CONTRIBUTIONS TO THE STUDY OF OSCILLATORY TIME-SERIES

By

M. G. KENDALL

CAMBRIDGE
AT THE UNIVERSITY PRESS
1946

CONTRIBUTIONS TO THE STUDY OF OSCILLATORY TIME-SERIES

By

M. G. KENDALL

CAMBRIDGE
AT THE UNIVERSITY PRESS
1946

Occasional Papers

This Series is designed to serve as a medium for the publication of several types of research results—that is to say: preliminary and partial results of studies later to be published in full as books; new material to bring existing reports up to date; and small-scale independent reports complete in themselves. The Papers will appear as occasion demands, not at regular intervals; and they will vary in length. Their purpose is to make quickly available any results of research which are of particular current interest.

The Institute assumes no responsibility for the theories and opinions expressed in any publications prepared under its auspices.

I. The New Population Statistics.
By R. R. Kuczynski.

II. The Population of Bristol.
By H. A. Shannon and E. Grebenik.

III. Standards of Local Expenditure.
By J. R. Hicks and U. K. Hicks.

IV. War-time Pattern of Saving and Spending.
By Charles Madge.

V. Standardized Accountancy in Germany.
By H. W. Singer.

VI. Ten Years of Controlled Trade in South-Eastern Europe.
By N. Momtchiloff.

VII. The Problem of Valuation for Rating.
By J. R. Hicks, U. K. Hicks and C. E. V. Leser

VIII. The Incidence of Local Rates in Great Britain.
By J. R. Hicks and U. K. Hicks.

CAMBRIDGE
UNIVERSITY PRESS
LONDON : BENTLEY HOUSE
NEW YORK TORONTO BOMBAY
CALCUTTA MADRAS : MACMILLAN

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRESS, CAMBRIDGE

To

GEORGE UDNY YULE

"To borrow a striking illustration from Abraham Tucker, the substructure of our convictions is not so much to be compared to the solid foundations of an ordinary building, as to the piles of the houses of Rotterdam which rest somehow in a deep bed of soft mud."

J. A. VENN, *The Logic of Chance*

CONTENTS

PREFACE

I am greatly indebted to the National Institute for Social and Economic Research for putting at my disposal sufficient computational assistance to enable me to work out a few ideas on the analysis of oscillatory series which are presented in the following pages. It hardly needs to be stated that this is only a small contribution to a very large subject, but I hope that the conclusions which are reached will do something towards clearing the ground for further research and will arouse interest among statisticians in a wide field awaiting their closer attention.

My thanks are specially due to Miss Muriel Potter and Miss Lysbeth Holbrook of the Institute who did much of the computing.

This booklet was sent to press some months ago, and in the meantime the restrictions on disclosure of information which might be useful to the enemy have been removed. In particular, the various instruments and machines constructed during the war for correlogram analysis are no longer secret. An account of some of these is given in a symposium introduced by Dr Bartlett, Dr Cunningham, Mr Hynd and Mr Foster before the Research Section of the Royal Statistical Society on 29th January 1946. The papers and discussions on that occasion will be published in the *Supplement* to the Society's Journal and contain much that is relevant to the further study of some of the problems touched on herein.

M. G. K.

LONDON
February 1946

CHAPTER 1

INTRODUCTION

1·1. The following pages give an account of some researches, mainly experimental, into the adequacy of current methods of investigating oscillatory movements in time-series. It is not my purpose to give a connected account of those methods or to describe previous work in any detail; but in order to make the results intelligible without a lot of back-reference I give this brief introductory account. The reader who is unwilling to grapple with the statistical and mathematical analysis of Chapters 2–6 will, I hope, be able to get some general idea of the reason for the inquiry and its impact on the study of time-series by reading this chapter in conjunction with Chapter 7.

Definitions

1·2. I shall throughout consider 'stationary' series, that is to say series with no trend present and with a degree of variation which does not alter systematically with time—the fluctuation is supposed not to get systematically bigger or smaller as we go along the series. As a matter of practical convenience the mean value of the series is usually taken to be zero without loss of generality. I shall also confine attention to series which are defined at equidistant intervals of time, which may be taken as our time units. The nth term of the series will be written u_n, and our series consists of the sequence of values $u_1, u_2, u_3, \ldots$. Sometimes we require to consider past history measured from the point $t = 1$ and can then write the previous terms as $\ldots u_{-3}, u_{-2}, u_{-1}, u_0, \ldots$, etc.

1·3. By a 'peak' of the series I mean a value which is greater than the two neighbouring values; by a 'trough' a value which is less than the two neighbouring values. If a number of successive maximal values are equal they will be regarded as defining one peak located half-way between the first and the last.

1·4. By an 'upcross' I mean the point at which the series, measured about its mean, changes sign from negative to positive; and similarly, by a 'downcross' the point at which it changes sign from positive to negative. Since, in our present convention, the series is defined at equidistant intervals the upcross and downcross will lie *between* integral values of the time unit, and conventionally we shall regard the crosses as half-way between neighbouring points of opposite sign. For instance, if the series is positive at $t = 4$ and negative at $t = 5$ we shall regard the downcross as located at $t = 4·5$.

This may be a slight approximation but does not affect the accuracy of the work to any material extent.

If a series passes from positive to zero values and then to positive again without becoming negative the points will not be regarded as defining an upcross and a downcross; and similarly for values proceding from negative to zero and again to negative. But if the series proceeds from positive to zero and then to negative, the point at which it is zero will be regarded as a downcross. If there are several consecutive zeros the downcross will be located half-way between the first and the last.

1·5. We now come to a more difficult question of definition: what do we mean by a *period* in a time-series? Most economists, I think, adopt the natural but rather crude method of measuring the average distance between peaks or between troughs. Such a measure I shall call the *mean-distance* (*peaks*) or *mean-distance* (*troughs*) as the case may be. I shall interpret this definition strictly by including all peaks or troughs in the count. There is a very dangerous tendency, in counting peaks in a series (dealt with in Chapter 7), to ignore certain peaks which do not look very important. This involves a substantial element of subjective judgment as to what is 'important' and is, from the scientific viewpoint, indefensible unless it can be shown that the peaks fall into two well-marked groups, major and minor. Series for which this is so are the exception rather than the rule.

1·6. There are certain more sophisticated definitions of 'periods' in time series which are associated with the mathematical methods used to detect them. For instance, harmonic or periodogram analysis seeks to exhibit the series as a sum of certain sine or cosine terms, and the periods of these terms are known as the 'periods' of the series. To distinguish these elements I shall refer to them as *harmonic periods*.

1·7. Again, the method of correlogram analysis seeks to detect periodic elements by examining fluctuations in the correlogram. A period identified in this way I shall call the *correlogram period*. Similarly, a basic period of a solution of an autoregressive scheme will be called an *autoregressive period*.*

1·8. It is of the first importance to draw these distinctions because of the confusion which has arisen in the past (and is still arising at present) among various writers who do not define their terms clearly, or indeed at all. Possibly this has been due to the belief that the various definitions are equivalent in the sense, for example, that the mean-distance (peaks) is the same as the harmonic period of a series with a single harmonic. In general, such a belief has no foundation in fact. Whether one finds 'periods' in a

* Hitherto Yule and I have called this the *fundamental* period of the autoregressive scheme. As this may be held to be tendentious, there being nothing particular to mark out this period as more fundamental than others, I consider it preferable to adopt the neutral terminology proposed above.

series and if so what are their 'lengths' is to a great extent dependent on how one defines them.

1·9. A further point to be made in this connection is that many observed time series (one might almost say most series) do not exhibit a regular recurrence of peaks, troughs or upcrosses. The intervals between successive maxima in economic series, for example, are not equal but show very substantial variation. We cannot, then, speak of *the* period of a series, since it possesses a whole distribution of intervals between successive events of a similar kind. This is my reason for speaking of mean-distance in this connection. The harmonic periods, the correlogram periods and the autoregressive periods are unique in themselves, but again they cannot do more than summarize within the compass of single numbers the essential variation in the oscillatory behaviour of the series. In statistical language, they are measures of location of the frequency distribution of intervals.

1·10. It is for such reasons that I shall avoid altogether the word 'cycle', which carries with it a connotation of regularity of recurrence. Sir William Beveridge ([4], p. 285) differs from this view and considers that whatever the etymological origins of the word, it has now come to have a more general significance. He proposes to use the word 'period' in the sense of strict regularity of recurrence. He may be right as to the understanding of the word 'cycle' by economists but not, I think, as to the interpretation usually put upon it in the exact sciences when a temporal element is involved.* The systematic (but not necessarily cyclic) recurrence of phenomena in time series I shall call 'oscillation'. The more general stationary movement, whether systematic or not, I shall refer to as a 'fluctuation'. In this sense an oscillation is a particular case of fluctuation and a cycle is a particular case of oscillation. A periodic series is a particular case of a cyclic series, but if the occasion arises we may, in the usual mathematical terminology, describe a cyclic series as 'almost' periodic if it consists of the sum of a number of periodic terms with incommensurable periods.

Current methods of analysing oscillatory series

1·11. There are three main methods of studying oscillatory effects in time series:

(*a*) the method of counting peaks, troughs or crosses which has already been mentioned; I shall consider this method in Chapter 6;

(*b*) periodogram analysis, dealt with in Chapter 4;

* Sir William does say that he uses 'cycle' as when we speak of the 'life cycle' of a species —a regular succession of identifiable phases. It can be justified 'by showing in all the waves uniformities so important as to make a similar pattern for each of them and point to a persistent underlying cause'. I do not want to wrangle over questions of terminology, but this seems to me to beg the question. Cyclical movements in Sir William's sense can be generated by random effects without any persistent underlying cause.

(c) correlogram analysis, some aspects of which are dealt with in Chapter 3.

I shall also examine in Chapter 5 some aspects of

(d) variate-differencing, which purports to detect and estimate the magnitude of random elements superposed on a systematic movement.

1·12. The fundamental idea of periodogram analysis may be briefly explained in this way: suppose we wish to test a series to see whether it contains a period of length p. We write down the series in rows of p:

$$\begin{array}{lcccccc} & u_1 & u_2 & \cdots & \cdots & \cdots & u_p \\ & u_{p+1} & u_{p+2} & \cdots & \cdots & \cdots & u_{2p} \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & u_{(m-1)p+1} & u_{(m-1)p+2} & \cdots & \cdots & \cdots & u_{mp} \\ \hline \text{Totals} & U_1 & U_2 & \cdots & \cdots & \cdots & U_p \end{array} \tag{1·1}$$

If the number of terms N in the series is not an exact multiple of p we ignore the few that are left over after writing down as many rows as possible.

If there is present a term of period p in the series, the column totals U will cumulate the periodic effect; but if the remaining element is random, the effect of summing m rows will be to reduce the relative contribution of that element to the column totals, and similarly if there are other elements with different periods they will get out of step in successive rows and tend to cancel out in the totals. Hence, if there are enough rows, we may expect that the totals U will reveal the periodic effect and will reduce any masking effects due to random components or oscillatory components of different periods which would have prevented us from discerning it in the primary series.

1·13. The table of equation (1·1) is known as the Buys-Ballot table. Following Schuster [15], but with some changes of notation, we form the sums:

$$A(p) = \frac{2}{mp}\sum_{j=0}^{p-1}\left(U_{j+1}\cos\frac{2\pi j}{p}\right), \tag{1·2}$$

$$B(p) = \frac{2}{mp}\sum_{j=0}^{p-1}\left(U_{j+1}\sin\frac{2\pi j}{p}\right), \tag{1·3}$$

and form the *intensity*

$$I(p) = A^2 + B^2. \tag{1·4}$$

There is a value of $I(p)$ corresponding to each trial period p, and if we calculate a set of values for a certain range of trial periods we have what is known as a periodogram analysis. The graph of $I(p)$ as ordinate against p as abscissa is called the periodogram.

1·14. There are a few variants of these formulae which need not be noticed here, and it may be remarked that the method may be extended to give the intensity for rational non-integral trial periods. One useful modification of the above formula (1·4) is to divide the intensity by twice the variance of the primary series:

$$E = \frac{I}{2 \operatorname{var} u}, \tag{1·5}$$

so as to standardize it and permit of the comparison of different periodograms. The factor 2 is introduced for reasons connected with Fourier analysis.

1·15. In effect, what Schuster's method does is to find the correlation between the primary series and a harmonic term of period p. If this correlation is higher than for neighbouring trial periods there occurs a peak in the periodogram; and in particular, if the primary series contains a harmonic term of period p the periodogram should exhibit a peak of width $2p^2/(mp)$. The Schuster technique consists essentially of examining the periodogram for peaks, dismissing minor ones as side-bands of the main peaks or as sampling effects, and attributing to each remaining peak a constituent harmonic term of corresponding period in the primary series.

1·16. There have been several modifications of Schuster's method proposed, mainly with the object of obviating the rather tedious arithmetic which is involved. All those I have seen are no better than indifferent approximations, and some of them are not even that. An attempt of rather a different kind was made by Whittaker[19], who derives from the column totals U in the Buys-Ballot table a function

$$\eta^2 = \frac{\operatorname{var}(U/m)}{\operatorname{var} u}, \tag{1·6}$$

and constructs a diagram by graphing η against p, a figure which he also calls a periodogram. He shows that if the primary series contains a harmonic of period p and the remaining constituents of the series are uncorrelated with this term there is a peak in his version of the periodogram at the trial value p. But it does not follow that if there is a peak there is any periodicity in the data, as is evident from the consideration that η is independent of the order of the sums U in the Buys-Ballot table. For this reason alone, I regard the Whittaker periodogram as unsatisfactory. A practical example is given later (4·5).

Correlogram analysis

1·17. The coefficient of product-moment correlation between members of the series k intervals apart is called the serial correlation of order k. For the infinite series we have the corresponding autocorrelation

$$\rho_k = \frac{\operatorname{cov}(u_j, u_{j+k})}{\operatorname{var} u}, \tag{1·7}$$

and for the finite series

$$r_k = \frac{\sum_{j=1}^{N-k} (u_j u_{j+k})}{\left\{\sum_{j=1}^{N-k} (u_j^2) \sum_{j=1}^{N-k} (u_{j+k}^2)\right\}^{\frac{1}{2}}}, \tag{1·8}$$

where the values of u are measured about their mean. For every k we may compute a value of r_k, and the diagram obtained by graphing r_k as ordinate against k as abscissa is called the correlogram. If the primary series is a sum of harmonics the correlogram will also consist of a sum of harmonics of the same periods as the original but with different relative amplitudes. On the other hand, if the series is autoregressive the correlogram shows damped oscillations, and in particular if the series is generated by a moving average of finite extent the serial correlations vanish after a certain order, within sampling limits. The correlogram thus provides a criterion for determining the nature of the oscillations. Or at least it would do so for an infinite series; the correlogram of a finite series is more difficult to interpret.

1·18. For an infinite series everywhere defined the correlogram and the periodogram are mathematically related, one being a kind of Fourier transform of the other (cf. Davis [6], p. 112). We should thus hardly expect them to give apparently inconsistent results for practical series of any great length. It will, nevertheless, be shown later that they can do so and that the differences are very serious for certain types of series.

Variate-differences

1·19. If a series consists of a systematic component which can be represented (at least locally) by a smooth function such a polynomial, together with a random component, we may obtain an estimate of the variance of the latter from the successive differences of the series. If, in fact, such a series is differenced k times the systematic element is reduced to very small proportions and the random element has its variance increased by a factor $\binom{2k}{k}$. The variate-difference technique consists of taking the successive differences, ascertaining the mean-square of kth differences, dividing by $\binom{2k}{k}$ to give the 'reduced' variances V_k and examining the run of values of the quotient for $k = 1, 2, 3, \ldots$, etc. At some point (usually quite a low value of k) the quotients show signs of tending to a limit; and this limiting value is taken to be an estimate of the variance of the random component.

I consider the application of this method to autoregressive time series in Chapter 5.

Models of an oscillatory series

1·20. The older methods of detecting oscillatory movements were based essentially on attempts to exhibit the series as a sum of periodic terms, usually sine or cosine expansions. If the fit between theory and observation was not perfect (and as a rule, for economic series, it was not even good) the differences were treated as errors of observation. That is to say, any discrepancy was regarded as a purely temporary aberration which disappeared after occurrence *without affecting the future motion of the system*.

1·21. In economic terms this involved regarding an economic system as vibrating, either under its own internal elastic forces or under the stimulus of some external rhythm, with unchangeable period, amplitude and phase. No casual events, however severe their incidence, could alter this scheme unless they were so severe as to disrupt the fundamental structure of the system. The divergences from the strictly periodic scheme were as remote from its essential behaviour as the astronomer's errors of observation are from the stars which he contemplates. Attempts to represent economic series by schemes of this kind were being made as late as 1936, and for all I know may be going on to-day.

1·22. In the meantime it was being slowly realized that models of the periodic kind could not adequately represent observed series without great artificiality. It is, of course, always possible to represent any scheme exactly by taking enough harmonic terms, just as any function defined at a finite number of points can be represented by a polynomial if again enough terms are taken. But the hypotheses on which the periodic scheme is based are neither plausible *a priori* nor verified by observation. It is typical of economic series that they vary substantially in the intervals between successive peaks, that the amplitude of oscillations vary to much the same extent, and that there is no tendency for movements to remain in phase. Furthermore, when a disturbance occurs it is not forgotten but continues to exert some effect on the future. It may be 'random' in the sense that its occurrence is quite unpredictable and that it forms a member of a distribution of disturbances which affect the system casually from time to time. But once it has appeared *it is incorporated into the system and influences its future motion*.

1·23. This conception forms the starting point of the modern theory of oscillatory time-series, which we owe mainly to the work of Udny Yule[21, 22, 23].

Yule considers a system which is capable of damped oscillations under its own internal forces but is subjected to a stream of external shocks which continually regenerate the oscillations. The qualities of the system itself are supposed to be expressible by the fact that the value at any point of time is a function of the values at previous points, and the shocks by a disturbance function which may be a random variable. Such a system I call autoregressive. For the case of a series defined at equidistant intervals of time we may write

$$u_{t+m} = f(u_t, u_{t+1}, \dots, u_{t+m-1}) + \epsilon_{t+m}, \qquad (1\cdot 9)$$

where ϵ is the disturbance function. In particular, we have the linear autoregressive scheme

$$u_{t+m}+b_1 u_{t+m-1}+\ldots+b_m u_t=\epsilon_{t+m}, \tag{1·10}$$

which has been used by Yule himself[23] on sunspot numbers, by Walker[18] on barometric pressure, and by myself[9] on agricultural prices, acreages, crop yields and livestock numbers. Yule and I found that our series were sufficiently closely represented by the simple form

$$u_{t+2}+au_{t+1}+bu_t=\epsilon_{t+2}, \tag{1·11}$$

which I have used for the experiments described later.

