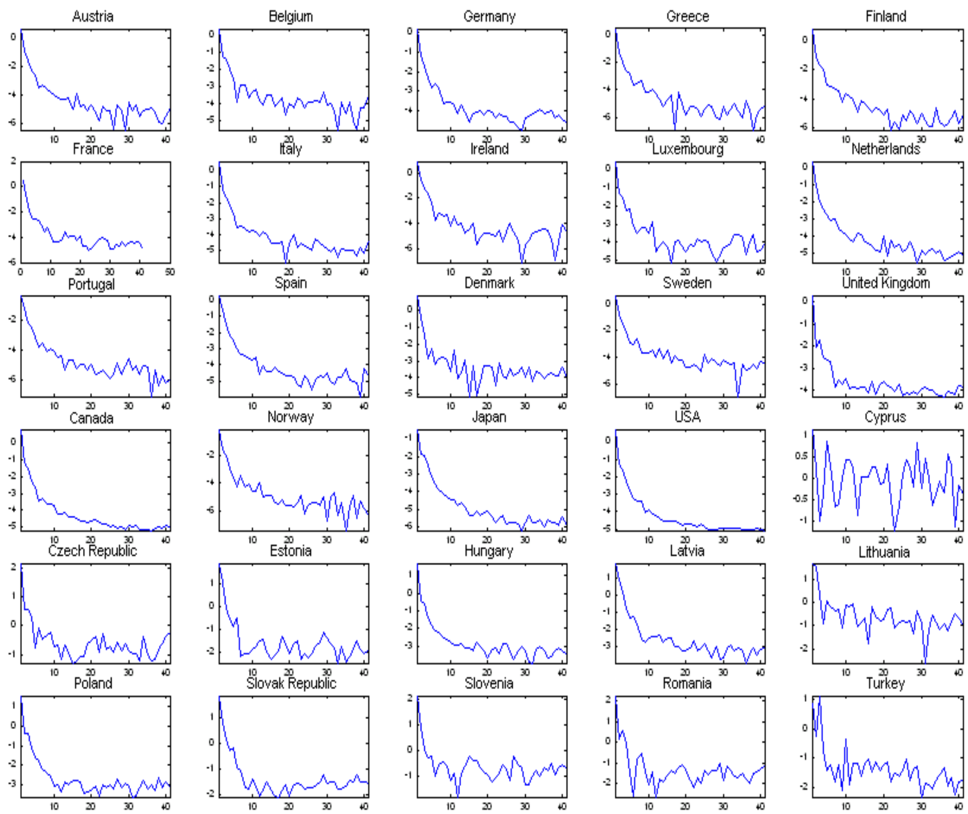

Figure 1. Log normalized interpolated periodograms of industrial production series of 30 European and some developed countries

large peaks at the same frequencies. It can 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 group of the new European Union countries and candidate countries (Estonia, Turkey, Slovak Republic, Romania, Lithuania, Slovenia, Czech Republic and Latvia). The picture shows a small group formed by Poland and Hungary that is very close to those countries of the European Union before the enlargement. We found also in the map two clear outliers Ireland and Cyprus.

5.2. Hierarchical clustering

We consider also the useful method of clustering the series by 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 grouped the series into three very reasonable clusters: Cluster 1=(CN, US, NL, IT, ES, FR, SD,

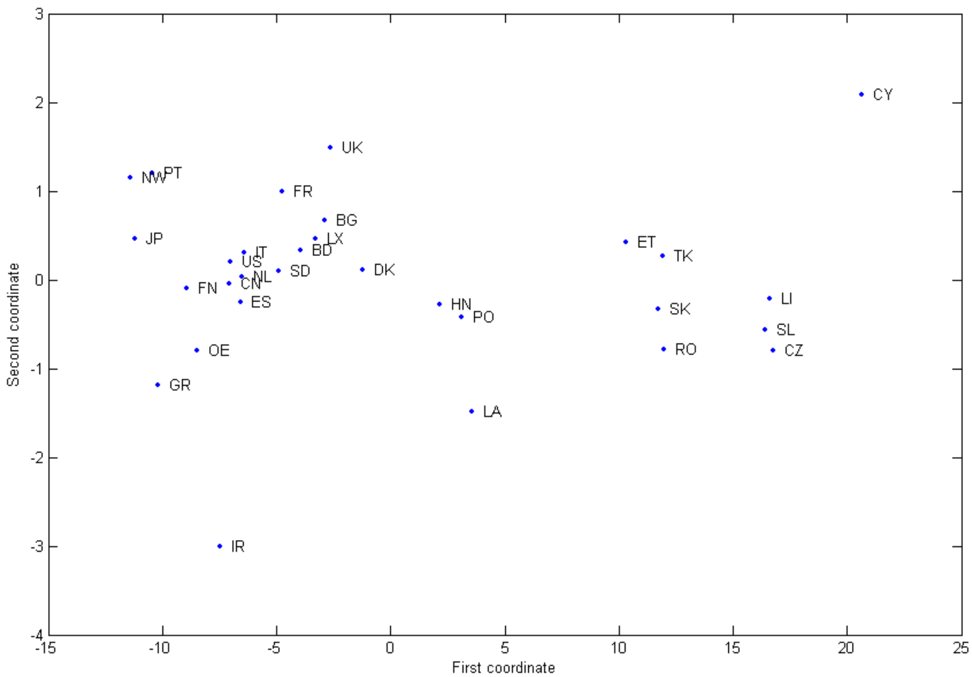

Figure 2. Principal coordinates of 30 countries using the interpolated periodogram based method

