A METHOD OF DISCRIMINATION IN TIME SERIES ANALYSIS—I*

By A. RUDRA
Indian Statistical Institute, Calcutta

1. Introduction

1.1. In an earlier paper bearing the same title and published elsewhere (Rudra, 1954), the author has outlined a method for discriminating on the basis of a given time series between stationary processes belonging to the autoregressive, the moving average, and the periodic types. For processes of the periodic type, a new model was suggested and used in our argument. The model is characterised by the relation

$$x_t = m_t + \epsilon_t = \Delta+\theta_t(p_1)+\theta_t(p_2)+\dots+\theta_t(p_k)+\epsilon_t \qquad \dots \quad (1.1)$$

$$\left(t = 1, 2, \dots, N;\ \sum_{t=1}^{p_j} \theta_t(p_j) = 0, \quad j = 1, 2, \dots, k\right)$$

where x_t is the observed value at time point t, and $\theta_t(p_j)$ (t to be reduced mod p_j ; $j = 1, 2, \dots, k$) the parametric constants; ϵ_t is an independent variable with zero mean and a constant variance σ^2. The model contains periodicities of lengths $p_1, p_2, \dots, p_k$. The quantity $\Omega(p_i) = \dfrac{1}{p_i-1} \sum_{t=1}^{p_i} \dfrac{\theta_t^2(p_i)}{\sigma^2}$ was termed the 'variance' of the periodicity $p_i (i = 1, 2, \dots k)$ and was suggested as a measure of its importance. The method in a nutshell is to draw up the observed series $x_1, x_2, \dots, x_N$ into a Buys-Ballot table for periodicity p in a chosen range (a, b) as follows :

Buys-Ballot table for Periodicity p

columns	(1)	(2)	...	(s)	...	(p)	
	x_1	x_2	...	x_s	...	x_p	
	x_{p+1}	x_{p+2}	...	x_{p+s}	...	x_{2p}	
	...		...		...	...	... (1.2)
	...		...		...	...	
	...		...		...	...	
	$x_{\overline{n(p)-1}\,p+1}$	$x_{\overline{n(p)-1}\,p+2}$	...	$x_{\overline{n(p)-1}\,p+s}$	...	...	
column means	$\bar{x}_{1\cdot}(p)$	$\bar{x}_{2\cdot}(p)$ ...		$\bar{x}_{s\cdot}(p)$ ...		$\bar{x}_{p\cdot}(p)$	$\bar{x}_{\cdot\cdot}$ (general mean)

where $n(p)$, and s are given by $N = \overline{n(p)-1}\,p+s$, $s<p$.

* Based on a thesis approved for the Ph. D. degree of the London University.

From each such table, an $F(p)$ is to be calculated by using the following formula :

$$F(p) = \frac{\text{between column sum of squares}}{\text{within column sum of squars}} \times \frac{N-p}{p-1} \quad \ldots \quad (1.3)$$

where the expressions 'between column sum of squares' and 'within column sum of squares' mean as in the analysis of variance for one way table.

The $F(p)$'s are then to be plotted against p and the F-diagram thus obtained used as the discriminating criterion. Decisions are to be made by following elaborate rules given in the aforementioned paper as to the occurrence of peaks in the three regions defined in the $\{p, F(p)\}$- space by the curves giving for each value of p the values $F_{\alpha}(p, N-p)$ and $F_{\beta}(p, N-p)$, which are the upper $\alpha\%$ and $\beta\%$ points of the standard F-distribution. (See Diagram 1).

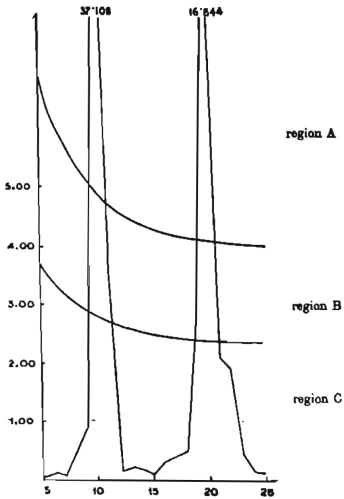

Diagram 1. Artificial cyclical series.
One periodicity : 10. (Kendall)

1.2. In the present paper, we endeavour to justify the intuitive grounds on which the rules of procedure were outlined. We also describe the methods of fitting a Linear Cyclical Model to a series and of Retesting a periodicity after the elimination of other existing periodicities. Only the final results of applying our method to a large number of series were given in the previous paper. In the present paper we provide a large number of the actual diagrams used, and more detailed tables and discussions, so that the actual working of the method under diverse circumstances may be rendered clear.

1.3. In order to maintain the continuity of arguments, we have thought it fit to take out of the body of the text proofs and demonstrations of certain assertions made and mathematical results used and keep them for later publication as Part II of the paper.

2. The distribution of $F(p)$ for a cyclical series

2.1. We shall write (1.1). as

$$x_t = m_t + \epsilon_t \qquad \ldots \quad (2.1)$$

and consider the following two tabular arrangements :

$$\begin{array}{r|cccccccc|l}
\text{columns} & (1) & (2) & \ldots & (s) & \ldots & (p) & & \\
 & m_1 & m_2 & \ldots & m_s & \ldots & m_p & & \\
 & m_{p+1} & m_{p+2} & \ldots & m_{p+s} & \ldots & m_{2p} & & \\
 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & & \\
 & m_{\overline{n(p)-1}\,p+1} & \ldots & \ldots & m_N & \ldots & \ldots & & \\
\hline
\text{column means} & m_1.(p) & m_2.(p) & & m_s.(p) & & m_p.(p) & & m.. \text{ general mean}
\end{array} \qquad \ldots \quad (2.2)$$

and

$$\begin{array}{r|cccccccc|l}
\text{columns} & (1) & (2) & \ldots & (s) & \ldots & (p) & & \\
 & \epsilon_1 & \epsilon_2 & \ldots & \epsilon_s & \ldots & \epsilon_p & & \\
 & \epsilon_{p+1} & \epsilon_{p+2} & \ldots & \epsilon_{p+s} & \ldots & \epsilon_{2p} & & \\
 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & & \\
 & \epsilon_{\overline{n(p)-1}\,p+1} & \ldots & \ldots & \epsilon_N & \ldots & & & \\
\hline
\text{column means} & \epsilon_1.(p) & \epsilon_2.(p) & & \epsilon_s.(p) & & \epsilon_p.(p) & & \epsilon.. \text{ general mean}
\end{array} \qquad \ldots \quad (2.3)$$

These are Buys-Ballot tables constructed out of the series (m_t) and (ϵ_t) $(t=1, 2, \ldots, N)$. For the sake of convenience we shall use the symbols $y_{ij}(p)$, $m_{ij}(p)$ and $\epsilon_{ij}(p)$ to denote the elements belonging to the jth row and the ith columns of (1.2), (2.2) and (2.3) respectively. In other words, we shall rewrite (1.2) as :

$$\begin{array}{r|cccccccc|l}
\text{columns} & (1) & (2) & \ldots & (s) & \ldots & (p) & & \\
 & y_{11}(p) & y_{21}(p) & \ldots & y_{s1}(p) & \ldots & y_{p1}(p) & & \\
 & y_{12}(p) & y_{22}(p) & \ldots & y_{s2}(p) & \ldots & y_{p2}(p) & & \\
 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & & \\
 & y_{1n(p)}(p) & y_{2n(p)}(p) & \ldots & y_{sn(p)}(p) & \ldots & \ldots & & \\
\hline
\text{column means} & \bar{y}_1.(p) & \bar{y}_2.(p) & & \bar{y}_s.(p) & & \bar{y}_p.(p) & & y.. \text{ general mean}
\end{array} \qquad \ldots \quad (2.4)$$

and similarly for (2.2) and (2.3). Then, as a result of (2.1),

$$\begin{aligned} y_{ij}(p) &= m_{ij}(p)+\epsilon_{ij}(p), \\ \bar{y}_{i\cdot}(p) &= m_{i\cdot}(p)+\epsilon_{i\cdot}(p) \\ \bar{y}_{\cdot\cdot} &= m_{\cdot\cdot}+\epsilon_{\cdot\cdot} \end{aligned} \qquad \ldots \quad (2.5)$$