1·24. It is not contended that the autoregressive scheme is the only one which will permit of the incorporation of disturbances into the system as they occur, or that it is the best scheme. It is a possible scheme. It is easily understandable in economic terms. It can be subjected to mathematical analysis. Its properties can be examined experimentally. It provides at least an approximation to the series. But perhaps better schemes will be devised as we progress in this difficult subject. For the present it seems to me to provide the best model of actual behaviour which has yet been proposed, and the work described in this brochure is intended to throw some light on its properties.

Experimental series

1·25. Owing to the uncertainty of the true nature of observed economic series I felt it desirable to work with artificial series about whose nature there could be no doubt. Four such series were constructed according to equation (1·11) for particular values of a and b. The disturbance function was in each case taken to be a random variable.*

Series 1 was constructed according to the formula

$$u_{t+2}=1·1u_{t+1}-0·5u_t+\epsilon_{t+2} \tag{1·12}$$

($a=-1·1$, $b=0·5$). It comprises 480 terms. The element ϵ was obtained by taking two-figure numbers from the *Tables of Random Sampling Numbers* by Babington Smith and myself[12]. The table on p. 9 illustrates the process.

Column (1) gives the values of the random numbers as taken from p. 2 of the tables (the number 00 being ignored so that the values range from 1 to 99). Column (2) shows the values of column (1) less 50, and thus reduces the element to one with a range from -49 to $+49$ and zero mean. Column (3) is obtained by applying equation (1·12), for instance the first term is

$$(1·1)(-35)-(0·5)(-27)+25=0,$$

* If the series is to be stationary it is necessary that b should be positive and should not exceed unity. Furthermore, a^2 must be less than $4b$ so that a cannot exceed 2 in absolute value. If a is negative the series has an autoregressive period less than 4 units, as in series 4.

and the second

$$(1\cdot1)(0)-(0\cdot5)(-35)-2 = 15\cdot5.$$

These two values 'start up' the series. The third value is then

$$(1\cdot1)(15\cdot5)-(0\cdot5)(0)+9 = 26\cdot05,$$

and so on. The final values were rounded off to the nearest unit.

(1)	(2)	(3)	(4)
23	−27		
15	−35		
75	25	0	0
48	− 2	15·5	15
59	9	26·05	26
1	−49	−28·09	−28
83	33	−10·92	−11
72	22	24·03	24
59	9	40·89	41
93	43	75·96	76
76	26	89·11	89
24	−26	34·04	34
97	47	39·89	40
8	−42	−15·14	−15
86	36	− 0·60	− 1
95	45	51·91	52
23	−27	30·40	30
3	−47	−39·51	−40
67	17	−41·66	−42
44	− 6	−32·07	−32

Series 2, this time of 240 terms, was constructed according to the formula

$$u_{t+2} = 1\cdot2u_{t+1}-0\cdot4u_t+\epsilon_{t+2}. \tag{1·13}$$

Series 3, also of 240 terms, was constructed according to the formula

$$u_{t+2} = 1\cdot1u_{t+1}-0\cdot8u_t+\epsilon_{t+2}. \tag{1·14}$$

Finally, series 4, also of 240 terms, had the formula

$$u_{t+2} = -u_{t+1}-0\cdot5u_t+\epsilon_{t+2}. \tag{1·15}$$

In each case the random function ϵ was taken from the tables of random numbers (different sets, of course, being used for different series) and the series constructed in the manner exemplified above.

1·26. The four series are shown in Tables 1·1–1·4. As they are rather too long to permit of being legibly graphed on a page of this size I have drawn the first 100 terms only of each series in Figs. 1·1–1·4. Certain constants of these series will continually be required and are brought together in Table 1·5.

The means, variances and standard deviations in Table 1·5 are those of the observed series. The autoregressive periods are the values for a series of infinite length with the constants a and b. They are calculated from the formula

$$\text{Autoregressive period} = \frac{2\pi}{\cos^{-1}\left(-\dfrac{a}{2\sqrt{b}}\right)}. \tag{1·16}$$

TABLE 1·1. Experimental series 1

No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t
1	0	81	−3	161	−68	241	−18	321	−17	401	−68
2	15	82	−65	162	−24	242	−70	322	12	402	−73
3	26	83	−102	163	−35	243	−82	323	55	403	−78
4	−28	84	−31	164	−34	244	−68	324	33	404	−55
5	−11	85	42	165	28	245	−9	325	11	405	−24
6	24	86	65	166	39	246	37	326	−31	406	−14
7	41	87	8	167	66	247	9	327	−82	407	−12
8	76	88	−3	168	72	248	31	328	−100	408	−13
9	89	89	36	169	49	249	30	329	−98	409	−36
10	34	90	16	170	50	250	−9	330	−85	410	−80
11	40	91	−38	171	77	251	−40	331	−42	411	−78
12	−15	92	−42	172	70	252	−89	332	−28	412	−66
13	−1	93	−36	173	66	253	−53	333	−44	413	6
14	52	94	−15	174	60	254	5	334	2	414	5
15	30	95	40	175	63	255	75	335	67	415	−38
16	−40	96	22	176	85	256	61	336	96	416	−24
17	−42	97	−40	177	38	257	15	337	47	417	35
18	−32	98	−93	178	−14	258	−36	338	49	418	22
19	−59	99	−121	179	−9	259	−60	339	50	419	−20
20	−45	100	−62	180	34	260	1	340	24	420	−54
21	−15	101	−32	181	18	261	−9	341	29	421	−94
22	6	102	39	182	−14	262	17	342	−11	422	−89
23	7	103	90	183	−38	263	64	343	11	423	−66
24	−35	104	73	184	10	264	101	344	53	424	−73
25	−39	105	19	185	61	265	70	345	59	425	−43
26	−2	106	27	186	28	266	8	346	55	426	−7
27	3	107	−22	187	19	267	−20	347	−3	427	53
28	−38	108	−15	188	−2	268	22	348	−13	428	100
29	−3	109	−24	189	33	269	48	349	−36	429	76
30	26	110	10	190	84	270	40	350	11	430	65
31	−1	111	46	191	111	271	32	351	80	431	46
32	−28	112	27	192	101	272	23	352	96	432	30
33	−36	113	−26	193	50	273	8	353	61	433	−16
34	−66	114	−70	194	−18	274	16	354	38	434	14
35	−8	115	−50	195	−47	275	−17	355	33	435	41
36	−4	116	0	196	−55	276	−20	356	30	436	70
37	−38	117	43	197	−7	277	−59	357	−23	437	27
38	−46	118	47	198	−23	278	−17	358	−78	438	6
39	−68	119	23	199	−63	279	26	359	−99	439	31
40	−15	120	−12	200	−83	280	13	360	−28	440	−3
41	−16	121	−63	201	−87	281	−22	361	−5	441	24
42	−57	122	−25	202	−12	282	−1	362	40	442	43
43	−55	123	13	203	4	283	18	363	61	443	−13
44	−50	124	−12	204	23	284	7	364	32	444	−4
45	−37	125	−69	205	44	285	30	365	21	445	−13
46	−23	126	−56	206	13	286	70	366	22	446	−34
47	5	127	−21	207	−51	287	64	367	58	447	−19
48	−10	128	−22	208	−106	288	80	368	36	448	30
49	−13	129	29	209	−86	289	71	369	32	449	84
50	−55	130	83	210	−8	290	34	370	35	450	122
51	−94	131	31	211	37	291	−13	371	48	451	90
52	−123	132	41	212	39	292	−75	372	3	452	71
53	−116	133	23	213	41	293	−73	373	−4	453	64
54	−105	134	19	214	9	294	−71	374	5	454	29
55	−54	135	−34	215	23	295	−38	375	−40	455	−17
56	−19	136	−57	216	24	296	−47	376	−90	456	−66
57	−36	137	−59	217	−15	297	−59	377	−85	457	−96
58	−46	138	6	218	−36	298	−57	378	−80	458	−117
59	−45	139	66	219	−51	299	−42	379	−51	459	−127
60	21	140	82	220	−78	300	1	380	−29	460	−36
61	62	141	11	221	−29	301	44	381	−44	461	22
62	57	142	−1	222	1	302	8	382	−77	462	−2
63	33	143	−18	223	51	303	8	383	−19	463	11
64	51	144	−44	224	44	304	−9	384	63	464	32
65	−5	145	−9	225	−24	305	−33	385	120	465	55
66	−64	146	6	226	−92	306	−10	386	123	466	61
67	−60	147	52	227	−87	307	52	387	104	467	55
68	−31	148	59	228	−45	308	70	388	68	468	−19
69	24	149	64	229	−5	309	56	389	72	469	−28
70	72	150	2	230	28	310	−14	390	48	470	−6
71	76	151	9	231	31	311	−64	391	28	471	2
72	−1	152	−9	232	59	312	−32	392	50	472	−33
73	5	153	−6	233	74	313	23	393	69	473	−56
74	−12	154	−6	234	30	314	41	394	98	474	−70
75	−24	155	2	235	−7	315	81	395	106	475	−50
76	17	156	−20	236	−26	316	113	396	116	476	−39
77	−3	157	−2	237	−14	317	87	397	78	477	−26
78	33	158	7	238	4	318	39	398	52	478	−23
79	85	159	12	239	25	319	−32	399	1	479	−39
80	58	160	−39	240	35	320	−65	400	−70	480	−39

TABLE 1·2. Experimental series 2

No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t
1	32	41	−28	81	−70	121	−170	161	17	201	−90
2	17	42	21	82	−56	122	−149	162	62	202	−58
3	42	43	82	83	−40	123	−96	163	24	203	−4
4	78	44	91	84	−60	124	−90	164	−36	204	−23
5	39	45	56	85	−91	125	−42	165	−60	205	−7
6	−13	46	66	86	−108	126	31	166	−83	206	33
7	−22	47	10	87	−141	127	82	167	−106	207	58
8	−51	48	9	88	−128	128	75	168	−82	208	29
9	−86	49	23	89	−120	129	39	169	−23	209	3
10	−63	50	33	90	−98	130	1	170	28	210	−18
11	−31	51	−9	91	−73	131	29	171	12	211	−57
12	−6	52	−47	92	−43	132	8	172	−15	212	−16
13	11	53	−15	93	−28	133	36	173	−38	213	28
14	60	54	3	94	−12	134	33	174	−25	214	80
15	22	55	36	95	−17	135	18	175	−26	215	110
16	11	56	48	96	−16	136	46	176	−2	216	139
17	14	57	84	97	28	137	4	177	55	217	129
18	9	58	81	98	37	138	9	178	76	218	124
19	−35	59	73	99	69	139	9	179	38	219	89
20	−65	60	7	100	116	140	41	180	60	220	71
21	−89	61	−46	101	132	141	4	181	56	221	57
22	−109	62	−52	102	63	142	−53	182	29	222	3
23	−56	63	−89	103	3	143	−33	183	26	223	−34
24	3	64	−48	104	−12	144	−63	184	−28	224	−82
25	19	65	−31	105	−55	145	−64	185	−43	225	−85
26	35	66	−65	106	−89	146	−24	186	−84	226	−105
27	28	67	−68	107	−62	147	−17	187	−72	227	−52
28	0	68	−26	108	−29	148	5	188	−97	228	25
29	−23	69	25	109	2	149	56	189	−38	229	64
30	−67	70	55	110	25	150	17	190	−28	230	53
31	−97	71	65	111	7	151	43	191	6	231	62
32	−50	72	7	112	−29	152	51	192	64	232	72
33	−4	73	2	113	−54	153	40	193	56	233	21
34	−28	74	27	114	−34	154	−19	194	−3	234	9
35	−48	75	62	115	−53	155	−35	195	1	235	−13
36	−14	76	49	116	−88	156	−49	196	−14	236	18
37	−15	77	50	117	−123	157	−62	197	7	237	26
38	−34	78	28	118	−124	158	−43	198	−35	238	54
39	−14	79	−4	119	−146	159	−43	199	−72	239	9
40	−52	80	−57	120	−171	160	−20	200	−103	240	38

TABLE 1·3. Experimental series 3

No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t	No. of term, t	Value, u_t
1	−102	41	42	81	6	121	−29	161	50	201	77
2	−117	42	−16	82	52	122	54	162	27	202	−20
3	−35	43	−51	83	80	123	123	163	15	203	−85
4	70	44	−9	84	48	124	111	164	22	204	−92
5	133	45	48	85	−3	125	−22	165	57	205	−23
6	117	46	60	86	−49	126	−97	166	14	206	66
7	20	47	22	87	−77	127	−96	167	−53	207	131
8	−89	48	−33	88	−35	128	−11	168	−82	208	111
9	−113	49	−41	89	7	129	17	169	−20	209	20
10	−76	50	−16	90	83	130	39	170	56	210	−81
11	−20	51	−7	91	133	131	−3	171	101	211	−73
12	−9	52	−4	92	114	132	−16	172	63	212	−9
13	−31	53	4	93	26	133	25	173	−35	213	98
14	15	54	−17	94	−35	134	−7	174	−110	214	79
15	36	55	−28	95	−65	135	−65	175	−127	215	41
16	−10	56	−26	96	−23	136	−111	176	−90	216	18
17	7	57	−31	97	42	137	−115	177	17	217	−4
18	16	58	−22	98	36	138	−22	178	99	218	−33
19	−34	59	21	99	−7	139	104	179	69	219	2
20	−41	60	90	100	−47	140	172	180	−32	220	−20
21	30	61	75	101	−69	141	136	181	−86	221	−18
22	−34	62	−38	102	−12	142	35	182	−74	222	9
23	26	63	−93	103	69	143	−119	183	24	223	64
24	43	64	−100	104	55	144	−110	184	128	224	111
25	58	65	−26	105	−15	145	−19	185	165	225	103
26	70	66	15	106	−24	146	63	186	61	226	57
27	12	67	1	107	15	147	93	187	−78	227	24
28	2	68	−7	108	60	148	35	188	−152	228	−18
29	−57	69	0	109	26	149	−59	189	−96	229	−42
30	−90	70	−31	110	−27	150	−61	190	47	230	−25
31	−80	71	−81	111	−5	151	−38	191	96	231	10
32	−3	72	−16	112	34	152	1	192	25	232	−9
33	12	73	46	113	38	153	80	193	−22	233	4
34	−9	74	111	114	33	154	106	194	−23	234	−6
35	−58	75	122	115	−7	155	32	195	−21	235	23
36	−101	76	50	116	8	156	−2	196	−47	236	53
37	−65	77	−13	117	29	157	−39	197	−34	237	42
38	58	78	−77	118	31	158	−61	198	−16	238	−21
39	119	79	−51	119	−8	159	−57	199	47	239	−8
40	98	80	−7	120	−58	160	33	200	106	240	55

TABLE 1·4. Experimental series 4

No. of terms, t	Value, u_t	No. of terms, t	Value, u_t	No. of terms, t	Value, u_t	No. of terms, t	Value, u_t	No. of terms, t	Value, u_t	No. of terms, t	Value, u_t
1	−20	41	38	81	−32	121	−19	161	−3	201	39
2	12	42	−51	82	35	122	13	162	39	202	−83
3	−30	43	47	83	−4	123	22	163	−85	203	46
4	1	44	−22	84	9	124	−29	164	55	204	38
5	58	45	18	85	−14	125	27	165	18	205	−75
6	−65	46	35	86	−34	126	15	166	−49	206	66
7	79	47	−54	87	−2	127	−32	167	44	207	−37
8	−33	48	66	88	−9	128	9	168	−18	208	29
9	−53	49	−84	89	46	129	23	169	25	209	16
10	42	50	41	90	−40	130	4	170	14	210	−58
11	−59	51	−16	91	58	131	3	171	5	211	23
12	8	52	−3	92	−4	132	44	172	−28	212	43
13	46	53	15	93	−18	133	−69	173	37	213	−48
14	−89	54	−25	94	62	134	71	174	−72	214	31
15	83	55	28	95	−38	135	−58	175	12	215	27
16	7	56	−36	96	28	136	47	176	51	216	−25
17	−58	57	3	97	−48	137	−52	177	−77	217	48
18	87	58	1	98	50	138	−4	178	63	218	−43
19	−97	59	−25	99	−1	139	67	179	4	219	59
20	99	60	−1	100	5	140	−92	180	−23	220	−71
21	−28	61	31	101	34	141	100	181	−22	221	83
22	−8	62	−50	102	−32	142	−61	182	63	222	−34
23	41	63	65	103	57	143	7	183	−66	223	−6
24	−27	64	−69	104	−54	144	30	184	34	224	45
25	18	65	34	105	34	145	−65	185	26	225	−27
26	23	66	−38	106	−24	146	1	186	−88	226	45
27	−45	67	6	107	−13	147	27	187	26	227	−37
28	−15	68	−1	108	17	148	15	188	−2	228	8
29	−3	69	−48	109	−41	149	−66	189	−51	229	31
30	−14	70	86	110	27	150	82	190	52	230	−13
31	−1	71	−94	111	1	151	−10	191	4	231	−36
32	−40	72	100	112	−40	152	−57	192	−47	232	90
33	−2	73	−26	113	34	153	76	193	94	233	−52
34	−27	74	−25	114	−20	154	−73	194	−42	234	20
35	16	75	36	115	−11	155	60	195	−37	235	46
36	0	76	−25	116	63	156	19	196	77	236	−74
37	15	77	−14	117	−43	157	−14	197	−34	237	98
38	16	78	47	118	−28	158	17	198	−4	238	−37
39	−26	79	−58	119	25	159	−4	199	−27	239	−42
40	−7	80	38	120	2	160	−42	200	−4	240	36

TABLE 1·5. Constants of the four experimental series

	Series			
	1	2	3	4
Number of terms	480	240	240	240
Mean	8·821	−8·179	3·729	0·317
Variance	2535·110	3414·422	3900·939	2001·200
Standard deviation	50·350	58·433	62·457	44·735
a	−1·1	−1·2	−1·1	1·0
b	0·5	0·4	0·8	0·5
Autoregressive period	9·25	19·53	6·92	3·00
m.d. (peaks)	5·05	5·57	5·52	2·69
m.d. (upcrosses)	8·30	12·39	6·38	2·76

