Tomáš Cipra

Investigation of periodicity for dependent observations

Aplikace matematiky, Vol. 29 (1984), No. 2, 134–142

Persistent URL: http://dml.cz/dmlcz/104076

INVESTIGATION OF PERIODICITY FOR DEPENDENT OBSERVATIONS

TOMÁŠ CIPRA

(Received May 10, 1983)

It is proved that Hannan's procedure [9] for statistical test of periodicity in the case of time series with dependent observations can be combined with Siegel's improvement of the classical Fisher's test of periodicity. Simulations performed in the paper show that this combination can increase the power of Hannan's test when at least two periodicities are present in the time series with dependent observations.

1. INTRODUCTION

The first exact test for periodic components in time series was proposed by Fisher [5], [6]. It deals with observations $x_1, \ldots, x_n$ arising from the model

$$x_t = \zeta_t + \varepsilon_t\,, \quad t = 1, \ldots, n\,, \tag{1.1}$$

where ζ_t represents the deterministic unobservable component of the series and ε_t is the normal white noise (i.e. $\varepsilon_t \sim$ iid $N(0, \sigma^2)$) representing random errors due to measurement or other sources. When investigating periodicity of the series we are interested in periodic activity of ζ_t. The null hypothesis is that there is no periodic activity

$$H_0 : \zeta_1 = \zeta_2 = \ldots = \zeta_n\,. \tag{1.2}$$

Let $I_n(\lambda, x)$ be the periodogram of x_t defined as

$$I_n(\lambda, x) = a_n^2(\lambda, x) + b_n^2(\lambda, x)\,, \tag{1.3}$$

where

$$a_n(\lambda, x) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} x_t \cos \lambda t\,, \tag{1.4}$$

$$b_n(\lambda, x) = \sqrt{\frac{2}{n}} \sum_{t=1}^{n} x_t \sin \lambda t\,.$$

Fisher's test uses the values of this periodogram only for the frequencies

$$\lambda_j = 2\pi j/n\,, \quad j = 1, \ldots, s\,, \tag{1.5}$$

where s is the integer part of $\frac{1}{2}(n-1)$. These values must be normalized to the form

$$Y_j = I_n(\lambda_j, x)/\sum_{i=1}^{s} I_n(\lambda_i, x) \tag{1.6}$$

to eliminate the effect of σ^2. The hypothesis H_0 in (1.2) is rejected at the given significance level α if

$$\max_{j=1,\ldots,s} Y_j > g_F(n, \alpha)\,, \tag{1.7}$$

where $g_F(n, \alpha)$ is the appropriate critical value calculated according to the exact distributional formula for $\max_{j=1,\ldots,s} Y_j$ given in [1] or [5] and tabulated in [1], [5] or [14]. The use of the rth largest value from $Y_1, \ldots, Y_s$, $r > 1$, in testing for periodicity is discussed in [7]. Walker [15] demonstrated that the dropping of the normality condition has little effect on the large sample distributions of the previous test statistics.

When there is a suspicion of activity at several frequencies in (1.1) (i.e., ζ_t is composed of several periodic components), Siegel's extension of Fisher's test can be used (see [14]) since it has generally higher power for such compound periodicity. Its test statistic has the form

$$T_\lambda = \sum_{j=1}^{s} (Y_j - \lambda g_F(n, \alpha))_+\,, \tag{1.8}$$

where $(t)_+$ denotes $\max(t, 0)$ and λ is a parameter chosen between 0 and 1. The exact distributional formula for T_λ under the null hypothesis (1.2) can be derived and used for the calculation of the critical values $t_\lambda(n, \alpha)$ tabulated in [14], such that H_0 is rejected for

$$T_\lambda > t_\lambda(n, \alpha)\,. \tag{1.9}$$

Simulation studies demonstrate that the most advantageous value for λ is 0·6 (see [14]). It is shown in [13] that the asymptotic null distribution of T_λ is the so called noncentral chi-squared distribution with zero degrees of freedom.