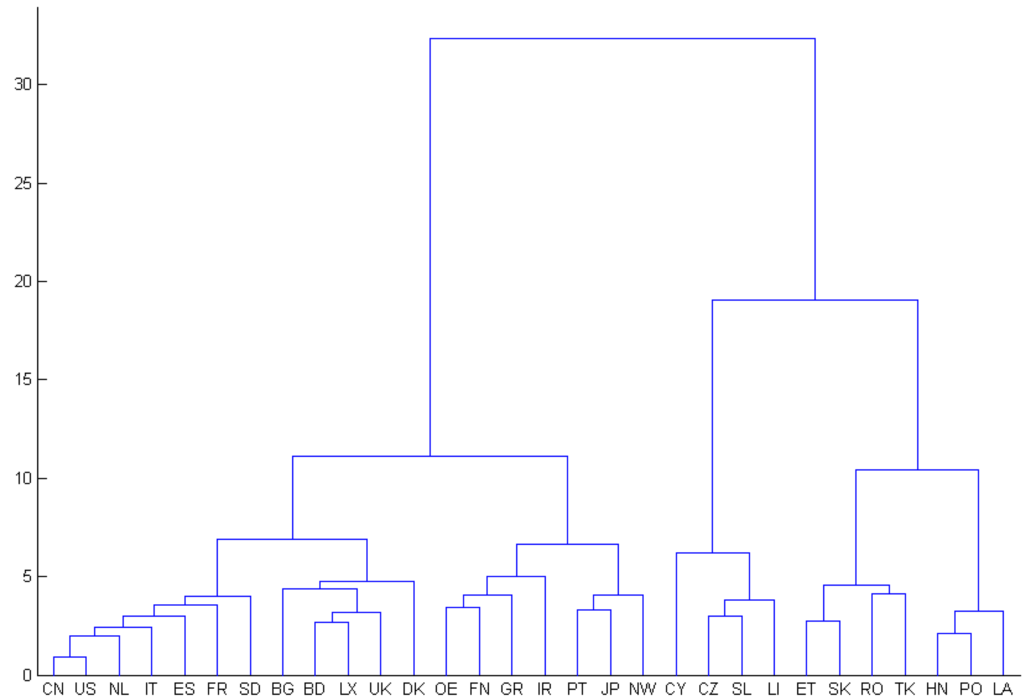

Figure 3. Dendrogram of industrial production series of 30 countries using the interpolated periodogram based method

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

These results differ slightly from the ones of Camacho, Pérez-Quiróz and Saiz (2004). They found a cluster that includes most of the old European countries and the new European countries Cyprus, Lithuania, Slovenia and Hungary together, a cluster formed by the industrialized countries USA, Canada, United Kingdom, Japan, and a cluster formed by the other new European countries, the candidates countries Latvia, Estonia, Slovenia, Czech Republic, Romania, Turkey and Poland, and the industrialized country Norway.

6. Concluding remarks

This paper focus on development of a periodogram-based method and a hypothesis testing procedure for comparison of time series with unequal length. The proposed method is based on the linear interpolation of the individual periodogram ordinates at different Fourier frequencies. It can perform very well for comparing stationary processes with similar sample properties, for comparing nonstationary and near nonstationary processes, and for comparing short-memory and long-memory processes. The estimated power and sizes of the tests for both equal and unequal lengths give similar results, which shows the robustness of the proposed approach. One application to industrial production series also demonstrates the merits of the method.

We also consider a time domain parametric approach based on the distance between parameter estimates of the same model. We found that the parametric test had very high power to distinguish between two distinct processes. However, contrarily to the periodogram-based test, which is easy to implement and computational fast, the parametric approach needs ad-hoc ARMA modelling of several time series.

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

REFERENCES

1. Caiado, J., Crato, N., Peña, D. (2006). "A periodogram-based metric for time series classification", *Computational Statistics & Data Analysis* 50, 2668-2684.
2. Camacho, M., Pérez-Quiróz, G., Saiz, L. (2004). "Are European business cycles close enough to be just one?", Banco de España, Working Paper n.º 0408.
3. Coates, D. S., 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., Harte, D. S. (1987). "Tests for Hurst effect", *Biometrika* 74, 95-102.
6. Diggle, P. J., Fisher, N. I. (1991). "Nonparametric comparison of cumulative periodograms", *Applied Statistics* 40, 423-434.
7. Diggle, P. J., al Wasel, I. (1997). "Spectral analysis of replicated biomedical time series", *Applied Statistics* 46, 31-71.
8. Enders, W. (1995). *Applied Econometric Time Series*. Wiley, New York.
9. Galeano, P., Peña, D. (2000). "Multivariate analysis in vector time series", *Resenhas* 4, 383-404.
10. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Princeton NJ.
11. Jenkins, G. M. (1961). "General considerations in the analysis of spectra", *Technometrics* 3, 133-166.
12. 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.
13. Kakizawa, Y., Shumway, R. H., Taniguchi, M. (1998). "Discrimination and clustering

for multivariate time series", *Journal of the American Statistical Association* 93, 328-340.
14. Maharaj, E. A. (2002). "Comparison of non-stationary time series in the frequency domain", *Computational Statistics* & *Data Analysis* 40, 131-141.
15. Peña, D. and Rodriguez, J. (2005). "Detecting non linearity in time series by model selection criteria", *International Journal of Forecasting* 21, 731-748.
16. Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Academic Press, London.
17. Whichern, D. W. (1973). "The behaviour of the sample autocorrelation function for an integrated moving average process", *Biometrika* 60, 235-239.