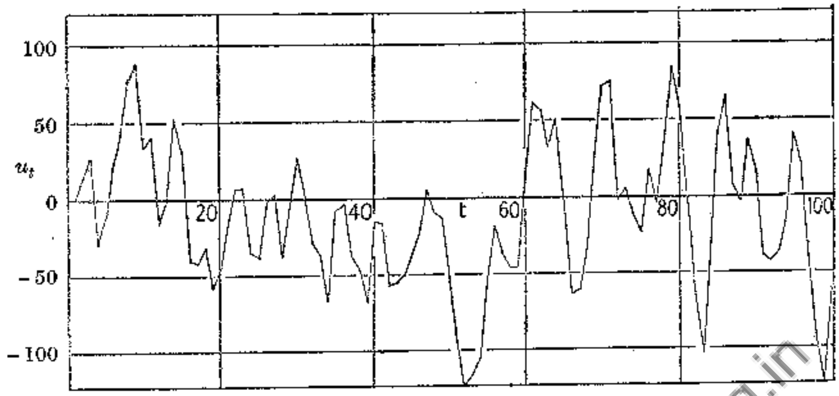

Fig. 1·1. Graph of the first 100 terms of series 1

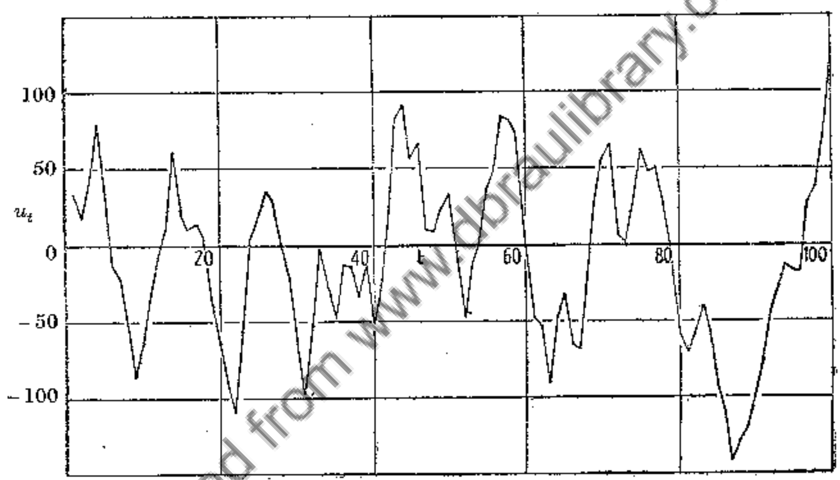

Fig. 1·2. Graph of the first 100 terms of series 2

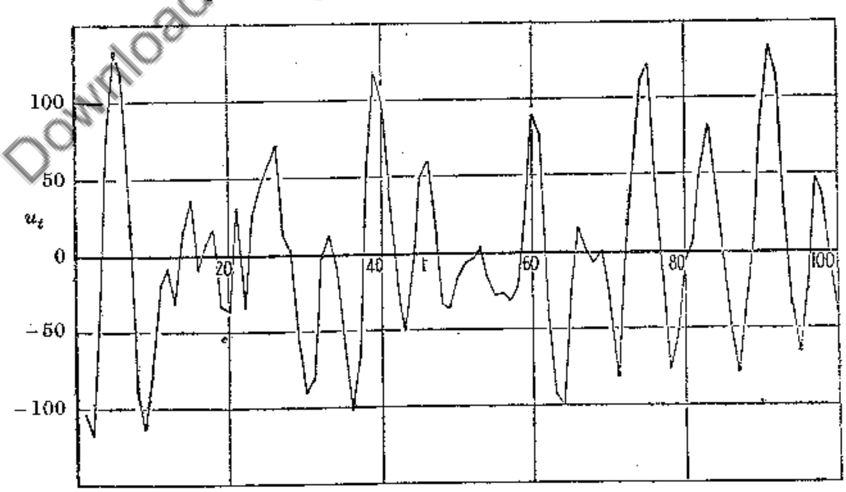

Fig. 1·3. Graph of the first 100 terms of series 3

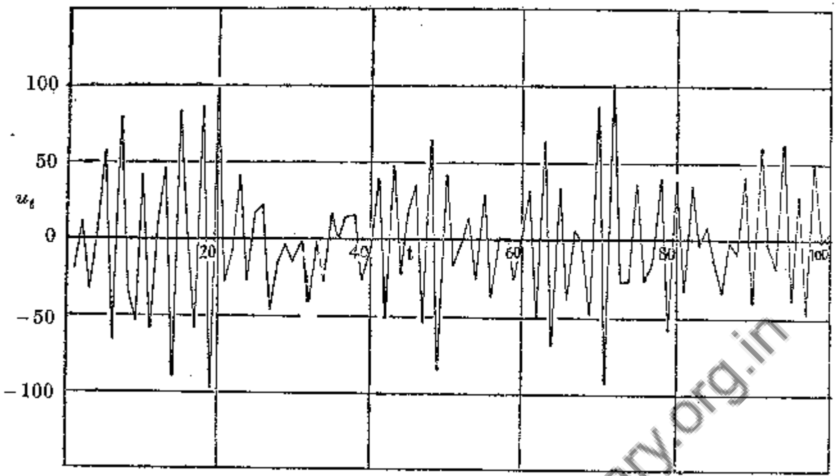

Fig. 1·4. Graph of the first 100 terms of series 4

1·27. I shall also have occasion to refer to a famous series of wheat-price index numbers compiled by Sir William Beveridge [2] and for the sake of completeness give the actual series in Table 1·6.

1·28. Chapter 2 discusses the arithmetical operations required in correlogram and variate-difference analysis. In Chapter 3 I consider the deviations of correlograms for short autoregressive series from theoretical expectation—a troublesome point in the interpretation of observed correlograms. Chapter 4 shows that for autoregressive series periodogram analysis is not only of no value but may be dangerously misleading. Chapter 5 deals with the interpretation of variate-differences for autoregressive series. In Chapter 6 I discuss the method of counting peaks, show that it is very insensitive, and suggest further possible lines of inquiry in this direction. Finally, Chapter 7 summarizes the results in non-technical language.

TABLE 1·6. Values of the Beveridge series of trend-free wheat-price index numbers. (Data from reference [**2**])

Year t	Index u_t	Year t	Index u_t	Year t	Index u_t	Year t	Index u_t	Year t	Index u_t	Year t	Index u_t
1500	106	1562	105	1624	106	1686	75	1748	109	1810	104
1501	118	1563	90	1625	121	1687	66	1749	104	1811	140
1502	124	1564	78	1626	105	1688	62	1750	90	1812	121
1503	94	1565	112	1627	84	1689	76	1751	99	1813	96
1504	82	1566	100	1628	97	1690	79	1752	95	1814	96
1505	88	1567	86	1629	109	1691	97	1753	90	1815	130
1506	87	1568	77	1630	148	1692	134	1754	80	1816	178
1507	88	1569	80	1631	114	1693	169	1755	85	1817	126
1508	88	1570	93	1632	108	1694	111	1756	117	1818	94
1509	68	1571	112	1633	97	1695	109	1757	112	1819	86
1510	98	1572	131	1634	92	1696	111	1758	95	1820	84
1511	115	1573	158	1635	97	1697	128	1759	91	1821	76
1512	135	1574	113	1636	98	1698	163	1760	88	1822	77
1513	104	1575	89	1637	105	1699	137	1761	100	1823	71
1514	96	1576	87	1638	97	1700	99	1762	97	1824	71
1515	110	1577	87	1639	93	1701	85	1763	88	1825	69
1516	107	1578	79	1640	99	1702	72	1764	95	1826	82
1517	97	1579	90	1641	99	1703	88	1765	101	1827	93
1518	75	1580	90	1642	107	1704	77	1766	106	1828	114
1519	86	1581	87	1643	106	1705	66	1767	113	1829	103
1520	111	1582	83	1644	96	1706	64	1768	108	1830	110
1521	125	1583	85	1645	82	1707	69	1769	108	1831	105
1522	78	1584	76	1646	88	1708	125	1770	131	1832	82
1523	86	1585	110	1647	116	1709	175	1771	136	1833	80
1524	102	1586	161	1648	122	1710	108	1772	119	1834	78
1525	71	1587	97	1649	134	1711	103	1773	106	1835	82
1526	81	1588	84	1650	119	1712	115	1774	105	1836	88
1527	129	1589	106	1651	136	1713	134	1775	88	1837	102
1528	130	1590	111	1652	102	1714	108	1776	84	1838	117
1529	129	1591	97	1653	72	1715	90	1777	94	1839	107
1530	125	1592	108	1654	63	1716	89	1778	87	1840	95
1531	139	1593	100	1655	76	1717	89	1779	79	1841	101
1532	97	1594	119	1656	75	1718	94	1780	87	1842	92
1533	90	1595	131	1657	77	1719	107	1781	88	1843	88
1534	76	1596	143	1658	103	1720	89	1782	94	1844	92
1535	102	1597	138	1659	104	1721	79	1783	94	1845	115
1536	100	1598	112	1660	120	1722	91	1784	92	1846	139
1537	73	1599	99	1661	167	1723	94	1785	85	1847	90
1538	86	1600	97	1662	126	1724	110	1786	84	1848	80
1539	74	1601	80	1663	108	1725	111	1787	93	1849	74
1540	74	1602	90	1664	91	1726	103	1788	108	1850	78
1541	76	1603	90	1665	85	1727	94	1789	108	1851	86
1542	80	1604	80	1666	73	1728	101	1790	86	1852	105
1543	96	1605	77	1667	74	1729	90	1791	78	1853	138
1544	112	1606	81	1668	80	1730	96	1792	87	1854	141
1545	144	1607	98	1669	74	1731	80	1793	85	1855	138
1546	80	1608	115	1670	78	1732	76	1794	103	1856	107
1547	54	1609	94	1671	83	1733	84	1795	130	1857	82
1548	69	1610	93	1672	84	1734	91	1796	95	1858	81
1549	100	1611	100	1673	106	1735	94	1797	84	1859	97
1550	103	1612	99	1674	134	1736	101	1798	87	1860	116
1551	129	1613	100	1675	122	1737	93	1799	120	1861	107
1552	100	1614	94	1676	102	1738	91	1800	139	1862	92
1553	90	1615	88	1677	107	1739	122	1801	117	1863	79
1554	100	1616	92	1678	115	1740	159	1802	105	1864	81
1555	123	1617	100	1679	113	1741	110	1803	94	1865	94
1556	156	1618	82	1680	104	1742	90	1804	125	1866	119
1557	71	1619	73	1681	92	1743	81	1805	114	1867	118
1558	71	1620	81	1682	84	1744	84	1806	98	1868	93
1559	81	1621	99	1683	86	1745	102	1807	93	1869	102
1560	84	1622	124	1684	101	1746	102	1808	94		
1561	97	1623	106	1685	74	1747	100	1809	94		

CHAPTER 2

THE CALCULATION OF SERIAL CORRELATIONS AND VARIATE-DIFFERENCES

Serial correlations

2·1. One of the most serious handicaps in the investigation of oscillations in time series is the amount of arithmetical work required. In this chapter I describe some endeavours to reduce the labour of computing serial correlations and variate-differences.

2·2. The method originally given by Yule for the calculation of serial correlations was as follows:

The series is typed in duplicate on a Burroughs recording tabulator and the slip cut up the middle, thus giving two records of the series. For the calculation of, say, the kth coefficient, one slip is pinned so that one series is k units below the other, i.e. the jth term of one is on the same level as the $(j+k)$th term of the second. (This, of course, means that k terms of one series at the beginning and k terms of the other at the end have no corresponding member. These are ignored in the calculation, which is thus based on $N-k$ pairs, N being the number in the original series.)

The computer then goes down the series, squaring the difference of each pair of terms (either in his head or by reference to a table of squares) and recording the square on the Burroughs machine. The squares are summed, checked by reading back against the series, and hence the resultant, $\Sigma(x_j - x_{j+k})^2$, gives the product sum $\Sigma(x_j x_{j+k})$ by a simple application of the identity:

$$2\Sigma(xy) = \Sigma(x^2) + \Sigma(y^2) - \Sigma(x-y)^2. \tag{2·1}$$

It is also, of course, necessary to compute $\Sigma(x)$ and $\Sigma(x^2)$ which are further required in calculating variances.

2·3. An alternative method is to calculate the sum $\Sigma(xy)$ directly by cumulating the individual products on the machine. Personally, I find that for a single operator this is more difficult than Yule's method:

(*a*) because one cannot memorize all the possible products as one can with squares and hence has to carry out multiplications,

(*b*) because one is apt to overlook the sign of the product, whereas of course the squares are all positive, and

(*c*) because checking by repetition of the operation does not reveal the source of an error if the repeat does not give the same answer as the first, whereas with the help of the record provided by the Burroughs an error can usually be located fairly easily.

2·4. In practical cases the means of the series are not exactly zero, and to calculate r_k accurately we have to use some such formula as

$$r_k = \frac{(N-k)\operatorname{cov}(x_j x_{j+k})}{\{(N-k)\operatorname{var} x_j\,(N-k)\operatorname{var} x_{j+k}\}^{\frac{1}{2}}}, \tag{2·2}$$

where
$$(N-k)\operatorname{cov}(x_j x_{j+k}) = \Sigma(x_j x_{j+k}) - \frac{1}{N-k}\Sigma(x_j)\,\Sigma(x_{j+k}), \tag{2·3}$$

$$(N-k)\operatorname{var} x_j = \Sigma(x_j^2) - \frac{1}{N-k}\Sigma^2(x_j), \tag{2·4}$$

$$(N-k)\operatorname{var} x_{j+k} = \Sigma(x_{j+k}^2) - \frac{1}{N-k}\Sigma^2(x_{j+k}), \tag{2·5}$$

the summations extending from $j = 1$ to $j = N-k$. If we choose an origin so that $\Sigma(x_j)$ is nearly zero, an approximation to r_k is given by

$$r_k = \frac{\Sigma(x_j x_{j+k})}{\{\Sigma(x_j^2)\,\Sigma(x_{j+k}^2)\}^{\frac{1}{2}}}, \tag{2·6}$$

that is to say, the terms on the extreme right in equations (2·3)–(2·5) may be neglected. This is quite a useful saving in work, and except for the first two or three serial correlations, which are usually required accurately, I find that the approximation given by (2·6) is satisfactory. In the first 60 serial correlations of the Beveridge series (Table 1·6), for example, the approximation was correct to three places of decimals in 15 cases, one unit in error in the third place in 35 cases, and two units in error in the third place in the remaining 10 cases. The effect on the correlogram was entirely negligible.

2·5. The magnitude of the error involved in the approximation is easily ascertained. We have, for the true value of r_k, writing the variates as x and y and the number of pairs entering into the correlation as n,

$$r_k = \frac{\frac{1}{n}\Sigma(xy) - \bar{x}\bar{y}}{\left[\left\{\frac{1}{n}\Sigma(x^2) - \bar{x}^2\right\}\left\{\frac{1}{n}\Sigma(y^2) - \bar{y}^2\right\}\right]^{\frac{1}{2}}}$$

$$= \frac{\Sigma(xy)}{\{\Sigma(x^2)\,\Sigma(y^2)\}^{\frac{1}{2}}}\left\{\left(1 - \frac{n\bar{x}\bar{y}}{\Sigma(xy)}\right)\left(1 + \frac{1}{2}\frac{n\bar{x}^2}{\Sigma(x^2)}\right)\left(1 + \frac{1}{2}\frac{n\bar{y}^2}{\Sigma(y^2)}\right)\right\}$$

approximately,

$$= \frac{\Sigma(xy)}{\{\Sigma(x^2)\,\Sigma(y^2)\}^{\frac{1}{2}}}\left\{1 + \frac{1}{2}\frac{n\bar{x}^2}{\Sigma(x^2)} + \frac{1}{2}\frac{n\bar{y}^2}{\Sigma(y^2)} - \frac{n\bar{x}\bar{y}}{\Sigma(xy)}\right\}. \tag{2·7}$$

The last three terms in curly brackets on the right represent the correction to be applied to the approximate value to reach the true value. Unless $\Sigma(xy)$ is small (in which case r_k itself is small) the correction itself is slight.

2·6. Apart from this approximation I have not found any substantial improvement on Yule's methods; nor have I found any simple arithmetical check on the accuracy of the working except by repeating the operations.

2·7. In order to avoid errors, however slight, I have not used the approximation of equation (2·6) in working out the serial correlations given in this brochure. This imposed a standard of arithmetical austerity which I cannot really recommend to others, but it seemed to me worth the extra labour at this stage to avoid complicating an already complicated issue.

2·8. The question then arises whether simplification might be effected by using some coefficient of correlation other than the ordinary product-moment coefficient. There are theoretical objections to such a course, but they might have to give way to arithmetical necessities in the analysis of long series. I consider in the first place an alternative used by Schumann and Hofmeyer [14] who argue as follows:

We have
$$\operatorname{var}(x-y) = \operatorname{var} x + \operatorname{var} y - 2r\sqrt{(\operatorname{var} x \operatorname{var} y)},$$
$$r = \frac{\operatorname{var} x + \operatorname{var} y - \operatorname{var}(x-y)}{2\sqrt{(\operatorname{var} x \operatorname{var} y)}}.$$

Now suppose the series long enough, and the order of the serial correlation low enough, to justify us in writing $\operatorname{var} x = \operatorname{var} y = \sigma^2$, say. Then
$$r = \frac{2\sigma^2 - \operatorname{var}(x-y)}{2\sigma^2} = 1 - \frac{\operatorname{var}(x-y)}{2\sigma^2}. \tag{2·8}$$

Suppose further that the distribution is such that the ratio of the mean deviation M_x to the standard deviation is the same for the series as for the difference $x-y$. This will, in particular, be true if the deviations are normal. Then
$$M_x = k\sigma, \quad M_{x-y} = k\sqrt{\operatorname{var}(x-y)},$$
and hence, from (2·8),
$$r = 1 - \frac{M_{x-y}^2}{2M_x^2}. \tag{2·9}$$

An approximate value of r can then be found from the mean deviations of the series.

2·9. I tested this method on the first few correlations of the Beveridge series of trend-free wheat-price index-numbers (Table 1·6). The results were as follows:

Order of correlation	r_k	Estimate of r_k from (2·9)
1	0·562	0·609
2	0·103	0·168
3	−0·075	−0·034
4	−0·092	−0·052

The agreement is far from satisfactory. In fact, for many purposes it would be considered definitely unsatisfactory. The cause does not lie wholly in the failure of the Beveridge series to attain normality, as is shown in Table 2·1, giving the distribution of values of the series itself and the first differences. Here the ratio of mean deviation to standard deviation for the two cases was 0·772 and 0·725, against a theoretical value of 0·798. The differences are not large, and it looks as if the Beveridge series is as near normality as one has any right to expect, at least in economic series.