2.2. We shall refer to (1.2), (2.2) and (2.3) as the Y-table, the M-table and the E-table respectively. For each of these the 'between column sum of squares' and the 'within column sum of squares' and their interrelations are shown in the following where $n_i(p) = n(p)$ if $i < s$ and $= n(p)-1$, if $i > s$.

	d.f.	M-table	E-table	Y-table
between column SS.	$p-1$	$\sum_{i=1}^{p}(m_{i\cdot}(p)-m_{\cdot\cdot})^2 n_i(p)$	$\sum_{i=1}^{p}(\epsilon_{i\cdot}(p)-\epsilon_{\cdot\cdot})^2 n_i(p)$	$\sum_{i=1}^{p}\{(m_{i\cdot}(p)-m_{\cdot\cdot})+(\epsilon_{i\cdot}(p)-\epsilon_{\cdot\cdot})\}^2 n_i(p)$
within column SS.	$N-p$	$\sum_{i,j}^{N}(m_{ij}(p)-m_{i\cdot}(p))^2$	$\sum_{i,j}^{N}(\epsilon_{ij}(p)-\epsilon_{i\cdot}(p))^2$	$\sum_{i,j}^{N}\{(m_{ij}(p)-m_{i\cdot}(p))+(\epsilon_{ij}(p)-\epsilon_{i\cdot}(p))\}^2$
total SS.	$N-1$	$\sum_{i,j}^{N}(m_{ij}(p)-m_{\cdot\cdot})^2$	$\sum_{i,j}^{N}(\epsilon_{ij}(p)-\epsilon_{\cdot\cdot})^2$	$\sum_{i,j}^{N}\{(m_{ij}(p)-m_{\cdot\cdot})+(\epsilon_{ij}(p)-\epsilon_{\cdot\cdot})\}^2$

... (2.6)

The three sums of squares from the Y-table, divided by σ^2 (which is the variance of the random component $\epsilon_{ij}(p)$)

$$\frac{\sum_{i,j}^{N}\{y_{ij}(p)-\bar{y}_{\cdot\cdot}\}^2}{\sigma^2},\ \frac{\sum_{i,j}^{N}\{y_{ij}(p)-\bar{y}_{i\cdot}(p)\}^2}{\sigma^2} \text{ and } \frac{\sum_{i=1}^{p}\{\bar{y}_{i\cdot}(p)-\bar{y}_{\cdot\cdot}\}^2 n_i(p)}{\sigma^2} \qquad \ldots \quad (2.7)$$

are therefore distributed as noncentral chi-squares with $N-1$, $N-p$, and $p-1$ degrees of freedom, and non centrality parameters

$$\left.\begin{aligned} (N-1)\beta &= \sum_{i,j}^{N}\frac{\{m_{ij}(p)-m_{\cdot\cdot}\}^2}{\sigma^2} \\ (N-p)\eta(p) &= \sum_{i,j}^{N}\frac{\{m_{ij}(p)-m_{i\cdot}(p)\}^2}{\sigma^2} \\ \text{and} \quad (p-1)\lambda(p) &= \sum_{i=1}^{p}\frac{\{m_{i\cdot}(p)-m_{\cdot\cdot}\}^2 n_i(p)}{\sigma^2} \end{aligned}\right\} \qquad \ldots \quad (2.8)$$

respectively, provided ϵ_i is assumed to be normal.

We shall denote these three noncentral chi-squares by

$\chi^2_{N-1}[\beta]$, $\chi^2_{N-p}[\eta(p)]$, and $\chi^2_{p-1}[\lambda(p)]$ respectively.

(This notation does not agree with that used by Tang (1938), Patnaik (1949), etc. who would write the same as $\chi^2_{N-1}[(N-1)\beta]$, $\chi^2_{N-p}[(N-p)\eta(p)]$ and $\chi^2_{p-1}[(p-1)\lambda(p)]$ respectively). $F(p)$ therefore can be regarded as the ratio of two noncentral chi-squares, multiplied by a constant factor :

$$F(p) = \frac{\chi^2_{p-1}[\lambda(p)]}{\chi^2_{N-p}[\eta(p)]} \cdot \frac{N-p}{p-1}. \qquad \ldots \quad (2.9)$$

2.3. It will be shown in Part II by an easy extension of Cochran's theorem that the two noncentral chi-squares $\chi^2_{p-1}[\lambda(p)]$ and $\chi^2_{N-p}[\eta(p)]$ are independent. The distribution of $F(p)$ therefore can be obtained from that of the ratio of two independent noncentral chi-squares given by Tang (1938), and it is as follows :

$$P[F(p)] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{\lambda(p)^i}{i!2^i} \frac{\eta(p)^j}{j!2^j} \frac{(p-1)^i (N-p)^j}{B\left(i+\frac{p-1}{2}, j+\frac{N-p}{2}\right)} \left\{\frac{\frac{p-1}{N-p}F(p)}{1+\frac{p-1}{N-p}F(p)}\right\}^{p+i-1} \times \left\{\frac{1}{1+\frac{p-1}{N-p}F(p)}\right\}^{N-p+j+1} \qquad \ldots \quad (2.10)$$

For constant p and large $N-p$, $F(p)$ is asymptotically distributed as

$$\frac{\chi^2_{p-1}[\lambda(p)]}{[1+\eta(p)](p-1)} \qquad \ldots \quad (2.11)$$

so that its first four asymptotic cumulants are

$$\left.\begin{aligned} \kappa_1[F(p)] &= \frac{1+\lambda(p)}{1+\eta(p)}, \\ \kappa_2[F(p)] &= \frac{2[1+2\lambda(p)]}{(p-1)(1+\eta(p))^2}, \\ \kappa_3[F(p)] &= \frac{8[1+3\lambda(p)]}{(p-1)^2[1+\eta(p)]^3}, \\ \kappa_4[Fp)] &= \frac{48[1+4\lambda(p)]}{(p-1)^3[1+\eta(p)]^4}. \end{aligned}\right\} \qquad \ldots \quad (2.12)$$

3. *F*-DIAGRAM FOR A SERIES HAVING A SINGLE PERIODICITY

3.1. Though we have found the distribution of a single $F(p)$, it is difficult to obtain the distribution of a collection of $F(p)$'s which is the F-diagram, especially so as they are mutually dependent. We shall however show in Part II that arguments as to the behaviour of the F-diagram under different hypotheses can be carried out in terms of the first moment of $F(p)$ only. If $\pi(p)$ be the first moment of $F(p)$, and if $\pi(p)$ regarded as a function of p between a and b be called the π-diagram, the latter can be regarded as an indicator of the probable behaviour of the F-diagram as to points of occurrence of peaks, and their relative magnitudes. In the present section, we shall study in terms of the π-diagram the probable behaviour of the F-diagram for a Linear Cyclical series having a single periodicity.