Bølviken (see [2] or [4]) has tried, similarly to Siegel, to increase the power of Fisher's test for the compound periodicity. He has suggested to replace Y_j in (1.6) by $I_n(\lambda_j, x)/\sum_{i=1}^{s-a} I_n(\lambda_{(i)}, x)$, where a is a preselected constant and the periodogram ordinates $I_n(\lambda_1), \ldots, I_n(\lambda_s)$ are ordered so that $I_n(\lambda_{(1)}) \leqq \ldots, \leqq I_n(\lambda_{(s)})$. Experience shows that this modification of Fisher's test consisting in putting away the largest periodogram values in the denominator of Y_j can further increase the power of the test when more than one periodic component are present.

Another improvement of the classical Fisher's test is given in [3]. It is based on the fact that the maximum among $I_n(\lambda_1), \ldots, I_n(\lambda_s)$ is frequently much smaller than the maximum of $I_n(\lambda)$ for all $-\pi \leqq \lambda \leqq \pi$.

The test for periodicity in multiple time series based on the Euclidean norm of the matrix of the periodogram is derived in [10].

All the tests mentioned have the common feature that the null hypothesis H_0 supposes the independence of the observations of the series. In practice this demand is often nonrealistic and therefore tests of periodicity for dependent observations have been looked for. The most important of them are described in Section 2 of this paper. However, the main purpose of this paper is to propagate the combination of Hannan's test from Section 2 with Siegel's test described in this section as a test suitable for dependent observations with compound periodicity in the alternative hypothesis. The theoretical justification of this combination is given in Section 3 while the results of numerical simulations are reported in Section 4.

2. SOME TESTS OF PERIODICITY FOR DEPENDENT OBSERVATIONS

Let observations $x_1, \ldots, x_n$ arise from the model

$$(2.1) \qquad x_t = \zeta_t + u_t\,, \quad t = 1, \ldots, n\,,$$

where $u_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ is a normal linear process with a positive spectral density $f(\lambda)$. Since such a process u_t under general assumptions fulfils the relation

$$(2.2) \qquad I_n(\lambda, u) = 2\pi f(\lambda)\, I_n(\lambda, \varepsilon) + 0(n^{-1/2})$$

(see e.g. [8]), Whittle [16] suggested to replace $I_n(\lambda_j, x)$ in Fisher's test by

$$(2.3) \qquad K_n(\lambda_j, x) = I_n(\lambda_j, x)/\{2\pi f(\lambda_j)\}\,, \quad j = 1, \ldots, s$$

(the corresponding distributional formula for the test statistic then holds, of course, only asymptotically). However, we must know apriori the spectral density $f(\lambda)$ for this procedure.

In the case of an unknown spectral density Whittle [17], [18] recommended to use an estimate $\hat{f}(\lambda)$ instead of $f(\lambda)$ in (2.3) but this approach has a great disadvantage: if the null hypothesis of nonexistence of periodicities in a time series is not true the estimate of $f(\lambda)$ in the neighbourhood of the significant frequencies can be inflated remarkably. This inflation of $f(\lambda)$ can reduce the values of (2.3), i.e. it can reduce the power of the test. Therefore Hannan [9] modified Whittle's approach in such a way that the regression on the harmonic with the frequency λ_j is taken out before computing the estimate of the spectral density at that frequency. The regression mentioned can be approximately carried out by using the corresponding value $I_n(\lambda_j)$ of the periodogram: the previous estimate $\hat{f}(\lambda_j)$ is replaced by

$$f_n^*(\lambda_j) = \frac{\hat{f}_n(\lambda_j) - (2\pi/n)\, w_n(0)\, I_n(\lambda_j)}{1 - (8\pi^2/n)\, w_n(0)}, \tag{2.4}$$

where $w_n(\lambda)$ is the spectral window used for the construction of the estimate $\hat{f}_n$. A certain version of this Hannan's test is given in [11].