TABLE 2·1. Distribution of the values (*a*) of the Beveridge wheat-price index (Table 1·6) and (*b*) of the first differences of the index increased by 100

Values of index (or first difference plus 100)	(*a*) Frequency of values of index	(*b*) Frequency of values of first difference + 100	Values of index (or first difference plus 100)	(*a*) Frequency of values of index	(*b*) Frequency of values of first difference + 100
40–	—	1	120–	10	10
45–	—	1	125–	10	10
50–	1	4	130–	11	8
55–	—	—	135–	9	2
60–	3	4	140–	4	2
65–	6	13	145–	1	4
70–	16	8	150–	—	1
75–	26	10	155–	3	1
80–	37	24	160–	2	2
85–	38	30	165–	2	1
90–	48	39	170–	—	—
95–	35	43	175–	2	—
100–	35	56	180–	—	—
105–	34	41	185–	—	1
110–	21	34			
115–	16	19	Totals	370	369

2·10. I conclude that the approximation represented by equation (2·9) is unsafe. For short series it would evidently be liable to give even less satisfactory results than those above. Schumann and Hofmeyer, I ought to say, were working with a meteorological series of 3652 terms and only proceeded as far as r_4, being interested in the decay of serial correlation rather than in oscillatory movements. One must sympathize with their wishes to cut down the arithmetic as far as possible on a series of such a length, which is far longer than anything in economics except where such topics as daily stock quotations are being considered.

But I am not at all sure that the proposal to use mean-deviations does result in a very appreciable saving in labour. Unless the series consists of terms which run into four significant figures it is almost as easy, with a little practice, to add squares of deviations as the absolute deviations themselves, and easier to work out correlations if the series is so short that a new mean has to be taken for each correlation.

2·11. Another possibility which seemed worth examination was the replacement of product-moment coefficients by ranking coefficients.

However, for series of the length encountered in practice, say greater than 30, the calculation of pure ranking correlations entails greater labour than the method of product-moments itself, for the series has to be renumbered for each serial coefficient. Theoretical considerations apart, there seems to be loss of time and efficiency in the substitution of ranking methods for ordinary product-moments.

2·12. Finally, it occurred to me that perhaps a rough indication of serial correlations might be obtained by the consideration of signs alone. Suppose that in a series of n pairs of terms with a mean in the neighbourhood of zero p have the same sign and $n-p$ opposite signs. (If one of a pair is zero we may allot $\frac{1}{2}$ to the number of both positive and negative signs.) A rough indication of the extent of serial correlation is given by the ratio p/n, or preferably the ratio

$$w = \frac{2p}{n} - 1, \tag{2·10}$$

which can vary from -1 to $+1$ like an ordinary coefficient of association or correlation.

TABLE 2·2. Serial correlations and coefficients w of equation (2·10) for two series of 65 and 64 terms

Order of correlation	Series A		Series B	
	r	w	r	w
1	0·27	0·25	0·58	0·46
2	−0·18	−0·14	0·02	0·03
3	−0·18	−0·16	−0·27	−0·21
4	−0·27	−0·21	−0·40	−0·23
5	−0·37	−0·27	−0·39	−0·22
6	−0·07	−0·01	−0·34	−0·14
7	0·11	0·14	−0·17	−0·02
8	0·24	0·02	0·17	0·11
9	0·27	0·04	0·41	0·31
10	0·07	−0·09	0·37	0·33
11	−0·13	−0·11	0·17	0·25
12	−0·11	−0·02	−0·07	0·04
13	−0·23	−0·04	−0·30	−0·18
14	−0·24	−0·06	−0·33	−0·24
15	−0·06	0·12	−0·32	−0·31
16	0·13	0·10	−0·28	−0·21
17	0·16	0·04	−0·05	0·06
18	0·16	−0·06	0·20	0·09
19	0·13	−0·22	0·36	0·02
20	0·09	−0·02	0·34	0·00
21	0·04	0·23	0·11	0·12
22	−0·18	0·16	−0·14	−0·14
23	−0·28	0·05	−0·20	−0·07
24	−0·04	−0·02	−0·20	−0·10
25	−0·07	−0·20	−0·22	−0·03
26	−0·14	−0·18	−0·05	0·11
27	0·23	0·21	0·22	0·14
28	0·38	0·24	0·40	−0·03
29	0·17	0·06	0·36	−0·17
30	0·05	−0·09	0·02	−0·15

Table 2·2 gives a comparison of the values of w and r for two series of 65 and 64 terms. (The two series were those of wheat acreage and wheat prices in England and Wales, the original data being given in my paper of 1944[9].) The approximation provided by w is obviously very rough and one cannot say more in its favour than that it might be useful as a preliminary test to see whether correlogram analysis was likely to yield useful results.

Mechanical methods

2·13. When correlogram analysis is extensive enough to justify the construction of special machines or the modification of existing machines for special purposes it will be quite possible to calculate serial correlations to a sufficient degree of accuracy without going through the arithmetical labour dealt with in the foregoing paragraphs. Three possibilities may be mentioned:

(*a*) By suitable arrangements of punched cards the serial covariances may be worked out exactly. I do not know of any correlogram analysis which has been carried out by punched cards and mechanical analysis, possibly because of the fact that individual workers find it quicker and less expensive to calculate serial correlations direct; but the technique would no doubt be useful in extensive investigations requiring a high degree of accuracy.

(*b*) I understand that during the war machines have been devised from telegraphic tape-machines for the purpose of calculating serial covariances.

(*c*) An interesting optical device described by Martindale[13] gives the serial covariances quickly and easily, and if it can be brought to the requisite pitch of accuracy will go far to overcome the difficulty of correlogram analysis. I understand that consideration is being given in several quarters to the construction of improved machines on these lines. If satisfactory machines are built and made available for general use it will be possible to make a great deal of progress with some important questions which still remain unsolved.*

Variate-differences

2·14. The only method I know of carrying out a variate-difference analysis directly is to write the series down, together with its differences, up to the required order and then to sum the squares of appropriate terms.

* I have made a few inquiries to see whether existing machines which integrate the product of given functions are of any use for serial correlation analysis. The results were rather disappointing. Owing to the length of time-series and the angular appearance of graphs which represent them there seem to be serious practical difficulties in applying such instruments as the Bush differential analyser. The optical methods described by Burger and van Cittert[5] seem to me to present similar difficulties.

This is almost intolerably tedious for long series, besides throwing a heavy strain on the computer's accuracy.

2·15. Variate-differences can, however, be computed fairly easily from the product-sums which are reached in serial correlation analysis. If we have a series of values $x_1 \dots x_n$, the first differences are $x_1 - x_2$, etc., the second differences $x_1 - 2x_2 + x_3$, etc., and so on. Let S_j be the sum of the squares of the jth differences and write for the product-sums

$$P_j = \sum_{k=1}^{n-j} (x_k x_{k+j}). \tag{2·11}$$

The sums S are those appearing in variate-difference analysis, and the quantities P appear in serial correlation analysis. Either set can be expressed in terms of members of the other, as follows:

$$S_0 = P_0, \tag{2·12}$$

$$S_1 = \sum (x_j - x_{j+1})^2 = \sum_{j=1}^{n-1} x_j^2 - 2\sum_{j=1}^{n-1} x_j x_{j+1} + \sum_{j=2}^{n} x_j^2 = 2P_0 - x_1^2 - x_n^2 - 2P_1, \tag{2·13}$$

$$\begin{aligned} S_2 &= \sum (x_j - 2x_{j+1} + x_{j+2})^2 \\ &= \sum_{j=1}^{n-2} x_j^2 + 4\sum_{j=2}^{n-1} x_j^2 + \sum_{j=3}^{n} x_j^2 - 4\sum_{j=1}^{n-2} x_j x_{j+1} + 2\sum_{j=1}^{n-2} x_j x_{j+2} - 4\sum_{j=2}^{n-1} x_j x_{j+1} \\ &= 6P_0 - 8P_1 + 2P_2 - x_1^2 - x_n^2 - (2x_1 - x_2)^2 - (2x_n - x_{n-1})^2, \end{aligned} \tag{2·14}$$

and so on. For the purpose of expressing the general formulae of this kind it is convenient to modify the sums S. Suppose we write the series $x_1 \dots x_n$ preceded and followed by a number of zeros. The difference table will then appear as follows:

0				
	0			
0		0		
	0		$-x_1$	
0		$+x_1$		
	$-x_1$		$3x_1 - x_2$	etc.
x_1		$-2x_1 + x_2$		
	$x_1 - x_2$		$-3x_1 + 3x_2 - x_3$	
x_2		$x_1 - 2x_2 + x_3$		
	$x_2 - x_3$		$x_1 - 3x_2 + 3x_3 - x_4$	
x_3		$x_2 - 2x_3 + x_4$		

with a symmetrical effect at the other end. Writing now T_1, T_2, etc., for the sum of the squares of members in the first, second, etc., column of differences we see that

$$T_j = \sum_k \left\{x_k - \binom{j}{1} x_{k+1} + \binom{j}{2} x_{k+2} \dots\right\}^2, \tag{2·15}$$

where the summation now takes place over all values of x and there are no complications introduced by end effects. In fact, we have thrown the end effects into the sums T which replace the S's. In actually calculating the T's from the S's it is very little trouble to add the extra terms to the tables giving the latter; and when calculating the S's from the T's only the differences at the end of the table need be worked out.

We have then from (2·15), on expansion

$$T_j = P_0\left\{\binom{j}{0}^2 + \binom{j}{1}^2 + \dots\right\} + 2(-1)^k \sum_k \left[P_k\left\{\binom{j}{0}\binom{j}{k} + \dots + \binom{j}{j-k}\binom{j}{j}\right\}\right]. \tag{2·16}$$

The coefficients of the various P's are easily seen to be equal to corresponding powers of t in

$$\left\{\binom{j}{0} t^0 - \binom{j}{1} t + \binom{j}{2} t^2 \dots\right\}\left\{\binom{j}{0} t^j - \binom{j}{1} t^{j-1} + \dots\right\},$$

i.e. in $(-1)^j (1-t)^{2j}$, and we find, on substitution in (2·16),

$$T_j = P_0\binom{2j}{j} - 2P_1\binom{2j}{j-1} + 2P_2\binom{2j}{j-2} + \dots + 2(-1)^j P_j. \tag{2·17}$$

For example,

$$\left.\begin{aligned}
T_0 &= P_0,\\
T_1 &= 2P_0 - 2P_1,\\
T_2 &= 6P_0 - 8P_1 + 2P_2,\\
T_3 &= 20P_0 - 30P_1 + 12P_2 - 2P_3,\\
T_4 &= 70P_0 - 112P_1 + 56P_2 - 16P_3 + 2P_4,\\
T_5 &= 252P_0 - 420P_1 + 240P_2 - 90P_3 + 20P_4 - 2P_5,\\
T_6 &= 924P_0 - 1584P_1 + 990P_2 - 440P_3 + 132P_4 - 24P_5 + 2P_6,\\
T_7 &= 3432P_0 - 6006P_1 + 4004P_2 - 2002P_3 + 728P_4 - 182P_5\\
&\qquad + 28P_6 - 2P_7,\\
T_8 &= 12870P_0 - 22880P_1 + 16016P_2 - 8736P_3 + 3640P_4 - 1120P_5\\
&\qquad + 240P_6 - 32P_7 + 2P_8,\\
T_9 &= 48620P_0 - 87516P_1 + 63648P_2 - 37128P_3 + 17136P_4 - 6120P_5\\
&\qquad + 1632P_6 - 306P_7 + 36P_8 - 2P_9,\\
T_{10} &= 184756P_0 - 335920P_1 + 251940P_2 - 155040P_3 + 77520P_4\\
&\qquad - 31008P_5 + 9690P_6 - 2280P_7 + 380P_8 - 40P_9 + 2P_{10}.
\end{aligned}\right\} \tag{2·18}$$

The coefficients check in virtue of the fact that they sum to zero.

Conversely we have

$$
\left.\begin{aligned}
2P_0 &= 2T_0,\\
2P_1 &= -T_1+2T_0,\\
2P_2 &= T_2-4T_1+2T_0,\\
2P_3 &= -T_3+6T_2-9T_1+2T_0,\\
2P_4 &= T_4-8T_3+20T_2-16T_1+2T_0,\\
2P_5 &= -T_5+10T_4-35T_3+50T_2-25T_1+2T_0,\\
2P_6 &= T_6-12T_5+54T_4-112T_3+105T_2-36T_1+2T_0,\\
2P_7 &= -T_7+14T_6-77T_5+210T_4-294T_3+196T_2-49T_1+2T_0,\\
2P_8 &= T_8-16T_7+104T_6-352T_5+660T_4-672T_3+336T_2\\
&\qquad -64T_1+2T_0,\\
2P_9 &= -T_9+18T_8-135T_7+546T_6-1287T_5+1782T_4-1386T_3\\
&\qquad +540T_2-81T_1+2T_0,\\
2P_{10} &= T_{10}-20T_9+170T_8-800T_7+2275T_6-4004T_5+4290T_4\\
&\qquad -2640T_3+825T_2-100T_1+2T_0.
\end{aligned}\right\} \quad (2\cdot19)
$$

The coefficients sum in turn to 2, 1, −1, −2, −1, 1, 2, etc. The coefficient of T_k in $2P_j$ is

$$(-1)^{j-k}\frac{2j}{j+k}\binom{j+k}{2k}. \quad (2\cdot20)$$

2·16. For the purposes of the present study we shall require equations of type (2·18) up to the 20th order. The coefficients increase rapidly in magnitude above T_{10}, and, since we eventually have to divide by the coefficient of P_0, namely $\binom{2k}{k}$, we might as well carry out the division at once. Table 2·3 gives the necessary coefficients, after division, up to the 20th order. For instance, if we know P_0 to P_6 from the serial correlation analysis we shall have

$$
\begin{aligned}
\frac{T_6}{\binom{12}{6}} = P_0 - 1{\cdot}714{,}285{,}714P_1 + 1{\cdot}071{,}428{,}571P_2 - 0{\cdot}476{,}190{,}476P_3\\
+ 0{\cdot}142{,}857{,}143P_4 - 0{\cdot}025{,}974{,}026P_5 + 0{\cdot}002{,}164{,}502P_6,
\end{aligned}
$$

the coefficients being given in Table 2·3, column headed 6.

Summary

2·17. (1) For a computer working only with the usual machines no essential improvement was found on the method of calculating serial correlations proposed by Yule.

(2) Provided that the mean of the observed series is near to zero an adequate approximation to serial correlations may be obtained by using sums of squares and products instead of variances and covariances in the formula for product-moment correlations.

TABLE 2·3. Coefficients for converting the sums P into the sums T (see **2·16**)

	1	2	3	4	5
0	1·	1·	1·	1·	1·
1	1·	1·333,333,333	1·5	1·6	1·666,666,667
2	—	0·333,333,333	0·6	0·8	0·952,380,952
3	—	—	0·1	0·228,571,429	0·357,142,857
4	—	—	—	0·028,571,429	0·079,365,079
5	—	—	—	—	0·007,936,508

	6	7	8	9	10
0	1·	1·	1·	1·	1·
1	1·714,285,714	1·75	1·777,777,778	1·8	1·818,181,818
2	1·071,428,571	1·166,666,667	1·244,444,444	1·309,090,909	1·363,636,364
3	0·476,190,476	0·583,333,333	0·678,787,879	0·763,636,364	0·839,160,839
4	0·142,857,143	0·212,121,212	0·282,828,283	0·352,447,552	0·419,580,420
5	0·025,974,026	0·053,030,303	0·087,024,087	0·125,874,126	0·167,832,168
6	0·002,164,502	0·008,158,508	0·018,648,019	0·033,566,434	0·052,447,552
7	—	0·000,582,751	0·002,486,402	0·006,293,706	0·012,340,601
8	—	—	0·000,155,400	0·000,740,436	0·002,056,767
9	—	—	—	0·000,041,135	0·000,216,502
10	—	—	—	—	0·000,010,825

	11	12	13	14	15
0	1·	1·	1·	1·	1·
1	1·833,333,333	1·846,153,846	1·857,142,857	1·866,666,667	1·875
2	1·410,256,410	1·450,549,450	1·485,714,286	1·516,666,667	1·544,117,647
3	0·906,593,407	0·967,032,967	1·021,428,571	1·070,588,236	1·115,196,078
4	0·483,516,484	0·543,956,044	0·600,840,336	0·654,248,366	0·704,334,365
5	0·211,538,462	0·255,979,315	0·300,420,168	0·344,341,245	0·387,383,901
6	0·074,660,634	0·099,547,511	0·126,492,702	0·154,953,560	0·184,468,524
7	0·020,739,065	0·031,436,056	0·044,272,446	0·059,029,928	0·075,464,396
8	0·004,366,119	0·007,859,014	0·012,649,270	0·018,782,250	0·026,248,486
9	0·000,654,918	0·001,496,955	0·002,874,834	0·004,899,717	0·007,655,808
10	0·000,062,373	0·000,204,130	0·000,499,971	0·001,020,774	0·001,837,394
11	0·000,002,835	0·000,017,750	0·000,062,496	0·000,163,324	0·000,353,345
12	—	0·000,000,740	0·000,005,000	0·000,018,845	0·000,052,347
13	—	—	0·000,000,192	0·000,001,396	0·000,005,609
14	—	—	—	0·000,000,050	0·000,000,387
15	—	—	—	—	0·000,000,013

	16	17	18	19	20
0	1·	1·	1·	1·	1·
1	1·882,352,941	1·888,888,889	1·894,736,842	1·9	1·904,761,905
2	1·568,627,451	1·590,643,275	1·610,526,316	1·628,571,429	1·645,021,645
3	1·155,830,753	1·192,982,456	1·227,067,669	1·258,441,558	1·287,408,244
4	0·751,289,989	0·795,321,637	0·836,637,047	0·875,437,606	0·911,914,173
5	0·429,308,565	0·469,962,786	0·509,257,333	0·547,148,504	0·583,625,071
6	0·214,654,283	0·245,197,975	0·275,847,722	0·306,403,162	0·336,706,772
7	0·093,327,949	0·112,382,405	0·132,406,907	0·153,201,581	0·174,588,696
8	0·034,997,981	0·044,952,962	0·056,018,307	0·068,089,592	0·081,059,038
9	0·011,199,354	0·015,560,641	0·020,747,521	0·026,749,483	0·033,541,671
10	0·003,015,211	0·004,610,560	0·006,668,846	0·009,223,960	0·012,298,613
11	0·000,670,047	0·001,152,640	0·001,839,682	0·002,767,188	0·003,967,294
12	0·000,119,651	0·000,238,477	0·000,429,259	0·000,714,113	0·001,115,802
13	0·000,016,504	0·000,039,746	0·000,083,082	0·000,156,212	0·000,270,497
14	0·000,001,650	0·000,005,129	0·000,012,982	0·000,028,402	0·000,055,691
15	0·000,000,106	0·000,000,481	0,000,001,574	0·000,004,177	0·000,009,547
16	0·000,000,003	0·000,000,029	0·000,000,139	0·000,000,477	0·000,001,326
17	—	0·000,000,001	0·000,000,008	0·000,000,040	0·000,000,143
18	—	—	0·000,000,000	0·000,000,002	0·000,000,011
19	—	—	—	0·000,000,000	0·000,000,001
20	—	—	—	—	0·000,000,000

(3) Coefficients alternative to the product-moment correlation based on mean-deviations, rankings or distributions of signs do not provide an adequate substitute.