3.2. The nature of the π-diagram depends on the relative magnitudes of $\lambda(p)$ and $\eta(p)$. When the series has only one periodicity p_0,

$$m_t = \alpha_t(p_0) = \Delta+\theta_t(p_0).$$

As $\alpha_t(p_0) = \alpha_{t+sp_0}(p_0)$ for any integral s, the M-table for trial period p_0 has the following appearance :

$$\begin{array}{r|ccccccc}
\text{columns} & (1) & (2) & \cdots & (s) & \cdots & (p_0) \\
\hline
 & \alpha_1(p_0) & \alpha_2(p_0) & \cdots & \alpha_s(p_0) & \cdots & \alpha_{p_0}(p_0) \\
 & \alpha_1(p_0) & \alpha_2(p_0) & \cdots & \alpha_s(p_0) & \cdots & \alpha_{p_0}(p_0) \\
 & \vdots & \vdots & & \vdots & & \vdots \\
 & \alpha_1(p_0) & \alpha_2(p_0) & \cdots & \alpha_s(p_0). & & \\
\hline
\text{column means} & \alpha_1(p_0) & \alpha_2(p_0) & \cdots & \alpha_s(p_0) & \cdots & \alpha_{p_0}(p_0)
\end{array} \quad \ldots \quad (3.1)$$

It will be noticed that,

$$\eta(p_0) = 0,$$

and

$$\lambda(p_0) = \frac{N-1}{p_0-1}\,\beta,$$

so that

$$(n(p_0)-1)\,\Omega\,(p_0) \leqslant \lambda(p_0) < n(p_0)\,\Omega\,(p_0).$$

Hence

$$\pi(p_0) = 1+\lambda(p_0) = 1+\frac{N-1}{p_0-1}\,\beta \simeq 1+n(p_0)\Omega(p_0). \quad \ldots \quad (3.2)$$

The only other values of the trial period p for which $\eta(p)$ is zero are the multiples of p_0. Thus for any positive integer s,

$$\eta(sp_0) = 0,$$

$$\lambda(sp_0) = \frac{N-1}{sp_0-1}\,\beta,$$

$$\pi(sp_0) = 1+\frac{N-1}{sp_0-1}\beta = 1+\frac{(p_0-1)\lambda(p_0)}{(sp_0-1)} \simeq 1+\frac{(p_0-1)}{(sp_0-1)}\,n(p_0)\,\Omega(p_0). \quad \ldots \quad (3.3)$$

It should be noted that $\pi(sp_0) < \pi(\overline{s-1}p_0)$ for all s.

3.3. For all other values of p, it is reasonable to suppose that the variations within columns and between columns are more or less the same (as a result of remaining unaffected by the arbitrary grouping in columns which have no phase-relation with the periodicity of the data).

If it is so,

$$\lambda(p) \simeq \eta(p) \simeq \beta \qquad \text{for } p \neq p_0 \neq sp_0;$$

$$\pi(p) = \frac{1+\lambda(p)}{1+\eta(p)} \simeq \frac{1+\beta}{1+\beta} = 1. \qquad \ldots \quad (3.4)$$

Thus, for a series having only one periodicity p_0, the π-diagram will more or less follow the line $\pi(p) = 1$ at all points except $p = p_0$, $2p_0$, $3p_0$ etc., where there will be sharp peaks. The sharpness would depend on the magnitude of β, which would depend on $\Omega\,(p_0)$, the 'variance' of the periodicity, as well as the number of rows $n(p_0)$, which increases with the length of the series N; further, the peaks at p_0, $2p_0$, $3p_0$, etc., will be of progressively diminishing magnitude.

3.4. It is however quite possible for the relation (3.4) not to hold. $\eta(p)$ may in certain cases be much smaller than $\lambda(p)$ making

$$\pi(p) = \frac{1+\lambda(p)}{1+\eta(p)}$$

greater than unity. There is therefore the possibility of observing peaks at points that do not correspond to any real periodicity. Hence it is of importance to study the factors that affect $\pi(p)$. This will be done in detail in Part II. It is sufficient to note here that, as was mentioned in the previous papers, the value of $\pi(p)$ depends on, (i) the actual nature of oscillation of the periodic series $\theta_1(p_0)$, $\theta_2(p_0)$, ..., $\theta_{p_0}(p_0)$, $\theta_1(p_0)$, $\theta_2(p_0)$, ..., $\theta_{p_0}(p_0)$, $\theta_1(p_0)$, ... ; (ii) the number of rows $n(p_0)$; and (iii) the relative values of p and p_0. There is a smaller chance of having a spurious peak at $p \neq sp_0$ if p and p_0 are relatively prime than if they are not.

Though it is possible to have peaks in the π-diagram at points not corresponding to the true periodicity, it can be proved that as long as $p > p_0$, $\pi(p)$ cannot be greater than $\pi(p_0)$.

4. F-Diagram for a series having several periodicities

4.1. The F-diagram for a series having more than one periodicity is more complicated to study under different hypotheses than that for one with a single periodicity.

Suppose that there are k periodicities $p_1, p_2, \ldots, p_k$. When $p=p_i (i=1, 2, \ldots, k)$, the M-table can be written as

$$
\begin{array}{lcccc}
\text{columns} & (1) & (2) & \cdots\cdots & p_i \\
& m'_{11}(p_i)+\theta_1(p_i) & m'_{21}(p_i)+\theta_2(p_i) & \cdots\cdots & m'_{p_i1}(p_i)+\theta_{p_i}(p_i) \\
& m'_{12}(p_i)+\theta_1(p_i) & m'_{22}(p_i)+\theta_2(p_i) & \cdots\cdots & m'_{p_i2}(p_i)+\theta_{p_i}(p_i) \\
& \vdots & \vdots & & \vdots \\
\hline
\text{column means} & m'_{1\cdot}(p_i)+\theta_1(p_i) & m'_{2\cdot}(p_i)+\theta_2(p_i) & & m'_{p_i}(p_i)+\theta_{p_i}(p_i)
\end{array} \quad \ldots \quad (4.1)
$$

where

$$m'_{ts}(p_i) = m_{ts}(p_i)-\theta_t(p_i),$$

$$m'_{t\cdot}(p_i) = m_{t\cdot}(p_i)-\theta_t(p_i),$$

$$(t = 1, 2, \ldots, p_i; \quad s = 1, 2, \ldots, n(p_i)).$$

Thus, even when $p=p_i$, $\eta(p_i)$ does not vanish. But because there is in the t-th column an element $\theta_t(p_i)$ common to each row, we may reasonably expect the variation within column to be less than when there are no elements common to the different rows in the same column. The same observations might be made for the case when $p=sp_i$. On the other hand, we may reasonably expect that when $p \neq s\, p_i$, $\lambda(p) \simeq \eta(p) \simeq \beta$. In other words, we may expect the following relations to hold :

$$\pi(p) = \frac{1+\lambda(p)}{1+\eta(p)} \simeq \frac{1+\beta}{1+\beta} = 1, \quad \ldots \quad (4.2)$$

when

$$p \neq sp_i \quad (i = 1, 2, \ldots, k; \quad s = 0, 1, \ldots);$$

and

$$\eta(p) < \beta, \quad \lambda(sp_i) > \beta,$$

so that

$$\pi(p) = \frac{1+\lambda p)}{1+\eta(p)} > 1 \quad \ldots \quad (4.3)$$

when

$$p = sp_i.$$

4.2. Of course these relations may not always be satisfied. It can be imagined that in certain cases even for $p \neq sp_i$, $\lambda(p)$ will be considerably larger than $\eta(p)$, and then we shall have a peak at a point which does not correspond to a real periodicity. It is however reasonable to assume that this peak will be of less magnitude than those at the more important of the peaks at $p_i (i = 1, 2, \ldots, k)$. It can also be imagined that only the more important of the periodicities $p_i (i=1, 2, \ldots, k)$ will be marked by prominent peaks and that some of the less important ones may fail to produce any peaks at all. In exactly the same way as for the case of a single period, the F-diagram is also likely to have peaks at the multiples sp_i of the actual periods p_i. The peaks at these points should in the ordinary case be less pronounced than those at the actual period p_i, but if the same point happens to be a multiple of two periodicities p_i and

$p_j(i \neq j)$, then the peak at $sp_i = rp_j$ may be more pronounced than either of the peaks at p_i and p_j. Further references to the case when there is more than one periodicity in the series will be made in Part II. Our conjectures as to the occurrence and relative magnitudes of peaks for series having general periodicities will be found on the whole correct regarding quite a few artificial series discussed in section 8.