We must also mention the so called Bartlett's grouped periodogram test described e.g. in [12]. Here the periodogram values $I_n(\lambda_1), \ldots, I_n(\lambda_s)$ are subdivided into several groups so that the spectral density corresponding to the frequencies in any of these groups can be considered approximately constant. Therefore it is possible to use in each of these groups the classical Fisher's test based on the usual periodogram values for the frequencies in the considered group.

All the above mentioned procedures for the case of dependent observations are certain generalizations of the classical Fisher's test so that they may have rather low power in the case of compound periodicity similarly as the classical Fisher's test. So far no procedure considering this fact has been proposed for testing periodicity in dependent observations, although such a test would be desirable for practical purposes (see e.g. [4]). This has motivated this paper in which we try to combine the above mentioned Hannan's test with Siegel's improvement of Fisher's test described in Section 1. The numerical simulations in Section 4 show that this method can actually improve the power of Hannan's test in the case of compound periodicity.

3. THEORETICAL RESULTS

Let us consider the model (2.1), where the process u_t fulfils the following assumptions:

$$u_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}, \tag{3.1}$$

$$\varepsilon_t \sim \text{iid } \mathrm{N}(0, \sigma^2), \tag{3.2}$$

$$\sum_{j=0}^{\infty} |\alpha_j|\, j^{1/2} < \infty, \tag{3.3}$$

$$f(\lambda) = \frac{\sigma^2}{2\pi} \Big|\sum_{j=0}^{\infty} \alpha_j e^{ij\lambda}\Big|^2 > 0, \quad -\pi \leqq \lambda \leqq \pi. \tag{3.4}$$

The convergence of the sum in (3.3) occurs when e.g. $\alpha_j = o(j^{-3/2})$. If the spectral density (3.4) is even uniformly greater than zero in the interval $-\pi \leqq \lambda \leqq \pi$ then u_t can always be expressed in the form (3.1) (see e.g. [1] or [8]). In particular, the assumptions (3.1)–(3.4) are fulfilled for stationary normal ARMA processes which are important for practical purposes.

The following Lemma shows that in the case of dependent observations the expression Y_j in Siegel's statistic (1.8) can be replaced for large samples by the expression $K_n(\lambda_j, x)/\sum_{i=1}^{s} K_n(\lambda_i, x)$.

Lemma. *Let the assumptions* (3.1)–(3.4) *be fulfilled. Then under the null hypothesis* (1.2),

$$\lim_{n\to\infty} \mathsf{P}\left\{\sum_{j=1}^{s}\left(\frac{K_n(\lambda_j, x)}{\sum_{i=1}^{s} K_n(\lambda_i, x)} - \lambda g_F(n, \alpha)\right)_+ > t_\lambda(n, \alpha)\right\} = \alpha\,. \tag{3.5}$$

Proof. Since $|(y)_+ - (z)_+| \leqq |y - z|$ we can write

$$\sum_{j=1}^{s}\left(\frac{K_n(\lambda_j, u)}{\sum_{i=1}^{s} K_n(\lambda_i, u)} - \lambda g_F(n, \alpha)\right)_+ - \sum_{j=1}^{s}\left(\frac{I_n(\lambda_j, \varepsilon)}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)} - \lambda g_F(n, \alpha)\right)_+ \leqq \tag{3.6}$$

$$\leqq \sum_{j=1}^{s}\left|\frac{K_n(\lambda_j, u)}{\sum_{i=1}^{s} K_n(\lambda_i, u)} - \frac{I_n(\lambda_j, \varepsilon)}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)}\right| = \left|\frac{\sum_{i=1}^{s} K_n(\lambda_i, u)}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)}\right|^{-1}$$

$$\sum_{j=1}^{s}\left|\frac{K_n(\lambda_j,u)\sum_{i=1}^{s} I_n(\lambda_i,\varepsilon) - I_n(\lambda_j,\varepsilon)\sum_{i=1}^{s} I_n(\lambda_i,\varepsilon) + I_n(\lambda_j,\varepsilon)\sum_{i=1}^{s} I_n(\lambda_i,\varepsilon) - I_n(\lambda_j,\varepsilon)\sum_{i=1}^{s} K_n(\lambda_i,u)}{[\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)]^2}\right| \leqq$$