(4) The adaptation of existing machines or construction of special machines for the calculation of serial covariances is quite possible and may materially lighten the labour involved in direct computation.

(5) The arithmetic of variate-difference analysis is related to that of serial correlation analysis and formulae are given for calculating the basic quantities required for one from those calculated in the other.

CHAPTER 3

CORRELOGRAM ANALYSIS OF SHORT AUTOREGRESSIVE SERIES

3·1. For the series defined by

$$u_{t+2} + au_{t+1} + bu_t = \epsilon_{t+2}, \tag{3·1}$$

where ϵ is a random element, the theoretical values of the serial correlations (that is to say, the values for a series of infinite length) are given by

$$\rho_k = \frac{p^k \sin(k\theta + \psi)}{\sin \psi}, \tag{3·2}$$

where

$$p = |\sqrt{b}|, \tag{3·3}$$

$$\theta = \cos^{-1}\frac{-a}{2\sqrt{b}}, \tag{3·4}$$

$$\tan\psi = \frac{1+b}{1-b}\tan\theta.$$

For the methods of proof of this result see Kendall [10]. It is due essentially to Yule. In conformity with the usual statistical practice we write a Greek ρ for the population value of the quantity which in samples is represented by the Roman r.*

3·2. It follows from (3·2) that the correlogram of an infinite series of the simple autoregressive type is a damped harmonic. The factor ψ is a constant; the factor $\sin(k\theta + \psi)$ introduces an oscillation into the correlogram with harmonic period $2\pi/\theta$; and the factor p^k damps out the oscillation with greater or less rapidity according to the value of the constant b.

* It is convenient to refer to the population values as autocorrelations and the observed values as serial correlations, the Greek and Latin derivations of 'auto' and 'serial' preserving the same distinction as the symbols ρ and r; but I fear my own practice has not been very consistent in this respect.

For our experimental series the damping factor is considerable. In series 1, for example, $p = \sqrt{0 \cdot 5}$, and the twentieth correlation is therefore damped according to the factor $(\frac{1}{2})^{10} = 0 \cdot 001$ approximately, so that ρ_{20} is of this order of magnitude. In series 2 ($b = 0 \cdot 4$) the damping is even heavier, and even in series 3 ($b = 0 \cdot 8$) it is considerable, ρ_{20} being of the order of 0·1.

3·3. It has been found that for series which, at first sight, appeared as if they might be of the autoregressive type, the damping effect was very much less marked than theoretical considerations led one to expect (cf. Kendall [10]). I proceed to show by analysis of the experimental series that in autoregressive series of finite length the failure to damp is, in fact, a normal phenomenon. It will follow that if, in practical cases, a correlogram oscillates in a manner suggestive of autoregression but does not damp out as would the correlogram of an infinite series, we need not on that account reject the hypothesis of autoregression.

3·4. Series 1 was divided into 8 sets of 60 terms each, which will be denoted by 1*a*, 1*b*, ..., 1*h*. For each of these subseries the first 50 serial correlations were computed. The series were then joined in pairs, 1*a* to 1*b*, 1*c* to 1*d*, 1*e* to 1*f* and 1*g* to 1*h*, so as to form 4 series of 120 terms each and again the first 50 serial correlations computed. The resulting series were joined in pairs giving two series of 240 terms each, and the first 50 serial correlations computed. Finally, the two series were joined so as to form the single series of 480 terms and again the first 50 serial correlations computed. The results are given in Tables 3·1 and 3·2, and the correlograms in Figs. 3·1 and 3·2. The theoretical correlogram is also shown.

3·5. These figures bring out some remarkable results. For series of 60 terms the observed correlograms differ very markedly from the theoretical value in some cases; and in no case do they damp in the theoretical manner. For series of 120, 240 and 480 terms the differences are not so great, and they diminish as the series becomes longer, as we have every right to expect. But even for a series as long as 480 terms the failure to damp according to expectation is clear.

3·6. The experiment was repeated with each of the other three series, the only difference being that in these cases there were four series of 60 terms each denoted by *a*, *b*, *c* and *d* (the total length being 240 terms), and that only the first 30 serial correlations were computed. The results are given in Tables 3·3–3·5 and Figs. 3·3–3·5. Again we have the greater damping for the longer series and the failure to damp according to expectation for the longest.

3·7. It is thus clear that on occasion the correlogram of a short autoregressive series fails to conform to expectation; and that it is the rule rather than the exception for it to fail to damp out in the manner of the theoretical correlogram. This brings out very well a point which is often inadequately

TABLE 3·1. Serial correlations of series 1*a* to 1*h*, the eight subseries of 60 terms each composing series 1

Order of correlation	1*a*	1*b*	1*c*	1*d*	1*e*	1*f*	1*g*	1*h*
1	0·776	0·635	0·727	0·725	0·737	0·760	0·885	0·827
2	0·543	0·016	0·391	0·219	0·317	0·335	0·613	0·534
3	0·446	−0·439	0·183	−0·180	−0·026	−0·032	0·429	0·254
4	0·390	−0·570	0·061	−0·304	−0·166	−0·193	0·293	−0·018
5	0·376	−0·334	−0·019	−0·192	−0·153	−0·171	0·145	−0·232
6	0·366	0·002	0·042	−0·042	−0·095	−0·084	0·026	−0·386
7	0·334	0·229	0·115	0·040	0·014	0·026	−0·038	−0·472
8	0·342	0·348	0·193	0·052	0·090	0·115	−0·048	−0·449
9	0·316	0·299	0·209	−0·000	0·076	0·110	−0·051	−0·336
10	0·207	0·060	0·105	−0·031	−0·108	−0·033	−0·088	−0·148
11	0·250	−0·300	−0·141	−0·023	−0·374	−0·233	−0·175	0·037
12	0·342	−0·543	−0·347	−0·017	−0·540	−0·325	−0·281	0·141
13	0·361	−0·421	−0·418	−0·098	−0·425	−0·299	−0·416	0·188
14	0·392	0·025	−0·369	−0·278	−0·019	−0·174	−0·509	0·223
15	0·415	0·498	−0·181	−0·412	0·295	−0·034	−0·524	0·215
16	0·336	0·652	0·021	−0·338	0·374	0·030	−0·495	0·219
17	0·297	0·441	0·192	−0·086	0·300	−0·001	−0·423	0·147
18	0·333	0·010	0·232	0·185	0·186	−0·061	−0·318	0·069
19	0·354	−0·403	0·216	0·332	0·062	−0·066	−0·213	0·051
20	0·386	−0·564	0·099	0·301	−0·109	−0·031	−0·099	0·039
21	0·298	−0·419	−0·080	0·149	−0·181	−0·046	0·018	0·094
22	0·190	−0·168	−0·118	−0·094	−0·086	−0·102	0·105	0·117
23	0·099	0·083	−0·056	−0·208	0·101	−0·240	0·139	0·014
24	0·104	0·331	0·037	−0·108	0·210	−0·347	0·149	−0·079
25	0·097	0·379	0·209	0·093	0·143	−0·319	0·157	−0·167
26	0·133	0·212	0·218	0·226	−0·117	−0·164	0·180	−0·341
27	0·129	−0·112	0·238	0·164	−0·403	0·100	0·274	−0·518
28	−0·018	−0·462	0·210	0·009	−0·528	0·373	0·329	−0·601
29	−0·033	−0·537	0·086	−0·127	−0·509	0·512	0·305	−0·597
30	−0·039	−0·198	−0·149	−0·122	−0·349	0·465	0·241	−0·514
31	−0·054	0·335	−0·167	−0·068	−0·163	0·289	0·206	−0·356
32	0·073	0·671	−0·092	−0·107	−0·039	0·066	0·223	−0·103
33	0·216	0·521	0·027	−0·256	0·048	−0·081	0·212	0·214
34	0·227	0·024	0·125	−0·400	−0·015	−0·051	0·143	0·442
35	0·112	−0·411	0·332	−0·482	−0·157	0·144	−0·018	0·572
36	−0·042	−0·601	0·369	−0·429	−0·308	0·386	−0·133	0·548
37	−0·058	−0·506	0·301	−0·259	−0·316	0·631	−0·174	0·420
38	−0·004	−0·280	0·185	−0·092	−0·123	0·667	−0·193	0·198
39	−0·075	0·107	0·094	−0·001	0·124	0·381	−0·281	−0·115
40	−0·162	0·391	0·082	0·094	0·293	−0·078	−0·393	−0·390
41	−0·272	0·549	0·133	0·158	0·173	−0·426	−0·429	−0·518
42	−0·466	0·437	0·196	0·249	−0·026	−0·403	−0·294	−0·624
43	−0·740	−0·100	0·106	0·232	−0·227	−0·180	−0·205	−0·712
44	−0·858	−0·559	−0·020	0·060	−0·305	−0·000	−0·263	−0·752
45	−0·734	−0·429	−0·138	−0·027	−0·290	0·087	−0·444	−0·725
46	−0·494	0·067	−0·231	0·047	−0·300	0·119	−0·402	−0·558
47	−0·364	0·628	−0·357	0·290	−0·194	0·297	−0·459	−0·271
48	−0·327	0·701	−0·472	0·446	−0·029	0·238	−0·585	−0·148
49	−0·278	0·312	−0·545	0·295	0·278	−0·055	−0·503	−0·024
50	−0·304	−0·208	−0·508	−0·031	0·476	−0·367	−0·175	0·161

TABLE 3·2. Serial correlations of series 1, the two subseries of 240 each and the four subseries of 120 terms each

Order of correlation	$1a+1b$	$1c+1d$	$1e+1f$	$1g+1h$	$1a+1b+1c+1d$	$1e+1f+1g+1h$	Series 1 in total
1	0·702	0·726	0·748	0·842	0·713	0·801	0·762
2	0·260	0·292	0·319	0·579	0·272	0·462	0·377
3	−0·027	−0·024	−0·039	0·347	−0·033	0·172	0·079
4	−0·119	−0·135	−0·190	0·146	−0·132	−0·011	−0·067
5	−0·016	−0·099	−0·169	−0·019	−0·056	−0·095	−0·078
6	0·136	−0·011	−0·088	−0·144	0·067	−0·126	−0·039
7	0·218	0·056	0·017	−0·224	0·136	−0·123	−0·007
8	0·265	0·107	0·087	−0·221	0·176	−0·094	0·022
9	0·208	0·104	0·060	−0·150	0·141	−0·064	0·018
10	0·034	0·055	−0·092	−0·068	0·039	−0·073	−0·036
11	−0·116	−0·032	−0·285	−0·024	−0·055	−0·119	−0·103
12	−0·174	−0·101	−0·364	−0·040	−0·106	−0·159	−0·145
13	−0·065	−0·141	−0·291	−0·092	−0·080	−0·160	−0·128
14	0·182	−0·181	−0·101	−0·137	0·005	−0·106	−0·052
15	0·398	−0·153	0·049	−0·177	0·110	−0·062	0·029
16	0·398	−0·026	0·086	−0·193	0·165	−0·043	0·069
17	0·227	0·157	0·050	−0·177	0·161	−0·033	0·065
18	0·007	0·283	0·005	−0·132	0·109	−0·020	0·032
19	−0·179	0·324	−0·032	−0·077	0·046	−0·009	−0·007
20	−0·211	0·242	−0·073	−0·017	0·009	−0·012	−0·025
21	−0·143	0·098	−0·107	0·078	−0·007	0·013	−0·005
22	−0·049	−0·017	−0·124	0·118	−0·007	0·023	0·019
23	0·044	−0·088	−0·152	0·060	−0·007	−0·017	0·011
24	0·137	−0·074	−0·101	−0·031	0·022	−0·091	−0·011
25	0·100	0·031	−0·149	−0·133	0·039	−0·153	−0·041
26	0·007	0·065	−0·087	−0·244	0·042	−0·196	−0·079
27	−0·133	−0·001	0·028	−0·302	−0·006	−0·176	−0·108
28	−0·238	−0·109	0·148	−0·305	−0·082	−0·122	−0·126
29	−0·172	−0·192	0·200	−0·281	−0·109	−0·085	−0·116
30	−0·035	−0·225	0·184	−0·209	−0·098	−0·054	−0·075
31	0·117	−0·204	0·118	−0·094	−0·070	−0·011	−0·018
32	0·232	−0·191	0·067	0·027	−0·040	0·045	0·039
33	0·196	−0·190	0·091	0·126	−0·053	0·109	0·069
34	0·013	−0·176	0·135	0·143	−0·094	0·119	0·048
35	−0·174	−0·100	0·185	0·103	−0·117	0·090	0·008
36	−0·316	−0·028	0·193	0·051	−0·115	0·035	−0·037
37	−0·272	0·038	0·205	−0·011	−0·046	−0·009	−0·047
38	−0·074	0·092	0·158	−0·061	0·043	−0·036	−0·023
39	0·084	0·098	0·019	−0·080	0·056	−0·062	−0·017
40	0·154	0·088	−0·175	−0·079	0·038	−0·098	−0·030
41	0·118	0·076	−0·342	−0·044	0·014	−0·110	−0·038
42	−0·021	0·121	−0·365	0·054	0·009	−0·056	−0·012
43	−0·181	0·119	−0·247	0·136	0·012	0·032	0·033
44	−0·241	0·072	−0·124	0·125	0·009	0·085	0·055
45	−0·186	0·038	−0·057	0·078	0·001	0·099	0·048
46	0·004	0·000	−0·059	0·063	0·009	0·093	0·040
47	0·159	−0·023	−0·039	0·044	0·018	0·093	0·041
48	0·190	−0·080	−0·016	−0·019	−0·015	0·093	0·035
49	0·106	−0·199	0·021	−0·077	−0·087	0·099	0·022
50	−0·079	−0·252	0·096	−0·106	−0·164	0·115	0·007

stressed in statistical studies of time-series: even a number of terms which is 'large' in the usual sense of the theory of random sampling may be 'small' for time-series. If we have 480 *independently chosen* members of a population we can make estimates of its parameters with some exactitude; but 480 consecutive terms of a time-series may not provide at all an accurate idea of the process which generated it, even if that section of the series was chosen at random. The terms themselves are not independent one of the next.

3·8. So far as oscillatory effects are concerned the failure to damp does not appear to affect the period of the undulations in the correlogram so much as their amplitude, so that an estimate of the autoregressive period may be derived from an examination of the correlogram itself. Nevertheless, we may notice that this estimate is not always a very good one. For the full series of 480 or 240 terms as the case may be we have

Series	Autoregressive period from (3·4)	Mean distance between troughs in the correlogram
1	9·25	8·33
2	19·53	12·0
3	6·92	7·67
4	3·00	2·80

The agreement is fair for series 1, 3 and 4; but poor for series 2.

3·9. The explanation of these effects seems to me to be that, for short series, there are only a few oscillations available. The variance of a short series should not differ systematically from the parent variance to any serious extent; but the covariances may systematically exceed in absolute value the corresponding parent covariance. In fact, as we proceed along the series the oscillations change in phase and when we have gone far enough will be quite unrelated in phase to the initial oscillations; but if we only go so far as the second or third oscillation, the final oscillation may not, so to speak, have had time to get very much out of step with the first. In consequence the correlations for such a series will tend to be higher than those for the infinite series (which explains the failure to damp) and may fail to express the autoregressive period (which explains the differences between that period and observed oscillations in the correlogram). Furthermore, as has happened in series 1*a*, a short series of 60 terms may show what appears to be a trend, which is reflected in the correlogram and obscures the oscillatory effect.