5. F-Diagram for a linear regressive series

5.1. The distribution of $F(p)$ for a linear regressive series has not been studied, but it will be shown in Part II that the first two moments are asymptotically given by

$$\mu_1\{F(p)\} = 1 - n(p)\bar{\rho},$$

$$\mu_2\{F(p)\} = \frac{2}{p}\left(1 - \frac{p}{p-1} n(p)\bar{\rho}\right)^2\left(1 + \overline{p-1}\ \bar{\rho}^2_{ij}(p)\right), \quad \ldots \quad (5.1)$$

where $\rho_1, \rho_2, \ldots, \rho_{N-1}$ are the auto-correlations of the process, and

$$\bar{\rho} = \frac{2}{N(N-1)}\left\{\sum_{i=1}^{N-1} (N-i)\rho_i\right\}, \quad \ldots \quad (5.2)$$

and

$$\bar{\rho}^2_{ij}(p) = \frac{2}{p(p-1)} \sum_{i,j}^{p} \rho^2_{ij}(p),$$

where $\rho_{ij}(p)$ is the correlation between $(\bar{y}_{i.}(p) - \bar{y})$ and $(\bar{y}_{j.}(p) - \bar{y}_{..})$, $(i, j = 1, 2, \ldots, p)$. It will also be shown that $\bar{\rho}$ can never be less than $-\frac{1}{N-1}$, and this, for large N, is very near to zero. (The upper bound for $\bar{\rho}$ in the case of a moving average model is also very near to zero). The first moment of $F(p)$ will therefore be for all values of p at or below the level of unity; and the second moment will be of the order of $\frac{2}{p-1}$ which is the second moment for a random series.

5.2. It is reasonable to expect on the basis of these results that the F-diagram for a linear regressive series will not in general have prominent peaks and will on the whole lie at a lower level than that for a random series. There can of course be chance peaks, but they will not generally be as prominent as those for a cyclical series. That our intuitive reasoning is not incorrect will be clear from a comparison of diagrams 5 to 13 with diagrams 1 to 4. This is one point where our method is clearly superior to that of Schuster. It is well known that Schuster's Periodogram gives for linear regressive series spurious peaks of the same order of magnitude as for cyclical series.

6. F-DIAGRAM AS A DISCRIMINATOR

6.1. The analysis of the last few sections, we believe justify the method of discrimination we have outlined in our previous paper. We have found that, for cyclical series, there are likely to be peaks at the true periodicities, their sharpness depending on the strength of the periodicities. We have also found that there are likely to be spurious peaks at the multiples of true periodicities, and that it is possible for both cyclical and linear regressive series to produce spurious peaks of comparatively smaller prominence at any point whatsoever. It can also be surmised that in a F-diagram for a series having several periodicities, peaks due to less important periodicities may get damped or even quite obliterated due to the influence of the other periodicities. It is to take care of all these disturbing factors that the device for a cautionary zone and the technique of Retest have been introduced.

6.2. The justification for using probability points of the standard F-distribution to define the regions in $\{p, F(p)\}$ space lies in the fact that departures from randomness (for which hypothesis the use of these points is valid) towards linear cyclical processes and towards linear regressive processes have exacty opposite effects on the F-diagram.

The choice of α and β has of course got to be arbitrary: they cannot possibly have any precise meaning as to the chance of correct decision with regard to any specific hypothesis. But then, a little reflection will show that, for the very large number of hypotheses we have chosen for our field of discrimination it is impossible to have any method whatsoever that will give correct decisions in a specified proportion of cases.

6.3. We shall like to mention here two points which are of some technical interest in the F-analysis of a given series by our method.

(1) If a series is linearly cyclical, then, comparison between the linear regressive and the linear cyclical models when all the peaks in the F-diagram are in Region B, may not be possible owing to there being not any optimum linear regressive fit. Hence if we find that the serial correlations over more than a reasonable number of stages are being significant, we may abandon the sequence and decide in favour of the cyclical model.

(2) The method of obtaining residuals from a moving average fit is to express the moving average as an autoregressive, and to obain the residual from the fit in the usual way.

7. FITTING A LINEAR CYCLICAL MODEL TO DATA AND THE TECHNIQUE OF RE-TESTING

7.1. When it has been finally decided that a series is cyclical with certain definite periodicities, the next step is to fit a linear cyclical model to the data. Suppose we have decided on a model $M(p_1, p_2, ..., p_k)$. The problem now is of estimating the parameters Δ, $\theta_1(p_i)$, $\theta_2(p_i)$, ..., $\theta_{p_i}(p_i)$ $(i = 1, 2, ..., k)$ in the equation

$$x_t = \Delta+\theta_t(p_1)+\theta_t(p_2)+...+\theta_t(p_k)+\epsilon_t. \quad ... \quad (7.1)$$

The least squares approach requires the solution of the following set of equations :

$$\left.\begin{aligned} &\frac{\partial}{\partial\Delta}\sum_{t=1}^{N}\left\{x_t-\Delta-\theta_i(p_1)-\theta_i(p_2)-\ldots-\theta_i(p_k)\right\}^2=0,\\ &\frac{\partial}{\partial\theta_i(p_j)}\sum_{t=1}^{N}\left\{x_t-\Delta-\theta_i(p_1)-\theta_i(p_2)-\ldots-\theta_i(p_k)\right\}^2=0\\ &\qquad (i=1,2,\ldots,p_j;\quad j=1,2,\ldots,k). \end{aligned}\right\}\quad \ldots \quad (7.2)$$

If we write s_i for the remainder when N is divided by $p_i(i=1, 2,\ldots, k)$, the first equation is equivalent to

$$\begin{aligned} x_{..} = \Delta+\frac{1}{N}\Big[&\{\theta_1(p_1)+\theta_2(p_1)+\ldots+\theta_{s_1}(p_1)\}\\ &+\{\theta_1(p_2)+\theta_2(p_2)+\ldots+\theta_{s_2}(p_2)\} \qquad \ldots \quad (7.3)\\ &+\\ &+\{\theta_1(p_k)+\theta_2(p_k)+\ldots+\theta_{s_k}(p_k)\}\Big] \end{aligned}$$

since $\sum_{i=1}^{p_i}\theta_i(p_i)=0 \quad (i, = 1, 2, \ldots, k).$

The typical equation of the remaining set reduces to

$$\bar{x}_{i.}(p_j)=\Delta+\theta_i(p_j)+\delta_{i.}(p_j) \qquad \ldots \quad (7.4)$$

$$(i=1, 2, \ldots, p_j;\ j=1, 2, \ldots, k)$$

where $\delta_{i.}(p_j)=m_{i.}(p_j)-\theta_i(p_j)-\Delta.$

The above set of equations contain too many constants to be solved by any direct means, and it is suggested that the method of iteration be used, taking $\bar{x}_{..}$ as the first approximation to Δ, and $\bar{x}_{i.}(p_j)$ as the first approximation to $\Delta+\theta_i(p_j)$. This first approximation $\bar{x}_{i.}(p_j)$ will of course be the actual solution for $\Delta+\theta_i(p_j)$ when p_j is the only periodicity present in the series. Another situation, when the first approximations in the iterative procedure happen to be the actual solutions, is when $p_1, p_2, \ldots, p_k$ are all relatively prime to each other and N is a common multiple of each of them. Under these circumstances, $\delta_{i.}(p_j)=0 (i=1, 2, \ldots, p_j; j=1, 2, \ldots, k)$. Even if N is not a common multiple of $p_1, p_2, \ldots, p_k$, $\delta_{i.}(p_j)\rightarrow 0$ as N is increased, provided $p_1, p_2, \ldots, p_k$ are relatively prime (Proof given in Part II). But only in rare circumstances shall we have such an ideally suitable collection of $p_1, p_2, \ldots, p_k$ and N. If $p_1, p_2, \ldots, p_k$ are mutually prime, but N is not a common multiple of them, nor is it very large, it is possible to achieve the simplicity of the solution by adjusting the length N such that N is a common multiple of the periodicities, adjustment being done by rejecting some initial or end observations. The utility of this device would of course depend on the number of observations that would be lost in the process. Estimation would be less accurate if based on a smaller number of observations, but saving in computational labour would be enormous.