$$\leqq \left|1 + \frac{\sum_{i=1}^{s}[K_n(\lambda_i, u) - I_n(\lambda_i, \varepsilon)]}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)}\right|^{-1}$$

$$\left\{\sum_{j=1}^{s}\left|\frac{K_n(\lambda_j, u) - I_n(\lambda_j, \varepsilon)}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)}\right| + \sum_{j=1}^{s}\frac{I_n(\lambda_j, \varepsilon)}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)}\,\frac{\sum_{i=1}^{s}|K_n(\lambda_i, u) - I_n(\lambda_i, \varepsilon)|}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)}\right\} =$$

$$= 2\left|1 + \frac{\sum_{i=1}^{s}[K_n(\lambda_i, u) - I_n(\lambda_i, \varepsilon)]}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)}\right|^{-1}\frac{\sum_{j=1}^{s}|K_n(\lambda_j, u) - I_n(\lambda_j, \varepsilon)|}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)}$$

We have (see e.g. [1])

$$\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)/s = \sum_{t=1}^{n} \varepsilon_t^2/s$$

so that

$$\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)/s \to 2\sigma^2 \tag{3.7}$$

in probability for $n \to \infty$ according to the weak law of large numbers since s is the integer part of $(n - 1)/2$.

Further,

$$\max_{j=1,\dots,s} |K_n(\lambda_j, u) - I_n(\lambda_j, \varepsilon)| \to 0 \tag{3.8}$$

in probability for $n \to \infty$ according to [8].

Since $K_n(\lambda_j, x) = K_n(\lambda_j, u)$ under the null hypothesis (1.2), we have due to (3.6)–(3.8)

$$\sum_{j=1}^{s} \left(\frac{K_n(\lambda_j, x)}{\sum_{i=1}^{s} K_n(\lambda_i, x)} - \lambda g_F(n, \alpha) \right)_+ - \sum_{j=1}^{s} \left(\frac{I_n(\lambda_j, \varepsilon)}{\sum_{i=1}^{s} I_n(\lambda_i, \varepsilon)} - \lambda g_F(n, \alpha) \right)_+ \to 0$$

in probability for $n \to \infty$ so that (3.5) holds because this formula holds for $K_n(\lambda_i, x)$ replaced by $I_n(\lambda_i, \varepsilon)$, $i = 1, \dots, s$ (the critical values $t_\lambda(n, \alpha)$ converge to a constant for $n \to \infty$ since T_λ has under (1.2) the asymptotic distribution described in Section 1). This completes the proof of Lemma.

On the basis of the previous discussion we suggest in the case of dependent observations with a suspicion of compound periodicity to use the test of H_0 with the critical region

$$\sum_{j=1}^{s} \left(\frac{K_n^*(\lambda_j, x)}{\sum_{i=1}^{s} K_n^*(\lambda_i, x)} - \lambda g_F(n, \alpha) \right)_+ > t_\lambda(n, \alpha), \tag{3.9}$$

where

$$K_n^*(\lambda_j, x) = I_n(\lambda_j, x)/\{2\pi f_n^*(\lambda_j)\}, \quad j = 1, \dots, s, \tag{3.10}$$

and $f_n^*(\lambda_j)$ is defined in (2.4).

4. SIMULATION STUDY

To verify the result of our lemma we performed the simulations given in Table 2. Some simple normal MA or AR models without periodicities were chosen and 100 replications were carried out for each of these models at the computer ADT 4130 at the Department of Statistics of Charles University. The constants $\lambda = 0{\cdot}6$ and $\alpha = 0{\cdot}05$ were chosen. We used critical values taken from [14], which are given in Table 1. Since we have known the spectral density for each of these models we could calculate directly the values $K_n(\lambda_j, x)$ according to (2.3) without the modification

(3.10). The empirical significance levels corresponding to (3.5) are recorded in Table 2. They are really in very good accordance with the used theoretical significance level $\alpha = 0{\cdot}05$ so that our lemma is empirically justified.