3·10. In short, it appears that for oscillatory series the operative sample number is not the number of terms in the series but the number of oscillations which it covers. In series 1, for example, there are about 57 oscillations of the autoregressive period and it is 57, not 480, which is the effective sample number. Even this may be deceptively high, if there are traces of correlation

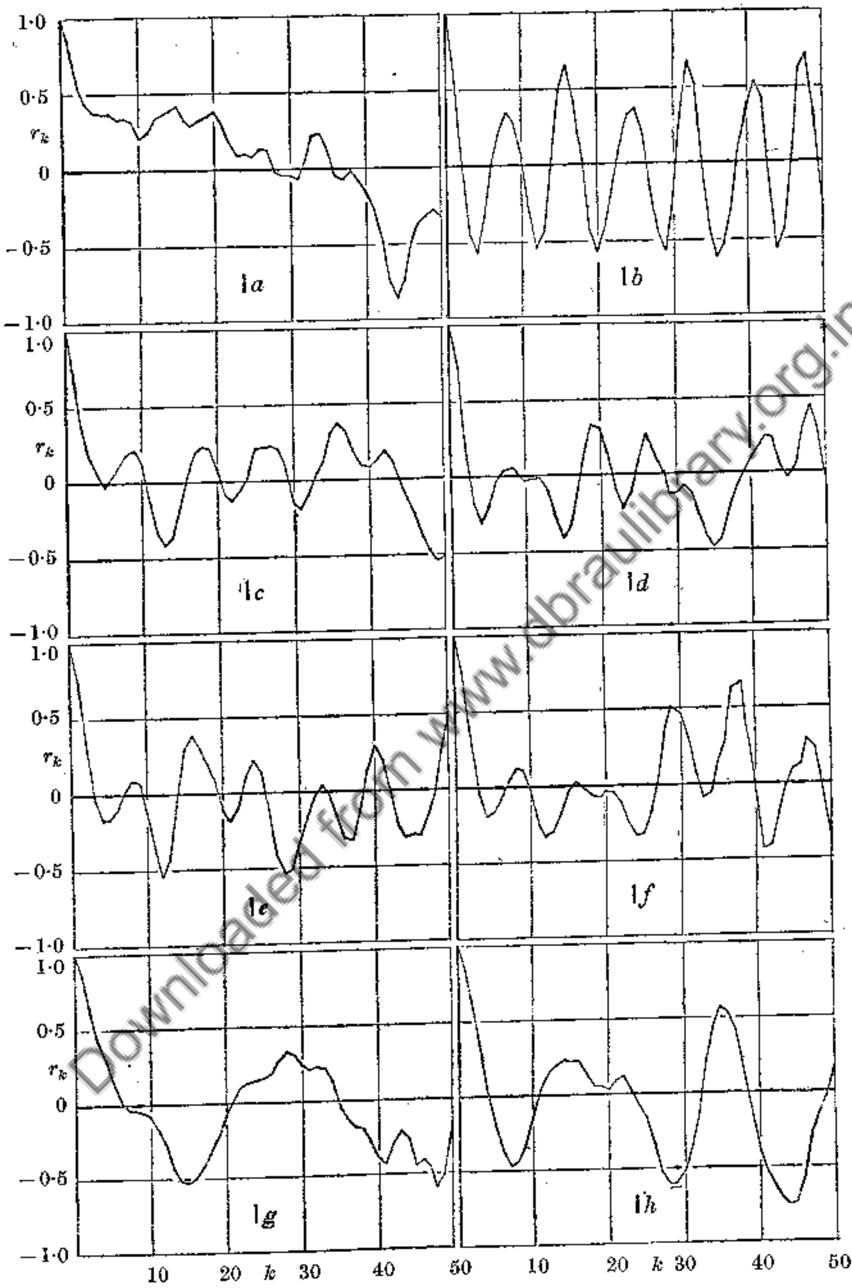

Fig. 3·1. Correlograms of series 1a to 1h (eight series of 60 terms each composing series 1)

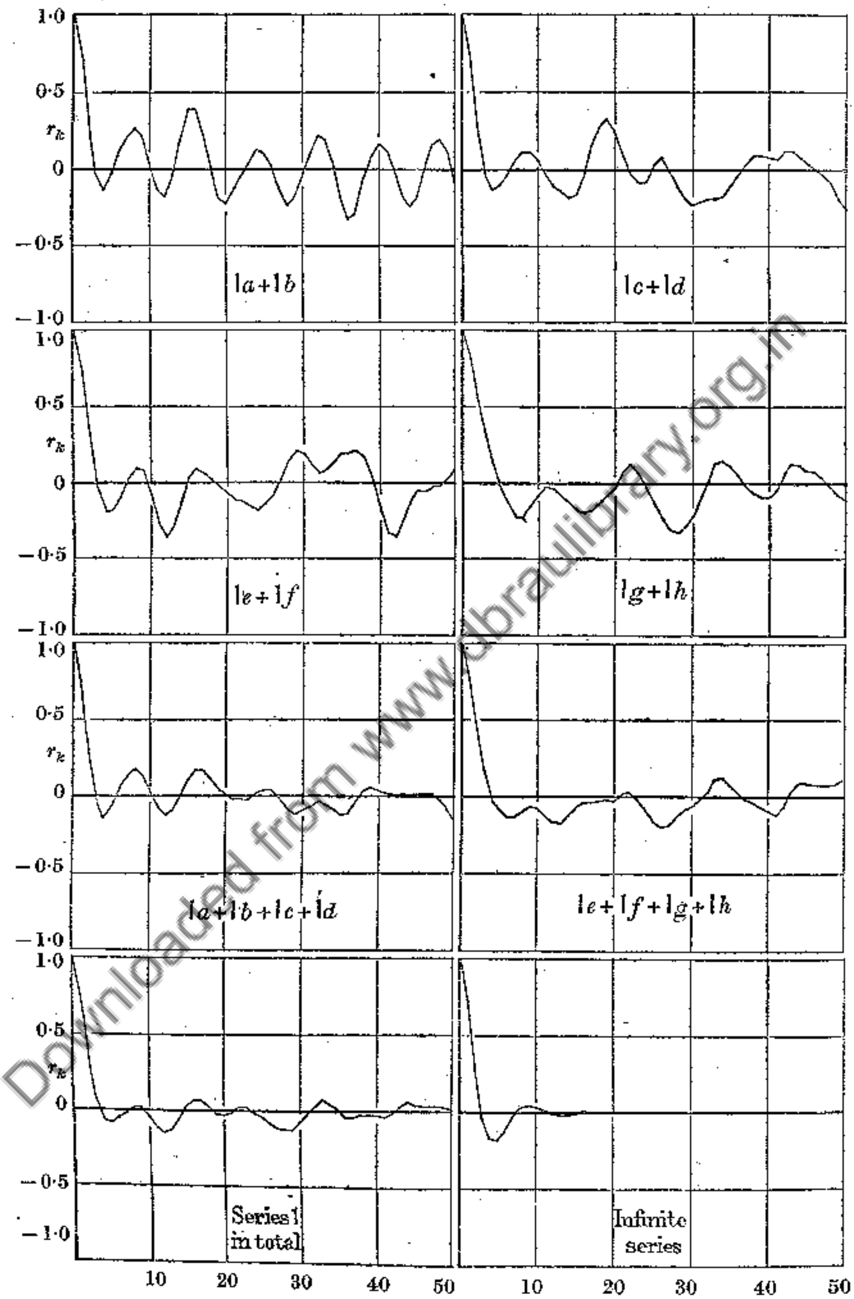

Fig. 3·2. Correlograms of four series of 120 terms and two series of 240 terms composing series 1 with correlogram of series 1 and that of an infinite series with the same constants between the periods (however defined) of successive oscillations. Again, in series 2, there are only about 12 oscillations of the autoregressive period, and it is not nearly so surprising as appears at first sight to find the correlogram possessing a mean distance between troughs of only 12 units against a theoretical value of 19·25.

TABLE 3·3. Serial correlations of series 2, the two subseries of 120 terms each and the four subseries of 60 terms each

Order of correlation	1st 60 terms (2*a*)	2nd 60 terms (2*b*)	3rd 60 terms (2*c*)	4th 60 terms (2*d*)	1st 120 terms (2*a*+2*b*)	2nd 120 terms (2*c*+2*d*)	Series 2 in total
1	0·774	0·910	0·804	0·839	0·863	0·825	0·850
2	0·437	0·731	0·475	0·584	0·618	0·542	0·606
3	0·131	0·550	0·125	0·256	0·379	0·218	0·350
4	−0·144	0·377	−0·219	−0·009	0·160	−0·076	0·127
5	−0·324	0·230	−0·392	−0·237	0·001	−0·288	−0·044
6	−0·307	0·130	−0·355	−0·337	−0·060	−0·354	−0·124
7	−0·190	0·048	−0·179	−0·370	−0·075	−0·321	−0·147
8	−0·069	−0·056	−0·008	−0·287	−0·104	−0·225	−0·138
9	0·154	−0·138	0·048	−0·151	−0·076	−0·120	−0·094
10	0·325	−0·202	−0·035	0·014	−0·047	−0·043	−0·057
11	0·374	−0·246	−0·105	0·148	−0·038	0·036	−0·032
12	0·368	−0·279	−0·137	0·207	−0·045	0·082	−0·015
13	0·320	−0·276	−0·131	0·223	−0·056	0·099	−0·009
14	0·165	−0·280	−0·190	0·197	−0·101	0·056	−0·040
15	0·007	−0·305	−0·211	0·111	−0·170	−0·003	−0·076
16	−0·116	−0·306	−0·180	−0·043	−0·216	−0·092	−0·128
17	−0·212	−0·254	−0·171	−0·207	−0·233	−0·170	−0·166
18	−0·208	−0·212	−0·136	−0·310	−0·209	−0·191	−0·172
19	−0·124	−0·150	−0·082	−0·318	−0·147	−0·147	−0·151
20	−0·054	−0·027	0·112	−0·202	−0·049	−0·016	−0·100
21	−0·055	0·141	0·266	−0·013	0·043	0·115	−0·029
22	−0·084	0·300	0·328	0·218	0·106	0·211	0·049
23	−0·201	0·429	0·308	0·382	0·140	0·248	0·098
24	−0·310	0·539	0·220	0·488	0·189	0·221	0·143
25	−0·317	0·581	0·105	0·467	0·240	0·162	0·190
26	−0·402	0·606	−0·019	0·265	0·250	0·060	0·186
27	−0·432	0·612	−0·187	−0·075	0·230	−0·078	0·135
28	−0·244	0·606	−0·282	−0·366	0·234	−0·171	0·099
29	0·037	0·630	−0·224	−0·557	0·269	−0·179	0·097
30	0·208	0·648	−0·269	−0·649	0·287	−0·206	0·094

3·11. These results are, in a way, very disappointing. They imply that before we can approach reasonable certainty about the true nature of the generating process of a series and estimate its fundamental constants a very large number of terms is required—something of the order of thousands rather than hundreds. This is unfortunate, particularly in economic work where series of such length are very rare; but the position has to be faced. As a general rule the correlogram even of short series will indicate whether damping exists and provide some evidence on the questions whether the series can be regarded as autoregressive. It is when we come to determine the damping factors and the other constants that difficulties arise. One such has been dealt with in the foregoing paragraphs. Another, the complications due to superposed variation, I have discussed elsewhere[10]. I proceed to consider a third.

3·12. Series 2 and 3 were combined to form a fifth series of 240 terms by adding the nth term of one to the nth term of the other. The serial correlations are shown in Table 3·6 and the correlogram of the full series in Fig. 3·6.

TABLE 3·4. Serial correlations of series 3, the two subseries of 120 terms each and the four subseries of 60 terms each

Order of correlation	1st 60 terms (3a)	2nd 60 terms (3b)	3rd 60 terms (3c)	4th 60 terms (3d)	1st 120 terms (3a+3b)	2nd 120 terms (3c+3d)	Series 3 in total
1	0·639	0·615	0·610	0·620	0·634	0·614	0·623
2	−0·012	−0·061	−0·142	−0·134	−0·040	−0·142	−0·105
3	−0·484	−0·502	−0·693	−0·632	−0·506	−0·673	−0·609
4	−0·563	−0·498	−0·688	−0·600	−0·540	−0·641	−0·596
5	−0·268	−0·197	−0·239	−0·163	−0·230	−0·167	−0·181
6	0·137	0·096	0·268	0·306	0·124	0·332	0·260
7	0·257	0·293	0·465	0·453	0·277	0·485	0·402
8	0·063	0·345	0·311	0·333	0·211	0·285	0·247
9	−0·167	0·207	0·041	0·102	0·037	−0·036	−0·011
10	−0·253	−0·079	−0·154	−0·079	−0·155	−0·248	−0·201
11	−0·122	−0·331	−0·194	−0·197	−0·239	−0·246	−0·225
12	0·025	−0·415	−0·161	−0·234	−0·199	−0·106	−0·135
13	0·090	−0·223	−0·150	−0·145	−0·039	0·030	−0·009
14	0·072	0·117	−0·094	0·105	0·157	0·123	0·125
15	−0·084	0·361	0·044	0·382	0·200	0·170	0·189
16	−0·231	0·388	0·234	0·496	0·092	0·177	0·159
17	−0·145	0·226	0·316	0·288	−0·020	0·102	0·055
18	0·057	0·002	0·173	−0·144	−0·075	−0·043	−0·069
19	0·234	−0·211	−0·149	−0·489	−0·041	−0·202	−0·144
20	0·401	−0·270	−0·409	−0·466	0·081	−0·235	−0·101
21	0·441	−0·110	−0·380	−0·067	0·197	−0·082	0·024
22	0·176	0·144	−0·063	0·396	0·178	0·176	0·141
23	−0·138	0·302	0·358	0·561	0·053	0·349	0·181
24	−0·309	0·251	0·588	0·281	−0·125	0·275	0·094
25	−0·269	0·005	0·364	−0·172	−0·234	−0·022	−0·074
26	−0·005	−0·269	−0·163	−0·366	−0·181	−0·313	−0·199
27	0·246	−0·362	−0·623	−0·225	−0·018	−0·380	−0·197
28	0·237	−0·257	−0·671	0·096	0·090	−0·168	−0·078
29	−0·111	−0·039	−0·228	0·261	0·038	0·175	0·075
30	−0·518	0·155	0·458	0·164	−0·102	0·402	0·174

Taking into account the known differences between theory and observation for series of this length we should have seen little or nothing in the correlogram of series 2 and 3 to indicate that it arose from the sum of two series with different autoregressive constants. On the face of it, this is a correlogram of a single damped series of the simple linear type defined by (3·1). The correlogram may, therefore, be insensitive. It indicates, correctly, that autoregression exists. It will give us an approximate idea of the kind of simple scheme which would account for the variation. But it does not establish that such a scheme is the only one and we are always left with the possibility that our model oversimplifies the reality. This, of course, is true of any scientific hypothesis, but whereas we can usually conduct further experiments to test a hypothesis in physical sciences such a course is rarely open to us in economics. The best we can do, as a rule, is to predict the next terms in the series and then wait to see how far the prediction is verified.

TABLE 3·5. Serial correlations of series 4, the two subseries of 120 terms each and the four subseries of 60 terms each

Order of correlation	1st 60 terms (4*a*)	2nd 60 terms (4*b*)	3rd 60 terms (4*c*)	4th 60 terms (4*d*)	1st 120 terms (4*a*+4*b*)	2nd 120 terms (4*c*+4*d*)	Series 4 in total
1	−0·699	−0·744	−0·675	−0·646	−0·719	−0·657	−0·684
2	0·257	0·387	0·129	0·027	0·312	0·069	0·174
3	0·249	−0·021	0·289	0·443	0·127	0·362	0·260
4	−0·560	−0·255	−0·397	−0·512	−0·414	−0·421	−0·414
5	0·569	0·350	0·210	0·305	0·454	0·206	0·305
6	−0·326	−0·352	−0·077	−0·008	−0·313	0·009	−0·127
7	−0·047	0·294	0·092	−0·252	0·088	−0·090	−0·007
8	0·313	−0·243	−0·252	0·309	0·081	−0·013	0·016
9	−0·376	0·144	0·430	−0·103	−0·156	0·224	0·064
10	0·218	−0·013	−0·413	−0·194	0·134	−0·334	−0·121
11	0·038	−0·049	0·131	0·388	−0·018	0·230	0·107
12	−0·310	0·145	0·244	−0·373	−0·104	0·002	−0·036
13	0·351	−0·177	−0·523	0·313	0·144	−0·151	−0·018
14	−0·351	0·108	0·537	−0·147	−0·211	0·174	−0·006
15	0·195	−0·116	−0·275	−0·076	0·172	−0·070	0·045
16	−0·026	−0·007	0·012	0·299	−0·145	0·018	−0·049
17	−0·100	0·141	0·090	−0·370	0·112	−0·069	−0·005
18	0·178	−0·280	−0·048	0·268	−0·066	0·153	0·070
19	−0·164	0·362	−0·006	−0·079	0·026	−0·180	−0·098
20	0·126	−0·317	−0·071	−0·084	0·049	0·065	0·063
21	0·029	0·189	0·211	0·174	−0·096	0·113	0·015
22	−0·118	0·061	−0·322	−0·146	0·159	−0·241	−0·068
23	0·275	−0·160	0·378	0·075	−0·116	0·279	0·113
24	−0·344	0·147	−0·329	−0·010	0·012	−0·218	−0·125
25	0·311	−0·214	0·140	−0·071	0·034	0·057	0·056
26	−0·038	0·273	0·147	0·218	0·017	0·190	0·103
27	−0·299	−0·354	−0·345	−0·328	−0·114	−0·314	−0·221
28	0·561	0·321	0·370	0·214	0·208	0·234	0·223
29	−0·638	−0·133	−0·235	0·041	−0·222	−0·001	−0·083
30	0·559	−0·172	0·127	−0·251	0·167	−0·158	−0·063

3·13. For purpose of illustration and later reference it may be useful to give at this stage the correlogram of the Beveridge series (Table 1·6). The first 60 serial correlations are shown in Table 3·7 and the correlogram in Fig. 3·7. It will be observed that the correlogram is heavily damped, indicating autoregression. There appear to be present two oscillatory movements, one with a period of about 15 years and a smaller one with a period of about 5 years. On this evidence alone we should see no reason to suspect the existence of more than two. In the next chapter we shall see how greatly this conflicts with the results obtained by Sir William Beveridge himself in an analysis of the same series based on the periodogram.

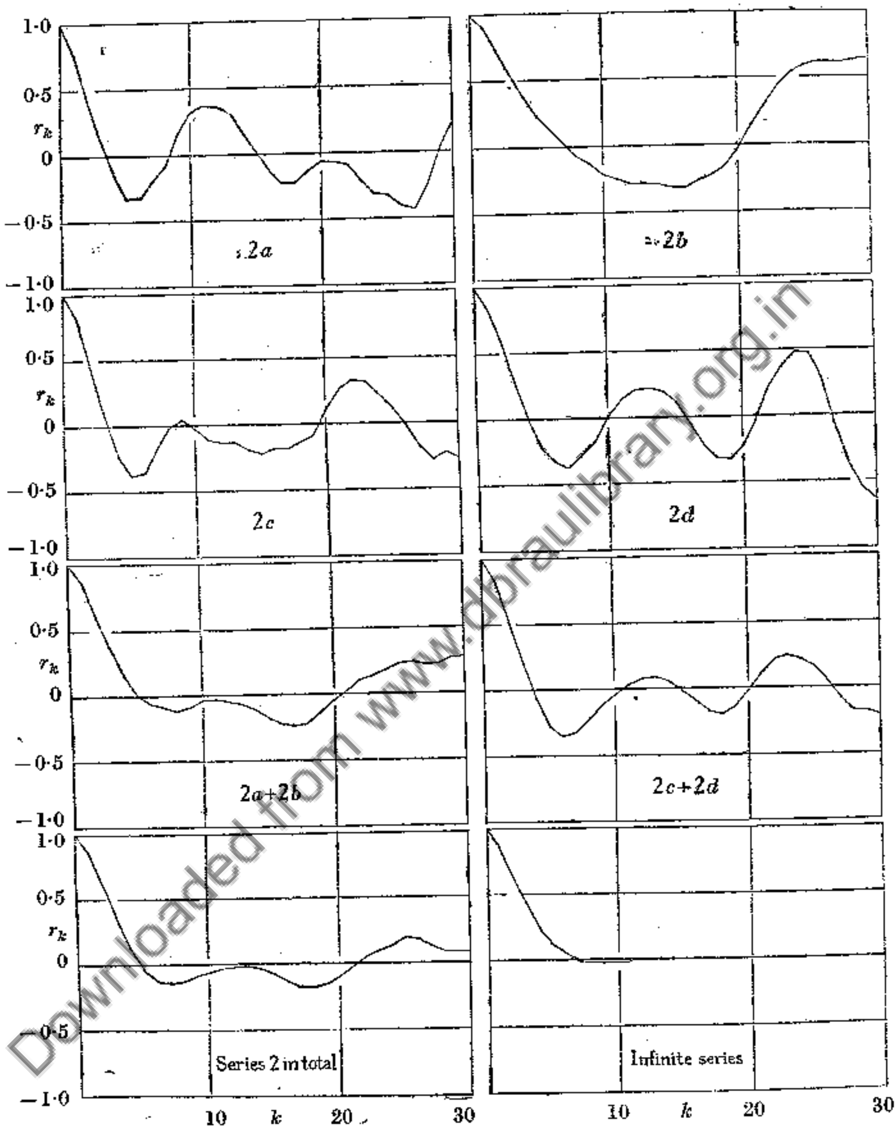

Fig. 3·3. Correlograms of series $2a$ to $2d$ (of 60 terms each), of series $2a+2b$ and $2c+2d$ (120 terms each), and series 2 (240 terms) together with the correlogram of the infinite series with the same constants