7.2. The idea of re-testing is as follows:

Suppose, that we unconditionally accept the periodicity p_1 and want to test whether the periodicity p_2 can also be considered as significant.

The test criterion yielded by the Likelihood Ratio method is

$$\frac{\sum_{t=1}^{N}\{x_t-\hat{\Delta}'-\hat{\theta}'_t(p_1)\}^2-\sum_{t=1}^{N}\{x_t-\hat{\Delta}-\hat{\theta}_t(p_1)-\hat{\theta}_t(p_2)\}^2}{\sum_{t=1}^{N}\{x_t-\hat{\Delta}-\hat{\theta}_t(p_1)-\hat{\theta}_t(p_2)\}^2} \quad \ldots \quad (7.5)$$

where $\hat{\Delta}'\hat{\theta}'_t(p_1)$ $(t=1, 2, \ldots, p_1)$ are the maximum likelihood estimates of the parameters in the model $M(p_1)$ and $\hat{\Delta}$, $\hat{\theta}_t(p_1)$ $(t=1, 2, \ldots, p_1)$ and $\hat{\theta}_t(p_2)$ $(t=1, 2, \ldots, p_2)$ the maximum likelihood estimates of the parameters in the model $M(p_1, p_2)$. There are two situations when the criterion (7.5) takes particularly simple forms.

(a) If p_1 and p_2 are relatively prime, and if N is a common multiple of p_1 and p_2, the least squares estimates $\hat{\Delta}$, $\hat{\theta}_t(p_1)$, $\hat{\theta}_t(p_2)$, and $\hat{\Delta}'$, $\hat{\theta}'_t(p_1)$ are given by

$$\hat{\Delta} = \hat{\Delta}' = \bar{x}_{..},$$

$$\hat{\theta}_t(p_1) = \hat{\theta}'_t(p_1) = \bar{x}_{t.}(p_1)-\bar{x}_{..}; \ (t=1, 2, \ldots, p_1),$$

$$\hat{\theta}_t(p_2) = \bar{x}_{t.}(p_2)-\bar{x}_{..}; \quad (t=1, 2, \ldots, p_2), \quad \ldots \quad (7.6)$$

so that (7.5) reduces to

$$\frac{\sum_{t=1}^{p_2}\{\bar{x}_{t.}(p_2)-\bar{x}_{..}\}^2 n_t(p_2)}{\sum_{t=1}^{N}(x_t-\bar{x}_{..})^2-\sum_{t=1}^{p_1}(\bar{x}_{t.}(p_1)-\bar{x}_{..})^2 n_t(p_1)-\sum_{t=1}^{p_2}\{\bar{x}_{t.}(p_2)-\bar{x}_{..}\}^2 n_t(p_2)} \quad \ldots \quad (7.7)$$

which we may treat as $F_{p_2-1, N-p_1-p_2+1}\times\frac{p_2-1}{N-p_1-p_2+1}$ under the null hypothesis. The result holds approximately true even if N is not a common multiple of p_1 and p_2 provided it is large.

(b) The second situation is when p_2 is a multiple of p_1. Let $p_2 = 2p_1$. It will be seen that only $2p_1$ out of the $3p_1$ parameters involved in the model can be estimated, as the least square equations (7.2) are not all independent. (7.5) will be found to reduce to

$$\frac{\sum_{t=1}^{2p_1} \{\bar{x}_t.(2p_1)-\bar{x}..\}^2 n_t(2p_1) - \sum_{t=1}^{p_1} \{x_t.(p_1)-\bar{x}..\}^2 n_t(p_1)}{\sum_{t=1}^{N} (x_t-\bar{x}..)^2 - \sum_{t=1}^{2p_1} \{x_t(2p_1)-\bar{x}..\}^2 n_t(2p_1)} \quad \dots \quad (7.8)$$

which we may treat as a $F_{p_1, N-2p_1} \times \frac{p_1}{N-2p_1}$ under null hypothesis. For all other values of p_1 and p_2, it is suggested that $\theta_t(p_1)$, $(t=1, 2, \dots, p_1)$ be estimated by $\{x_t.(p_1)-\bar{x}..\}$ $(t=1, 2, \dots, p_1)$. Then, let this estimate $\hat{\theta}_t(p_1)$ $(t=1, 2, \dots, N)$ (t reduced mod p) be subtracted from x_t $(t=1, 2, \dots, N)$, and let the residual series be subjected to a fresh F-analysis. If the recalculated F at p_2; which we may denote by $F(p_2/p_1)$ be significant, we accept p_2 as a real periodicity; if not, we reject the peak at p_2 as a spurious effect. The significance of a Retest F can be judged by comparing it with the lower significance line; that is what we have done in our illustrative examples.

8. Illustrations: artificial series

8.1. In the present section we shall discuss in detail the application of our method of discrimination to some artificially constructed Linear Cyclical and Linear Regressive series, and also illustrate the methods of fitting and that of retesting. The final results were provided without commentary in our previous paper. In all the examples we have chosen α and β to be 0.01 and 0.0001 respectively. The choice was made on our finding that cyclical series in general produce extremely sharp peaks, while there is considerable danger of choosing as significant peaks due to series that are purely random if the levels are lowered much further. We have chosen these figures after obtaining and studying our F-diagrams, and therefore, the arbitrary nature of the classifications needs no emphasis.

8.2. There are in all sixteen series. The first is a cyclical series given in Kendall (1946). The next six are artificial series constructed by the author, the first three being cyclical and the next three being moving averages. The remaining are all autoregressive series given by Kendall (1949). Diagrams 1 to 16 are the F-diagrams for these series. Table 1 summarises the actions and decisions taken on the basis of the diagrams in a self explanatory way.

8.3. Diagram 1 faithfully bears out everything we have said in section 3. The series is cyclical with a single periodicity. The diagram has only two peaks, one at the true value of the periodicity 10, and another at its multiple 20, the second peak being smaller in magnitude than the first.

8.4. The series for diagrams 2 to 4 have been built up in the following way. First, a random series with a rectangular distribution was taken; to it was added a cycle of length 8; this gave our series 3 (diagram 4). To this was then added a cycle of length 12 to give our series 2 (diagram 3). Lastly, series 1 (diagram 2) was obtained from series 2 by adding to the latter a cycle of length 5. Thus the three series have the same random part, and are well suited for studying the effect on the diagram of the simultaneous existence of several periodicities of different strengths. It will be observed that $F(8)$ in diagram 3 is less prominent than in diagram 4 due to the influence of cycle 12, and still more so in diagram 2 due to the added influence of the more powerful cycle 5. Also interesting is the fact that the cycle 12 fails to produce a peak in any of the diagrams. This cycle has the least variance of all, and its effect is completely obscured. Again in diagram 2, the more important cycle 8 produces a sharper peak than 5, as we expect.