Table 1. The used critical values $g_F(n, \alpha)$ and $t_\lambda(n, \alpha)$

n	λ	α	$g_F(n, \alpha)$	$t_\lambda(n, \alpha)$
41	0·4	0·01	0·330	0·222
41	0·6	0·05	0·270	0·116
51	0·4	0·01	0·278	0·194
51	0·6	0·05	0·228	0·0997
81	0·6	0·05	0·157	0·0721

Table 2. The frequencies of rejections of H_0 for models with known spectral densities and without periodicities ($\varepsilon_t \sim$ iid $N(0, 1)$, $\lambda = 0{\cdot}6$, $\alpha = 0{\cdot}05$)

Model	n	Empirical significance level corresponding to (3·5)
$x_t = \varepsilon_t + 0{\cdot}6\varepsilon_{t-1}$	51	0·06
$x_t = \varepsilon_t + 0{\cdot}8\varepsilon_{t-1}$	81	0·05
$x_t = 0{\cdot}5x_{t-1} + \varepsilon_t$	51	0·07
$x_t = 1{\cdot}1x_{t-1} - 0{\cdot}5x_{t-2} + \varepsilon_t$	51	0·07

Table 3 concerns the general case with unknown spectral density when the test (3.9) is recommended. Parzen's estimates of the spectral densities were used with points of truncation ranging from $n/6$ to $n/5$ in accordance with recommendations in the literature (see e.g. [1]). The power of the suggested test (3.9) can be compared with the power of Hannan's test by means of Table 3. One can see that the suggested tests has systematically higher power than Hannan's test in the case of compound periodicity. Under the null hypothesis H_0 the empirical significance levels of both tests exceed unpleasantly the theoretical level α (see first two rows of Table 3), which is caused by the choice of relatively small n in the simulations (the previous results are only asymptotical and, moreover, the estimates of spectral densities are imperfect for small n); nevertheless, also in this case the suggested test gives slightly better results than Hannan's test.

Table 3. The comparison of the power of the test (3·9) and Hannan's test

Model	n	λ	α	Point of truncation in $\hat{f}(\lambda)$	Power of test (3·9)	Power of Hannan's test
$x_t = \varepsilon_t + 0{\cdot}6\varepsilon_{t-1}$	51	0·6	0·05	10	0·18	0·19
$x_t = 0{\cdot}5x_{t-1} + \varepsilon_t$	81	0·6	0·05	15	0·16	0·18
$x_t = \varepsilon_t + 0{\cdot}6\varepsilon_{t-1} + 0{\cdot}75\cos(2\pi\, 5t/n) + 0{\cdot}75\cos(2\pi\, 12\, t/n)$	41	0·6	0·05	8	0·70	0·65
$x_t = \varepsilon_t + 0{\cdot}6\varepsilon_{t-1} + 0{\cdot}75\cos(2\pi\, 5t/n) + 0{\cdot}75\cos(2\pi\, 12\, t/n)$	51	0·6	0·05	10	0·63	0·57
$x_t = 0{\cdot}5x_{t-1} + \varepsilon_t + 0{\cdot}75\cos(2\pi\, 5\, t/n) + 0{\cdot}75\cos(2\pi\, 12\, t/n)$	41	0·4	0·01	8	0·64	0·56
$x_t = 0{\cdot}5x_{t-1} + \varepsilon_t + 0{\cdot}5\cos(2\pi\, 5\, t/n) + 0{\cdot}5\cos(2\pi\, 12\, t/n)$	51	0·6	0·05	10	0·42	0·32
$x_t = 1{\cdot}1x_{t-1} - 0{\cdot}5x_{t-2} + \varepsilon_t + 0{\cdot}5\cos(2\pi\, 5\, t/n) + \cos(2\pi\, 12\, t/n)$	51	0·4	0·01	10	0·67	0·59
$x_t = 1{\cdot}1x_{t-1} - 0{\cdot}5x_{t-2} + \varepsilon_t + 0{\cdot}75\cos(2\pi\, 5\, t/n) + 0{\cdot}5\cos(2\pi\, 12\, t/n)$	51	0·6	0·05	10	0·59	0·50
$x_t = 1{\cdot}1x_{t-1} - 0{\cdot}5x_{t-2} + \varepsilon_t + \cos(2\pi\, 5\, t/n) + \cos(2\pi\, 12\, t/n)$	51	0·6	0·05	10	0·92	0·82