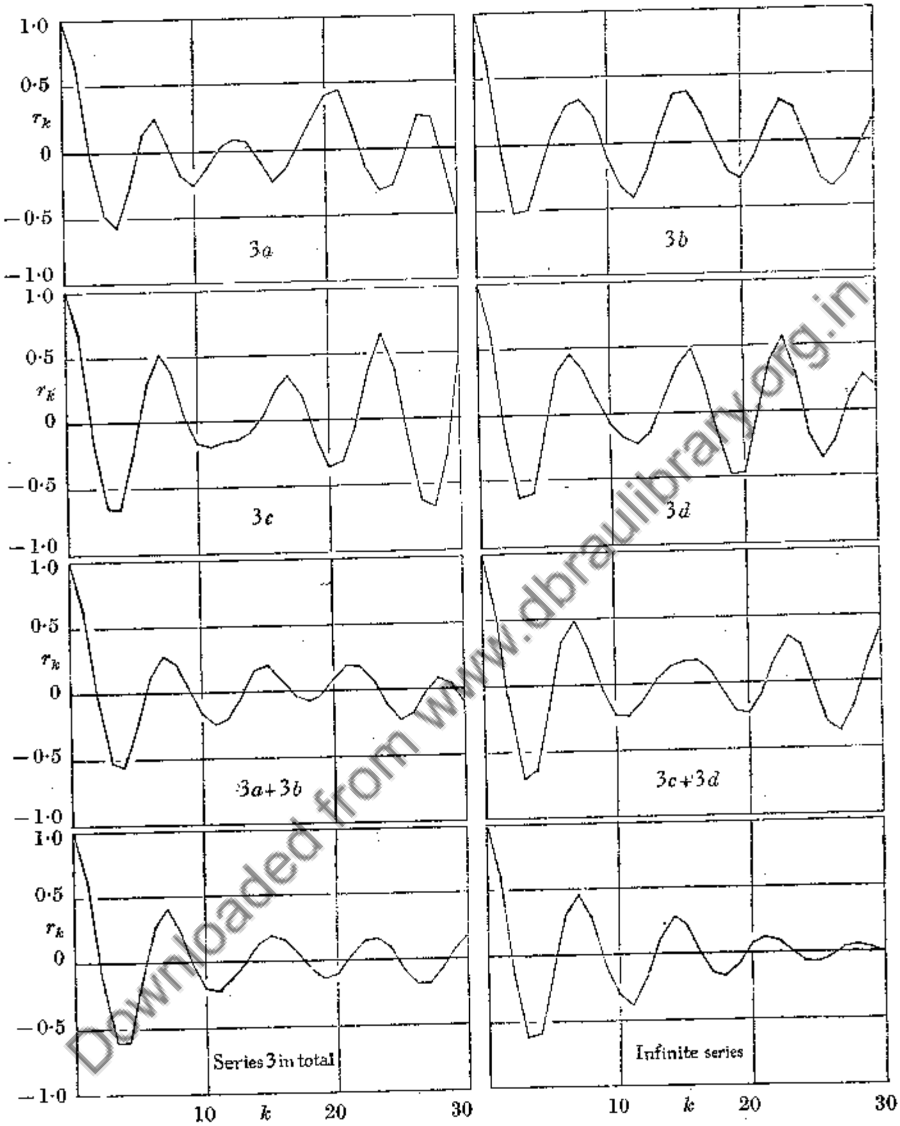

Fig. 3·4. Correlogram of series $3a$ to $3d$ (60 terms each), of series $3a+3b$ and $3c+3d$ (120 terms each), and series 3 (240 terms) together with the correlogram of the infinite series with the same constants

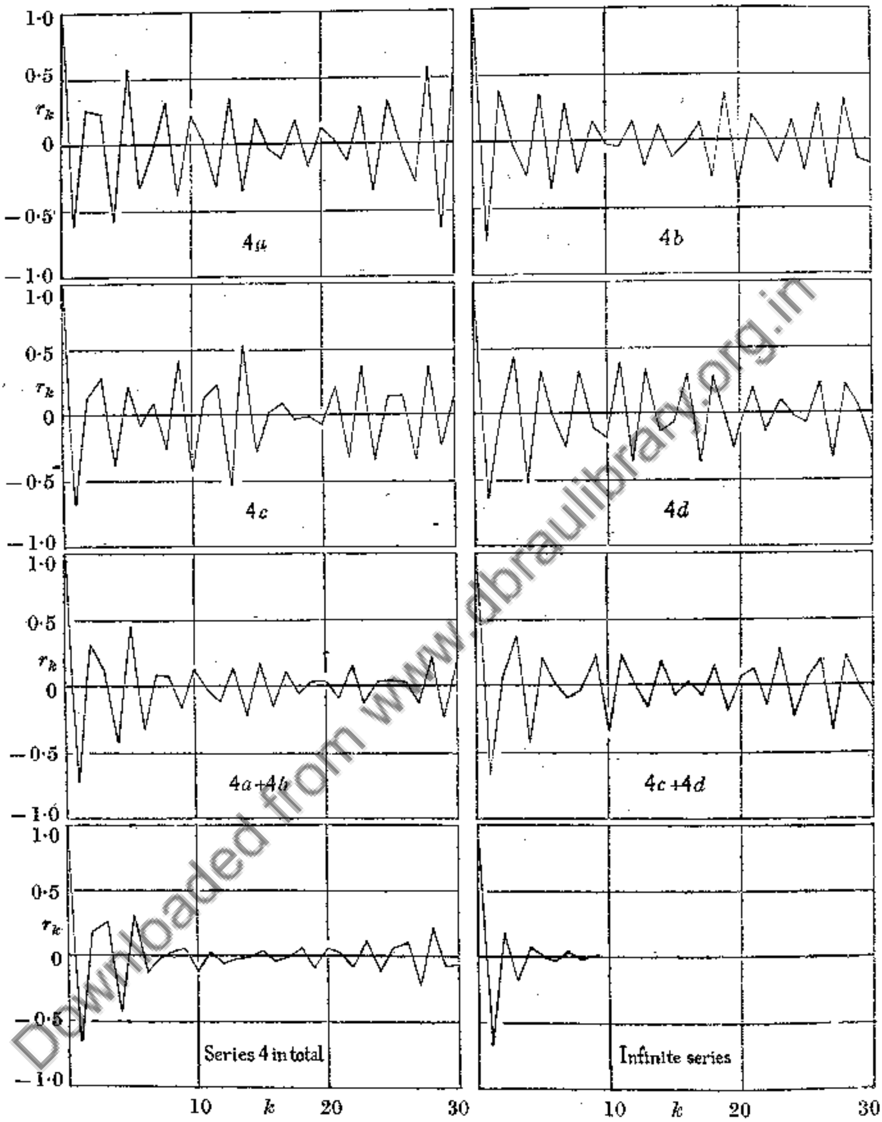

Fig. 3·5. Correlograms of series $4a$ to $4d$ (of 60 terms each), of series $4a+4b$ and $4c+4d$ (120 terms each) and series 4 (240 terms) together with the correlogram of the infinite series with the same constants

TABLE 3·6. Serial correlations of series 2 and 3, the two subseries of 120 terms each and the four subseries of 60 terms each

Order of correlation	1st 60 terms	2nd 60 terms	3rd 60 terms	4th 60 terms	1st 120 terms	2nd 120 terms	Series 2 and 3 in total
1	0·714	0·794	0·650	0·717	0·754	0·680	0·722
2	0·215	0·408	−0·049	0·189	0·301	0·061	0·197
3	−0·176	0·123	−0·566	−0·242	−0·055	−0·406	−0·216
4	−0·371	−0·015	−0·617	−0·319	−0·226	−0·457	−0·325
5	−0·335	−0·046	−0·314	−0·111	−0·221	−0·180	−0·171
6	−0·154	−0·015	0·073	0·200	−0·117	0·174	0·061
7	−0·022	0·056	0·324	0·358	−0·019	0·344	0·180
8	−0·028	0·065	0·314	0·357	−0·006	0·263	0·132
9	0·017	−0·011	0·121	0·189	−0·009	0·032	0·006
10	0·102	−0·195	−0·070	0·010	−0·046	−0·148	−0·106
11	0·203	−0·396	−0·130	−0·110	−0·096	−0·156	−0·133
12	0·227	−0·482	−0·104	−0·127	−0·134	−0·034	−0·086
13	0·171	−0·382	−0·049	−0·065	−0·106	0·100	−0·014
14	0·037	−0·228	−0·030	0·103	−0·060	0·163	0·050
15	−0·192	−0·144	0·017	0·291	−0·083	0·154	0·060
16	−0·323	−0·087	0·099	0·327	−0·129	0·077	0·006
17	−0·247	−0·012	0·061	0·117	−0·108	−0·077	−0·072
18	−0·017	−0·005	−0·087	−0·213	−0·048	−0·221	−0·133
19	0·195	−0·064	−0·254	−0·438	−0·002	−0·271	−0·146
20	0·402	−0·060	−0·258	−0·360	0·072	−0·148	−0·072
21	0·442	0·036	−0·094	−0·008	0·118	0·085	0·046
22	0·229	0·170	0·173	0·387	0·094	0·308	0·150
23	−0·089	0·298	0·395	0·549	0·035	0·380	0·175
24	−0·307	0·400	0·420	0·378	0·003	0·231	0·110
25	−0·335	0·370	0·141	−0·025	−0·005	−0·080	−0·018
26	−0·244	0·272	−0·279	−0·324	0·007	−0·344	−0·131
27	−0·072	0·189	−0·561	−0·407	0·044	−0·376	−0·144
28	−0·001	0·134	−0·490	−0·277	0·066	−0·160	−0·047
29	−0·080	0·179	−0·034	−0·095	0·071	0·168	0·106
30	−0·245	0·263	0·412	0·018	0·069	0·354	0·209

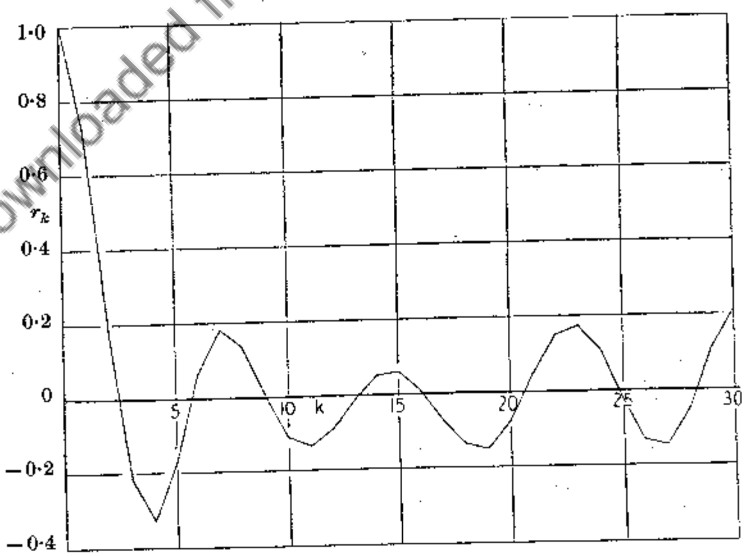

Fig. 3·6. Correlogram of the sum of series 2 and 3

TABLE 3·7. Serial correlations of the Beveridge trend-free wheat-price series

Order of corre-lation, k	Value of correlation, r_k	Order of corre-lation, k	Value of correlation, r_k	Order of corre-lation, k	Value of correlation, r_k
1	0·562	21	−0·021	41	0·008
2	0·103	22	−0·062	42	0·034
3	−0·075	23	−0·088	43	0·065
4	−0·092	24	−0·084	44	0·099
5	−0·082	25	−0·076	45	0·009
6	−0·136	26	−0·091	46	−0·036
7	−0·211	27	−0·052	47	−0·013
8	−0·261	28	−0·032	48	0·042
9	−0·192	29	−0·012	49	0·062
10	−0·070	30	0·059	50	0·065
11	−0·003	31	0·060	51	0·050
12	−0·015	32	−0·008	52	0·009
13	−0·012	33	−0·039	53	−0·027
14	0·047	34	0·007	54	−0·053
15	0·101	35	0·056	55	−0·073
16	0·158	36	0·010	56	−0·106
17	0·109	37	−0·004	57	−0·084
18	0·002	38	−0·015	58	−0·019
19	−0·075	39	−0·047	59	0·003
20	−0·062	40	−0·047	60	0·010

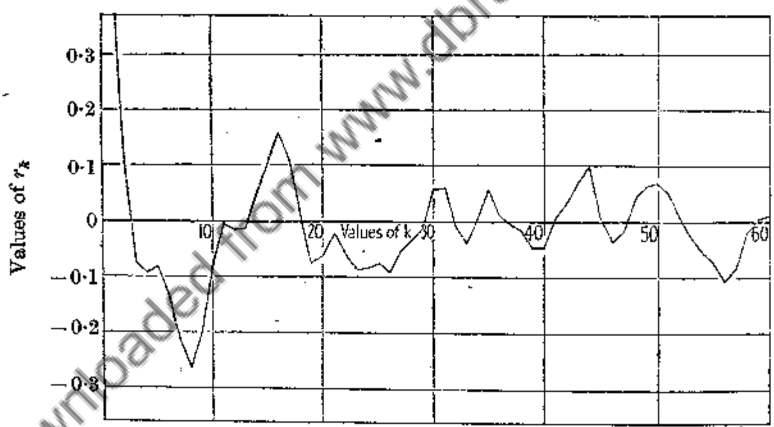

Fig. 3·7. Correlogram of the Beveridge trend-free wheat-price index

Summary

3·14. (1) For autoregressive series of finite length the correlogram differs systematically from the theoretical results appropriate to an infinite series, in that it fails to damp according to expectation.

(2) It appears likely that very long series are required to give decisive indications of the true nature of the generating process.

(3) Correlogram analysis appears to indicate reliably the existence of autoregressive effects, but for short series is somewhat insensitive.

(4) For some purposes in the theory of time series the material sampling number, from the point of view of judgment of experimental error, appears to be the number of periods covered, not the number of terms in the series.

CHAPTER 4

THE PERIODOGRAM OF AUTOREGRESSIVE SERIES*

4·1. The technique of periodogram analysis has been briefly described in Chapter 1 (**1·12–1·16**). I proceed to give the results of applying that technique to the first three experimental series.

4·2. For series 1, the standardized ordinates of the periodogram were calculated for each integral trial value from 2 to 50 inclusive and are shown in Table 4·1, the periodogram itself being given in Fig. 4·1.

TABLE 4·1. Standardized periodogram ordinates for series 1 and integral trial periods from 2 to 50 inclusive

Trial period, p	$E(p)$	Trial period, p	$E(p)$	Trial period, p	$E(p)$
2	0·0001	19	0·0000	35	0·0220
3	0·0002	20	0·0510	36	0·0174
4	0·0000	21	0·0051	37	0·0083
5	0·0040	22	0·0212	38	0·0041
6	0·0007	23	0·0011	39	0·0176
7	0·0145	24	0·0088	40	0·0255
8	0·0136	25	0·0283	41	0·0372
9	0·0062	26	0·0295	42	0·0423
10	0·0022	27	0·0222	43	0·0317
11	0·0052	28	0·0186	44	0·0095
12	0·0163	29	0·0022	45	0·0192
13	0·0097	30	0·0039	46	0·0188
14	0·0199	31	0·0005	47	0·0093
15	0·0061	32	0·0061	48	0·0047
16	0·0188	33	0·0033	49	0·0005
17	0·0107	34	0·0056	50	0·0077
18	0·0120				

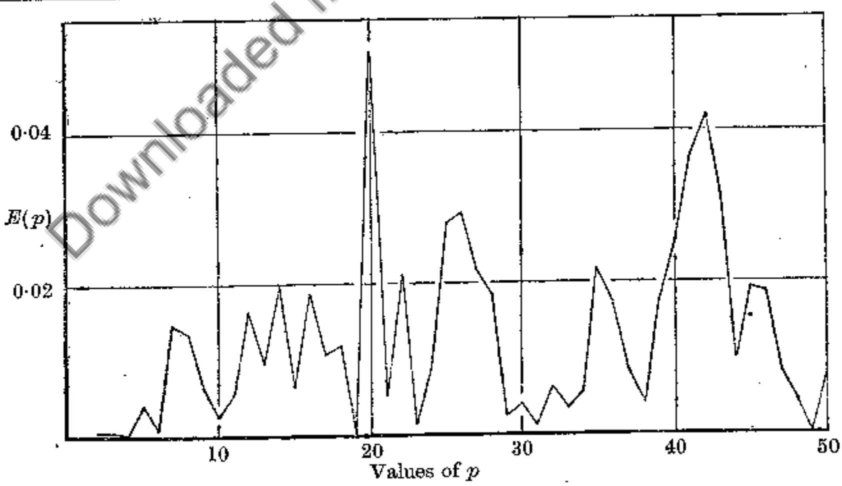

Fig. 4·1. Periodogram of series 1

* Most of this chapter was incorporated in a paper 'On the analysis of oscillatory time-series' read before the Royal Statistical Society on 17 January 1945. Reference may be made to that paper for the consideration of some further points and to the report of the ensuing discussion for additional comments.

This series has a m.d. (peaks) of 5·05, a m.d. (upcrosses) of 8·30 (see **6·4** and **6·10**) and an autoregressive period of 9·25 (**1·26**). A peak at any of these points in the periodogram would at least be understandable. Actually there are about a dozen peaks, two of them, at 20 and 42, standing out as offering substantial evidence of significant periods, but nothing at all striking in the very places where we might expect them.