8.5. Diagram 2 shows three peaks in region A, at 5, 8 and 16, and two in region B, at 10 and 25. According to our rules, we decide that the series is periodic. We unconditionally accept the values 5 and 8 and treat the peaks at 10, 16, and 25, with caution as they are multiples of 5 and 8. In order to retest these cycles, we should eliminate 5 and 8 simultaneously. But the fact that the peaks at the multiple points are smaller in sharpness than those at 5 and 8, suggest that the former are merely reflections of the latter. If that is so, the peak at 16 would vanish if we eliminate 8 only whether we eliminate 5 or not. Similarly, in retesting 10, we may eliminate 5 only. By this means we can save computational labour and use the convenient formula (7.8).

Re-test for cycle 16:

total sum of squares	:	95713.2
sum of squares due to period 8	:	47958.5
sum of squares due to period 16	:	48712.5
sum of squares due to 16 when 8 is eliminated	:	48712.5—47958.5 =754.0
residual sum of squares		95713.2—48712.5 =47000.7

$$F(16/8) = \frac{754.0}{47000.7} \times \frac{104}{8}$$

$$= 0.2086$$

(Note should be made of the fact that the error sum of squares is the same as the within column sum of squares in the Buys-Ballot Table for trial period 16. This is due to the fact that we cannot construct a model having two distinct periodicities 8 and 16, and having 24 parameters in all: we can fit at most 16 parameters and the fit is the same as if we were fitting a model involving one period of length 16 only.)

Re-test for cycle 10 :

total sum of squares	:	95713.2
sum of squares due to 10	:	16410.3
sum of squares due to 5	:	20010.7
sum of squares due to 10 when 5 is eliminated	:	3600.4
error sum of squares	:	75702.5

$$F(10/5) = \frac{3600.4}{75702.5} \times \frac{100}{5} = 1.0463$$

Both the re-tests give nonsignificant results. Similar result is obtained for 25. Hence we decide that the series contains only two periodicities, 5 and 8.

8.6. Diagram 3 is extremely interesting in that while we know that the series contains only two periodicities 8 and 12, there are in the diagram three peaks at 8, 16 and 24 ; the one at 16 is smaller than that at 8, but the one at 24 is larger than that at 8. Judging by the diagram, we would suspect that the peak at 8 is genuine and the one at 16 is a reflection of that at 8. But we cannot say that the peak at 24 is also a reflection of the one at 8, for if it were so, it would very probably be less than that at 8. We suspect that there are some genuine periodic elements at 24 and this is what the re-test establishes. Hence we conclude that the series is periodic with two periods, 8 and 24. Thus, the genuine period 12 is obscured by the period 8; but 24, being a multiple of both 8 and 12, becomes more pronounced than either of the actual ones.

8.7. The Linear Regressive series also on the whole behave according to our expectations. All but Kendall's 4, 12, 14, 16 and the artificial MA (2) (that is, moving average of order 2) series (author's 5) have diagrams lying entirely in region C. Series 12 and 14 produce peaks in the region A, and our method therefore gives wrong decisions regarding them. The others have peaks in region B. According to our method, we should first of all see which of the peaks may be considered significant, if the series is at all periodic. Then we have to compare the fit of the periodic model with the optimum linear regressive fit. This is done in Table 3. Series 16 has peaks at 7, 14 and 21, in region B, and the peak at 14 is found nonsignificant. Series 4 has two peaks in region B, at 13 and 15, the one at 15 being smaller. We subject the period 15 to a re-test. As 15 is not a multiple of 13, as suggested in the section 7 we carry out an approximate re-test by estimating the periodic elements of period 13 by the column means for period 13, subtracting them from the series, rearranging the residual series in a Buys-Ballot table for period 15, and calculating a fresh F-ratio for this table. There is, however, no need of actually obtaining the residual series. The total sum of squares for the residual series is just the residual sum of squares for period 13. $\hat{\theta}_t(13)$ (which is the estimated t-th periodic element of the periodicity 13, t being reduced mod 13, ($t=1, 2, \ldots, N$)) is arranged in a Buys-Ballot table for periodicity 15, $\theta_{i.}(15/13)$ being its i-th column mean and $\theta_{..}(15/13)$ its general mean. If

$$z_{i.} = \bar{x}_{i.}(15) - \hat{\theta}_{i.}(15/13) \text{ and } z_{..} = \bar{x}_{..} - \hat{\theta}_{..}(15/13), \qquad (i=1, 2, \ldots, 15),$$

then the between group sum of squares for the residual series is

$$\sum_{i=1}^{15} (\bar{z}_{i.} - \bar{z}_{..})^2 \, n_i(15).$$

The results are as follows :

total sum of squares	: 131212.40
between group sum of squares	: 46414.84

$$F(15/13) = \frac{46414.84}{84797.60} \times \frac{72}{14} = 2.0268.$$

This is not significant at the 0.01 point.

Hence if series 4 is at all periodic, it has one period, viz. 13. The Moving Average series (author's 5) has got only one period 11, and hence the problem of re-testing does not arise.

8.8. Table 3 carries out the comparison between the optimum periodic and the optimum linear regressive fits. Let p be the periodicity of the cyclical fit and k the order of the optimum autoregressive fit. Then, we have to compare

$$\frac{\text{within column sum of squares for period } p}{\text{total sum of squares}} \times \frac{N-1}{p-1} \qquad \ldots \quad (8.1)$$

with

$$\frac{1 - r^2_{k+1.k,k-1,\ldots,1}}{N-k}$$

where $r^2_{k+1.k,k-1,\ldots,1}$ is the multiple correlation coefficient of x_t and $x_{t-1}, x_{t-2}, \ldots x_{t-k}$. (Note that the abbreviation L. R. is used in the table to denote Linear Regressive Model.)

9. ILLUSTRATIONS : NATURAL SERIES

9.1. Rules of action having been decided upon on the basis of analysis of the artificial series, we proceeded to apply our method to a number of observed series. The final results were published in the previous paper. The most noteworthy feature in them is that our method does not seem to be biased in favour of any one of the types of processes: a fair number of series were obtained for each of them. Due to lack of space, we have not been able to provide for examination by the reader all the diagrams; only a few, representing all the different types, are presented. The most interesting feature of all these diagrams is that, except for a few (of which an example is diagram 17), the peaks of even those which have the most prominent peaks are hardly as large as those in the artificial cyclical series. The obvious conclusion is that natural series are hardly ever periodic in the same clear-cut way as the artificial series are. This is quite in line with the idea in vogue that a natural series is less likely to have either a purely continuous or a purely discontinuous spectrum than to have a mixture of the both.

9.2. Table 2 summarises in a self explanatory way the actions and decisions taken regarding the series on the basis of their F-diagrams. It is seen that the first four series are decidedly accepted as cyclical at the F-diagram stage. Of the rest, the next six have their diagrams entirely in region C, and are therefore decidedly accepted as linear regressive. The remaining twelve series have peaks in region B but not in region A. For these latter twelve series comparison has to be carried out between the Optimum Linear Regressive fit and the Optimum Linear Cyclical fit. This is done in Table 4. Those series which, on the basis of Tables 2 and 4, were thought to be Linear Regressive, were finally subjected to the test method of Rudra (1952) and the decisions as to their scheme are summarised in Table 5. A few series which should have been included in the Tables 4 and 5 are not there, as their serial correlations were not available to us.