References

[1] *J. Anděl:* Statistical Analysis of Time Series. SNTL, Prague, 1976 (in Czech).

[2] *E. Bølviken:* New tests of significance in periodogram analysis. Scand. J. Statist. 10 (1983), 1—10.

[3] *T. Cipra:* Improvement of Fisher's test of periodicity. Apl. mat. 28 (1983), 186—193.

[4] *E. Damsleth, E. Spjøtvoll:* Estimation of trigonometric component in times series. Journal of the American Statistical Association 77 (1982), 381—387.

[5] *R. A. Fisher:* Tests of significance in harmonic analysis. Proc. Roy. Soc. A 125 (1929), 54—59.

[6] *R. A. Fisher:* The sampling distribution of some statistics obtained from non-linear equations. Annals of Eugenics 9 (1939), 238—249.

[7] *R. A. Fisher:* On the similarity of the distribution found for the test of significance in harmonic analysis and in Steven's problem in geometrical probability. Annals of Eugenics 10 (1940), 14—17.

[8] *E. J. Hannan:* Time Series Analysis. Methuen, London, 1960.

[9] *E. J. Hannan:* Testing for a jump in the spectral function. J. R. Statist. Soc. B 23 (1961), 394—404.

[10] *J. B. Mac Neill:* Tests for periodic components in multiple time series. Biometrika 61 (1974), 57—70.

[11] *D. F. Nicholls:* Estimation of the spectral density function when testing for a jump in the spectrum. Aust. J. Statist. 9 (1967), 103—108.

[12] *M. B. Priestley:* Spectral Analysis and Time Series. Academic Press, London—New York, 1981.

[13] *A. F. Siegel:* The noncentral chi squared distribution with zero degree of freedom and testing for uniformity. Biometrika 66 (1979), 381—386.

[14] *A. F. Siegel:* Testing for periodicity in a time series. Journal of the American Statistical Association 75 (1980), 345—348.

[15] *A. M. Walker:* Some asymptotic results for the periodogram of a stationary time series. J. Aust. Math. Soc. 5 (1965), 107—128.

[16] *P. Whittle:* Hypothesis Testing in Time Series Analysis. Almqvist and Wiksell, Uppsala, 1951.

[17] *P. Whittle:* Test of fit in time series. Biometrika 39 (1952), 309—318.

[18] *P. Whittle:* The statistical analysis of a seiche record. Sears Foundation Journal of Marine Research 13 (1954), 76—100.

Souhrn

VYŠETŘOVÁNÍ PERIODICITY PRO ZÁVISLÁ POZOROVÁNÍ

TOMÁŠ CIPRA

V článku je dokázáno, že Hannanův postup [9] při statistickém testování periodicity v případě časových řad se závislými pozorováními lze zkombinovat se Siegelovým vylepšením [14] klasického Fisherova testu periodicity. Simulace, které byly provedeny, ukazují, že tato kombinace může zvýšit sílu Hannanova testu, když řada se závislými pozorováními obsahuje alespoň dvě periodické složky o různých frekvencích.

Author's address: RNDr. *Tomáš Cipra,* CSc., Matematicko-fyzikální fakulta UK, Sokolovská 83, 186 00 Praha 8.