9.3. An especially interesting case is Wolfer's Sunspots series (diagram 18). It is found by our method to be cyclical with periodicity 23. It is however well known that this data has a periodicity of length 11.5. As our method applies only to integral periodicities, we notice a peak at 23, and a minor one at 11, which latter vanishes on elimination of 23. It should be recalled that the series has lately been thought to follow an autoregressive scheme of order 2. In fact, we find that if the Sun-Spots series be subjected to the discriminatory test of Rudra (1952) we do decide on $AR(2)$, and the AR fit accounts for 81% of the variability while the periodic fit accounts for only 32%. Thus our present decision is wrong quite definitely. The reason why our method fails is that, the series, regarded as autoregresive, has a 'mean distance between upcrosses' of about 11.5. Thus, the series may also be regarded as genuinely periodic. Hence our decision, which agrees with that of Schuster, is understandable.

In conclusion I have to acknowledge my indebtedness to Dr. F. N. David of University College, London who helped me with preparation of this paper.

References

Davis, H. T. (1941): *The Analysis of Economic Time Series*, Bloomington Press, Indiana.

Kendall, M. G. (1943): Oscillatory movements in English agriculture. *J. Roy. Stat. Soc.*, **106**, 91.

——— (1946): *The Advanced Theory of Statistics*, Charles Griffin, London.

——— (1949): Tables of autoregressive series. *Biometrika*, **36**, 267.

Patnaik, P. B. (1949): The noncentral χ^2 and F-distributions and their applications. *Biometrika*, **36**, 202.

Rudra, A. (1952): Discrimination in time series analysis. *Biometrika*, **39**, 434.

——— (1954): A method of discrimination in time series analysis. *Cal. Stat. Ass. Bull.*, **18**, 59.

Schuster, A. (1906): On the periodicities of sun spots. *Phil. Trans.*, A, **206**, 69.

Tang, P. C. (1938): *Statistical Research Memoirs II*, 126. Dept. of Statistics, University of London.

Whittaker, E. T. and Robinson, G. (1940): *The Calculus of Observations*, Blackie & Sons, London.

Wold, H. (1938): *A study in the Analysis of Stationary Time Series*. Almquist & Wicksells, Uppsala.

Yule, G. U. (1927): On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sun spot numbers. *Phil. Trans.*, A, **226**, 267.

Paper received: December, 1953.

TABLE 1. SUMMARY OF THE ACTIONS AND DECISIONS TAKEN ON THE BASIS OF THE DIAGRAMS 1 TO 16

diagram no.	reference	description	length of series	periods which have peaks in region		periods subjected to retesting	result of retesting	decision
				A	B			
(1)	(2)	(3)	(4)	(5)		(6)	(7)	(8)
1	Kendall (1946)	M(10)* $\Omega(10)$ 44.44	60	10 20	none	20	not significant.	M(10)
2	author (1)	M(5,8,12) $\Omega(5)$ 132.50 $\Omega(8)$ 600.00 $\Omega(12)$ 34.02	120	5 8 16	10 25	10 16 25	not significant ,, ,,	M(5,8)
3	,, (2)	M(8,12) $\Omega(8)$ 600.00 $\Omega(12)$ 34.02	120	8 16 24	none	16 24	not significant significant	M(8,24)
4	,, (3)	M(8) $\Omega(8)$ 600.00	120	8 16 24	none	16 24	not significant ,,	M(8)
5	,, (4)	MA(1)§	100	none	none			L.R.†
6	,, (5)	MA(2)	100	none	11			M(11) or L.R.
7	,, (6)	MA(1)	100	none	none			L.R.
8	Kendall (1949) (1)	AR(2)‡	240	none	none			L.R.
9	,, (2)	AR(2)	240	none	none			L.R.
10	,, (3)	AR(2)	240	none	none			L.R.
11	,, (4)	AR(2)	100	none	13 15	15	not significant	L.R. or M(13)
12	,, (8)	AR(3)	200	none	none			L.R.
13	,, (10)	AR(3)	100	none	none			L.R.
14	,, (12)	AR(3)	100	21	none			M(21)
15	,, (14)	AR(3)	100	11	22	22	not significant	M(11)
16	,, (16)	AR(3)	100	none	7 14 21	14 21	not significant significant.	M(7,21) or L.R.

*M(*a*, *b*) : Periodic Model involving two periods, *a* and *b*.

† L.R. : Linear Regressive Model.

§ MA(*a*) : Moving Average of order *a*.

‡ AR(*b*) : Autoregressive of order (*b*).

TABLE 2. SUMMARY OF ACTIONS TAKEN AND DECISIONS REACHED ON THE BASIS OF F-DIAGRAMS FOR NATURAL SERIES NUMBERED 1 TO 22

series no.	diagram no.	reference	length	periods which have peak in region		periods subject to retesting	result of retesting	decision
				A	B			
1	17	weather Whittaker and Robinson (1940)	600	24 29				M(24,29)
2		egg Kendall (1946)	36	12				M(12)
3	18	Wolfer's sun spots Yule (1927)	176	23	11	11	not significant	M(23)
4	19	oats acres Kendall (1943)	65	24	12	24	significant	M(12,24)
5	20	cost of living index Wold (1938)	74					L.R.
6	21	Beveridge's wheat prices Kendall (1946)	370					L.R.
7†	22	cows	65					L.R.
8		horses	61					L.R.
9		potato acreage	65					L.R.
10	23	wheat yields	48					L.R.
11		oats prices	64		9			M(9) or L.R.
12		barley prices	64		13			M(13) or L.R.
13		oats yield	48		18			M(18) or L.R.
14		potato yield	48		24			M(24) or L.R.
15		barley acres	65		17			M(17) or L.R.
16		sheep	65		8 16 24	16 24	not significant not significant	M(8) or L.R.
17		wheat prices	64		9 19	19	not significant	M(9) or L.R.
18	24	wheat acres	65		9			M(9) or L.R.
19		pigs	65		13			M(15) or L.R.
20		barley yield	48		15			M(15) or L.R.
21		marriage Kendall (1946)	54		20			M(20) or L.R.
22	25	freight car loading Davis (1941)	168		12			M(12) or L.R.

† References to the series numbered 7 to 20 are the same as that of series 4.

TABLE 3. COMPARISON BETWEEN THE OPTIMUM CYCLICAL FIT AND THE OPTIMUM LINEAR REGRESSIVE FIT WITH REGARD TO THOSE ARTIFICIAL SERIES THAT HAVE PEAKS IN REGION *B* AND NOT IN REGION *A*

diagram no.	true model	optimum auto-regressive fit and the residual variance expressed as a proportion of the total variance	optimum periodic fit and the residual variance expressed as a proportion of the total variance	decision
6	MA(2) (author's 5)	AR(3) 0.57	M(11) 0.82	L.R.
11	AR(2) (Kendall's series 4)	AR(3) 0.37	M(13) 0.80	L.R.
16	AR(3) (Kendall's 16)	AR(3) 0.27	M(7,21) 0.68	L.R.

TABLE 4. COMPARISON BETWEEN THE OPTIMUM CYCLICAL FIT AND THE OPTIMUM LINEAR REGRESSIVE FIT FOR THOSE NATURAL SERIES THAT HAVE PEAKS IN REGION *B* BUT NOT IN *A*

series no.	series	proportion of residual variance to total variance of a cyclical fit	proportion of residual variance to total variance of a linear regressive fit	decision
12.	barley prices	0.68	0.61	L.R.
14.	potato yield	0.53	1.00	M(24)
15.	barley acre	0.69	1.00	M(17)
16.	sheep	0.75	0.26	L.R.
17.	wheat prices	0.78	0.55	L.R.
19.	pigs	0.73	0.65	L.R.
21.	marriage	0.62	0.39	L.R.

Series 11 and 18 are absent in the above table as it was not possible to find for them an optimum auto-regressive fit of reasonably small order. Verdict should be in favour of the cyclical type in pursuance of the principle laid down in 6.3.

The serial correlations for series 13, 20 and 22 being unavailable, optimum linear regression fit for them could not be obtained, and hence they have also been excluded from the above table.

TABLE 5. DISCRIMINATORY METHOD OF RUDRA (1952) APPLIED TO THOSE SERIES WHICH LIE ENTIRELY IN REGION *C* AND THOSE SERIES WHICH ON THE BASIS OF TABLE 4 ARE DECIDED TO BE LINEAR REGRESSIVE

series no.	series	decision
5	cost of living	AR (2)
6	Beveridge's wheat prices	MA (1)
7	cows	MA (4)
9	potato acreage	MA (2)
10	wheat yields	random
12	barley prices	MA (1)
16	sheep	AR (2)
17	wheat prices	MA (1)
19	pigs	MA (2) [or AR (2)]
21	marriages	AR (4) [or MA (4)]

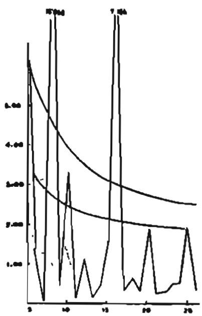

Diagram 2. Artificial cyclical series. Three periodicities : 5,8 and 12. (author)

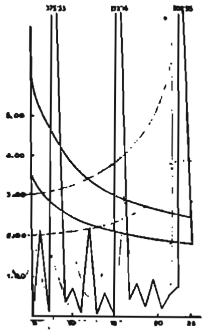

Diagram 3. Artificial cyclical series. Two periodicities : 8 and 12. (author)

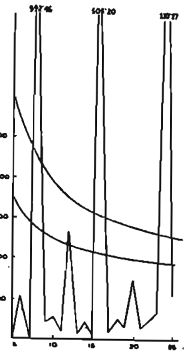

Diagram 4. Artificial cyclical series. One periodicity : 8. (author)

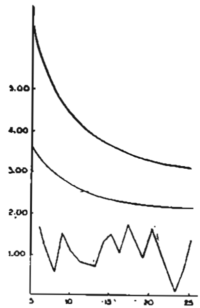

Diagram 5. Artificial moving average series. (author)

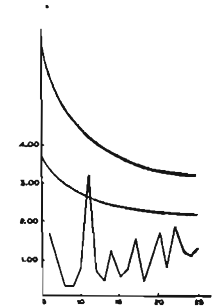

Diagram 6. Artificial moving average series. (author)

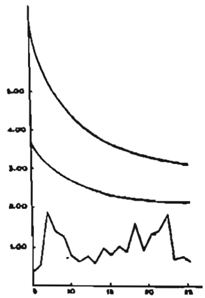

Diagram 7. Artificial moving average series. (author)

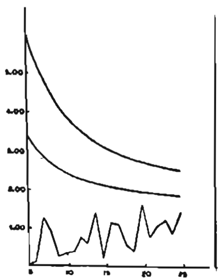

Diagram 8. Artificial autoregressive series. (Kendall's. 1).

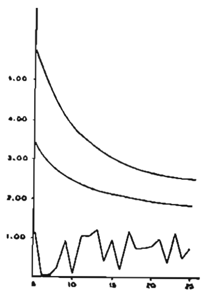

Diagram 9. Artipicial antoregressive series. (Kendall's 2).

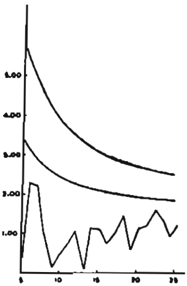

Diagram 10. Artificial autoregressivc series. (Kendall's 3).

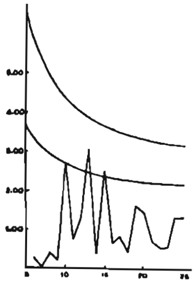

Diagram 11. Artificial autoregressive series. (Kendall's 4).

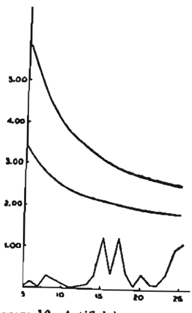

Diagram 12. Artificial autoregressiv series. (Kendall's 8).

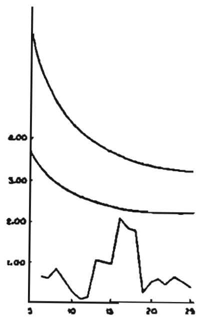

Diagram 13. Artificial autoregressive series. (Kendall's 10).

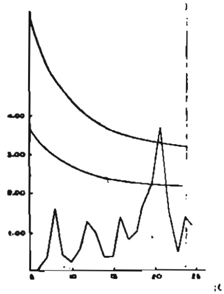

Diagram 14. Artificial autoregressive series. (Kendall's 12).

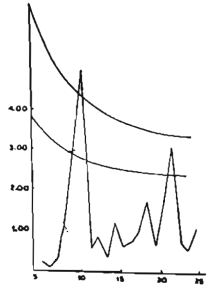

Diagram 15. Artificial autoregressive series. (Kendall's 14).

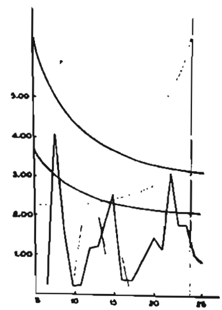

Diagram 16. Artificial autoregressive series. (Kendall's 16).

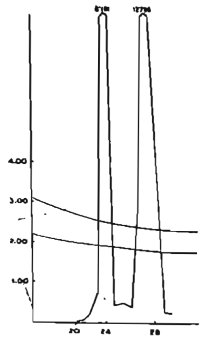

Diagram 17. Weather (Whittaker).

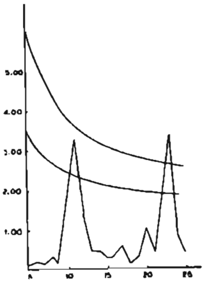

Diagram 18. Wolfer's sun spots. (Yule).

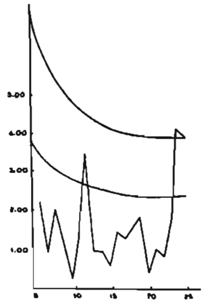

Diagram 19. Oats acres. (Kendall).

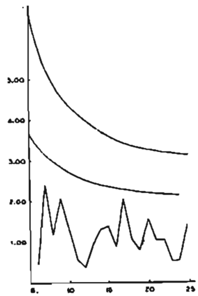

Diagram 20. Cost of Living. (Wold)

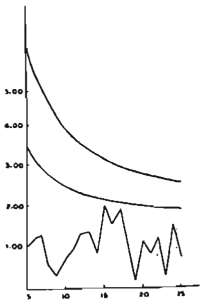

Diagram 21. Beveridge's wheat prices (Kendall)

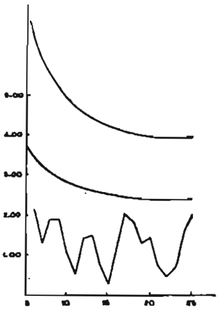

Diagram 22. Cows. (Kendall).

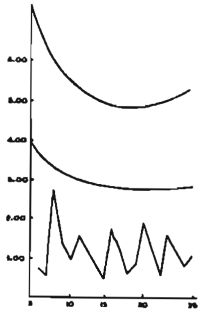

Diagram 23. Wheat yield. (Kendall).

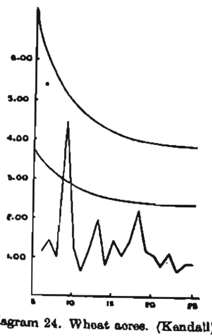

Diagram 24. Wheat acres. (Kendall).

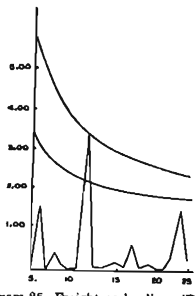

Diagram 25. Freight carloading. (Davis).