4·3. Certain tests have been advanced to express the probability that peaks of a given magnitude can arise in random sampling. In my view they are unsound,* but if they are applied they give significance to the periods

TABLE 4·2. Standardized periodogram ordinates for series 1 and certain non-integral trial periods between 8 and 9 units

Trial period, p	$E(p)$	Trial period, p	$E(p)$	Trial period, p	$E(p)$
8·000	0·0136	8·333	0·0177	8·667	0·0006
8·167	0·0115	8·400	0·0138	8·750	0·0105
8·200	0·0045	8·500	0·0199	8·800	0·0028
8·250	0·0129	8·600	0·0052	9·000	0·0062

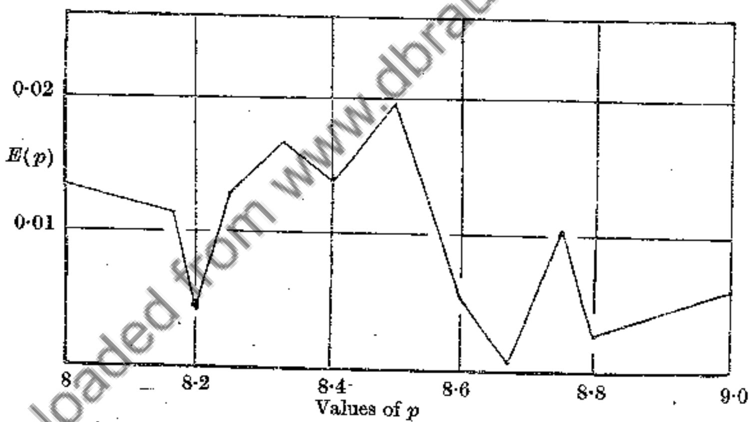

Fig. 4·2. Periodogram of series 1 for non-integral trial values between 8 and 9

at 20, 26 and 42. All three are illusory in the sense that they arise as sampling effects in a series generated by a process which gives no such periods either in the strict sense or as mean values in distributions. The periodogram is about as misleading as it could be.

4·4. According to the Schuster theory there will be, corresponding to a harmonic of period p in the original data, a peak in the periodogram of width $2p^2/(mp)$ or roughly $2p^2/N$, where N is the number of terms in the series. Calculating ordinates at integral trial periods only may therefore allow certain harmonics to slip through the net if $2p^2/N$ is much less than unity and narrower intervals are required for smaller values of p. Table 4·2

* See the paper referred to in the previous footnote.

and Fig. 4·2 show the periodogram for certain non-integral values from 8 to 9. We now find peaks at 8·33, 8·5 and 8·75, the latter probably being negligible. Even if we take the most favourable view and regard the other two as manifestations of a single period ($2p^2/N$ being about 0·3) the peak is far less than that of the three spurious periods and, indeed, less than those at 22 and 35 as well.

TABLE 4·3. Ordinates of the Whittaker diagram for series 1 and integral trial periods from 2 to 50 inclusive

Trial period, p	η_p^2	Trial period, p	η_p^2	Trial period, p	η_p^2
2	0·0000	19	0·0155	35	0·0632
3	0·0002	20	0·0629	36	0·0582
4	0·0006	21	0·0312	37	0·0843
5	0·0042	22	0·0463	38	0·0492
6	0·0009	23	0·0548	39	0·0563
7	0·0154	24	0·0421	40	0·1156
8	0·0137	25	0·0692	41	0·1207
9	0·0075	26	0·0518	42	0·1076
10	0·0066	27	0·0397	43	0·1146
11	0·0075	28	0·0692	44	0·1249
12	0·0176	29	0·0391	45	0·0639
13	0·0167	30	0·0150	46	0·1165
14	0·0367	31	0·0451	47	0·0441
15	0·0070	32	0·0612	48	0·0810
16	0·0346	33	0·0604	49	0·1103
17	0·0361	34	0·0657	50	0·1044
18	0·0207				

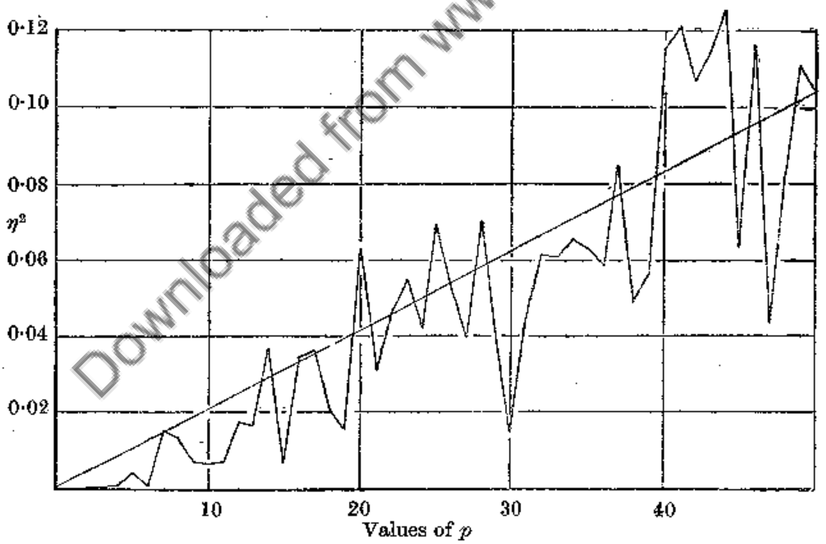

Fig. 4·3. Modified form of the Whittaker periodogram of series 1

4·5. As a matter of interest I give the ordinates of the Whittaker periodogram (**1·6**) in Table 4·3 and the periodogram itself in Fig. 4·3. It will be seen that the ordinate increases systematically and the real test lies in the

TABLE 4·4. Standardized periodogram ordinates for series 2 and integral trial periods from 2 to 40 inclusive

Trial period, p	$E(p)$	Trial period, p	$E(p)$	Trial period, p	$E(p)$
2	0·0000	15	0·0147	28	0·1448
3	0·0003	16	0·0082	29	0·0953
4	0·0004	17	0·0543	30	0·0425
5	0·0019	18	0·0147	31	0·0321
6	0·0000	19	0·0200	32	0·0151
7	0·0017	20	0·0592	33	0·0293
8	0·0044	21	0·0241	34	0·0465
9	0·0115	22	0·0090	35	0·0561
10	0·0006	23	0·0184	36	0·0635
11	0·0109	24	0·0221	37	0·0787
12	0·0257	25	0·0376	38	0·0747
13	0·0375	26	0·0842	39	0·0598
14	0·0163	27	0·1349	40	0·0529

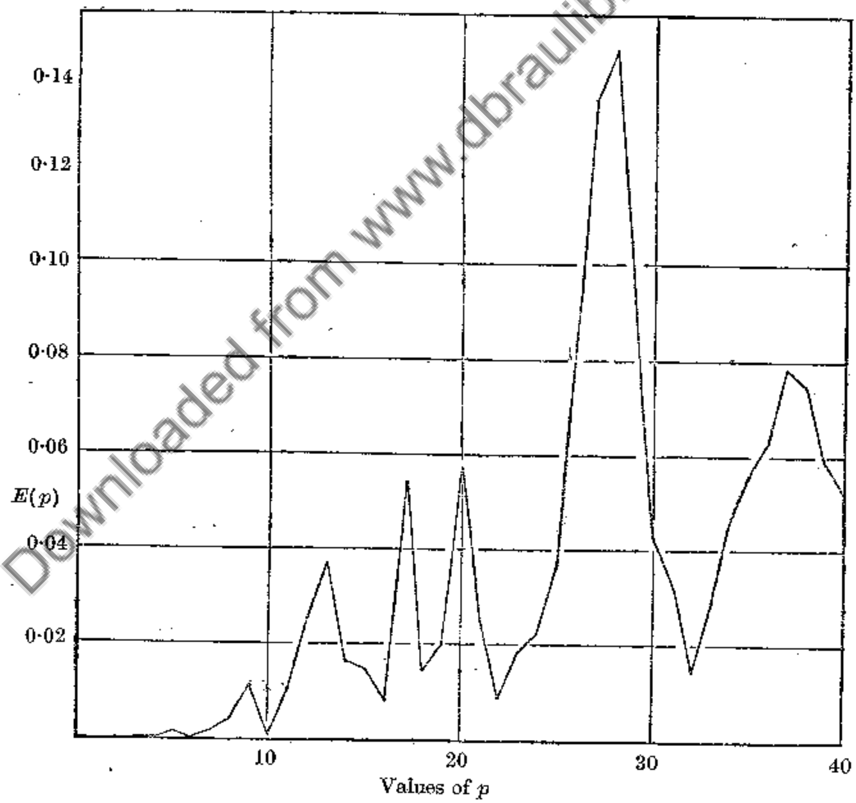

Fig. 4·4. Periodogram of series 2

divergence of particular values from the line $\eta^2 = kp$. There is much the same sort of fluctuation as in the Schuster form, and it is even more difficult to interpret; for in the latter a peak at trial period p can only correspond to a harmonic of period p (if it indicates a harmonic at all), whereas in the Whittaker form a peak can arise from harmonics with periods which are any integral multiple of p.

TABLE 4·5. Standardized periodogram ordinates for series 3 and integral trial periods from 2 to 40 inclusive

Trial period, p	$E(p)$	Trial period, p	$E(p)$	Trial period, p	$E(p)$
2	0·0000	15	0·0047	28	0·0009
3	0·0004	16	0·0167	29	0·0013
4	0·0003	17	0·0167	30	0·0011
5	0·0047	18	0·0165	31	0·0047
6	0·0457	19	0·0009	32	0·0057
7	0·0533	20	0·0099	33	0·0071
8	0·0277	21	0·0041	34	0·0068
9	0·0057	22	0·0013	35	0·0047
10	0·0135	23	0·0087	36	0·0034
11	0·0210	24	0·0063	37	0·0019
12	0·0009	25	0·0027	38	0·0031
13	0·0020	26	0·0017	39	0·0011
14	0·0068	27	0·0019	40	0·0012

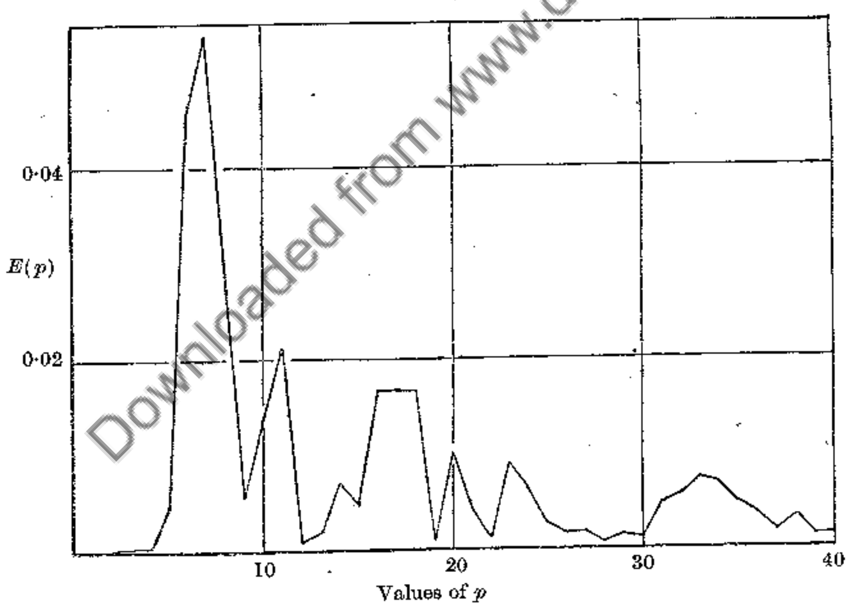

Fig. 4·5. Periodogram of series 3

4·6. For series 2 the Schuster periodogram was computed for all integral trial values from 2 to 40 inclusive, the ordinates being given in Table 4·4 and the periodogram in Fig. 4·4. There are fewer serrations in the periodogram, but peaks appear at 13, 17, 20, 28 and 37, the last but one being particularly striking. The m.d. (peaks) for the series is 5·57, the m.d. (upcrosses) 12·39 and the autoregressive period 19·53. Once again we reach the conclusion that the periodogram is grossly misleading.

TABLE 4·6. Standardized periodogram ordinates for series 2 and 3 and integral trial periods from 2 to 40 inclusive

Trial period, p	$E(p)$	Trial period, p	$E(p)$	Trial period, p	$E(p)$
2	0·0000	15	0·0055	28	0·0777
3	0·0003	16	0·0019	29	0·0494
4	0·0005	17	0·0323	30	0·0170
5	0·0023	18	0·0303	31	0·0079
6	0·0243	19	0·0120	32	0·0051
7	0·0194	20	0·0152	33	0·0125
8	0·0157	21	0·0034	34	0·0187
9	0·0004	22	0·0023	35	0·0198
10	0·0046	23	0·0015	36	0·0249
11	0·0305	24	0·0075	37	0·0285
12	0·0088	25	0·0139	38	0·0260
13	0·0109	26	0·0350	39	0·0219
14	0·0201	27	0·0770	40	0·0173

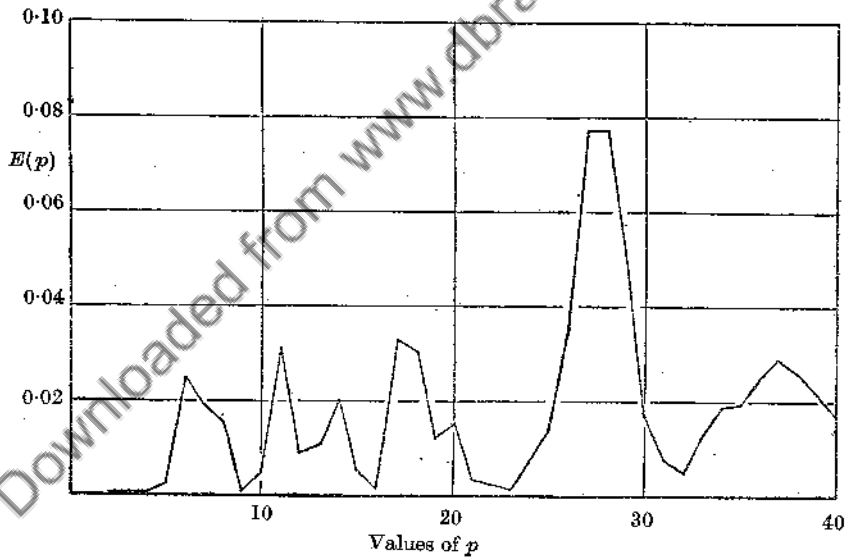

Fig. 4·6. Periodogram of the sum of series 2 and 3

4·7. The experiment was repeated with series 3, the ordinates being given in Table 4·5 and the periodogram in Fig. 4·5. This series has a m.d. (peaks) of 5·52, a m.d. (upcrosses) of 6·21 and an autoregressive period of 6·92. This time the periodogram is nearer to the kind of figure we are entitled to expect from a reliable guide. It has a major peak at 7, but there are still minor peaks at 11, 17, 20, 23 and 33.

4·8. It is noticeable that the periodogram of series 3, which is the least damped, is the least misleading, whereas that of series 2 for which the damping is greatest is the most misleading. General considerations would also lead us to the conclusion, supported by these experiments, that the greater the damping the greater the effect of the disturbance function and hence the greater the liability to spurious peaks in the periodogram due to chance effects.

4·9. The periodogram of the sum of series 2 and 3 is shown in Table 4·6 and Fig. 4·6. There are now appreciable peaks at 6, 11, 17, 28 and 37. The general conclusion of unreliability is confirmed.

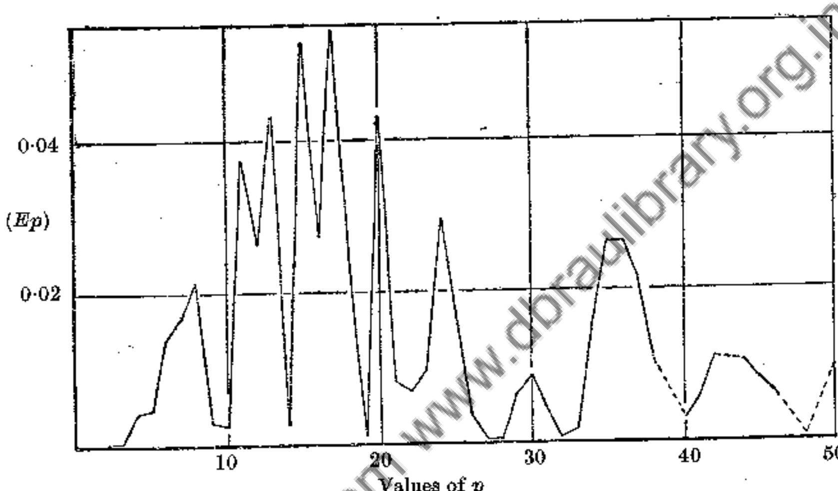

Fig. 4·7. Periodogram of the Beveridge series for integral trial periods up to 50

4·10. With these results in mind we may reconsider the periods proposed by Sir William Beveridge after an exhaustive analysis of the data of Table 1·6. In Fig. 4·7 I show the periodogram drawn for integral trial periods from 2 to 50. Sir William actually carried the analysis as far as 80 and examined a great many non-integral values. In consequence he suggested 19 periods, whereas our diagram reveals only 10. However, for purposes of comparison with the results already obtained, our figure is probably the better.

In fact, the general appearance of Fig. 4·7 is quite like those obtained from the artificial series. There is the same serrated course along the periodogram and the same occasional appearance of marked peaks. It seems to me quite evident that Sir William's periods could be just as unreal as those found in the artificial series. The collateral evidence points in the same direction. For instance, the correlogram (Fig. 3·7) affords strong evidence of a heavily damped series and shows only two periods—which, by the way, coincide with two of Sir William's 19. One can never prove mathematically that the series is not the sum of 19 harmonics and a random element. Mathematically it may be so. But it seems to me that the overwhelming balance of evidence is against this multiplicity of periods. Those who cling to the reality of their existence are under an obligation to bring forward much more convincing evidence before the statistician can regard them as even probable.

Summary

4·11. (1) The periodogram of an autoregressive series may give quite misleading results by failing to show existing oscillatory effects and, what is worse, by indicating the existence of significant effects where none exists.

(2) It seems that the heavier the damping in the series, the more likely are spurious effects to appear.

(3) Where, therefore, serial correlations or other evidence indicate damping in the dependence of successive terms in the series periodogram analysis is dangerous. I doubt whether in the majority of cases it is worth the labour of computation.

CHAPTER 5

VARIATE-DIFFERENCES OF AUTOREGRESSIVE SERIES

5·1. We shall require some expressions for the theoretical values of reduced variances derived from variate-differences of an autoregressive series of type

$$u_{t+2}+au_{t+1}+bu_t = \epsilon_{t+2}. \tag{5·1}$$

The solution of this equation is

$$u_t = \sum_{j=0}^{\infty} \xi_j \epsilon_{t-j+1}, \tag{5·2}$$

where

$$\xi_j = \frac{2}{\sqrt{(4b-a^2)}} p^j \sin \theta j,$$

and hence $\xi_0 = 0$, $\xi_1 = 1$. The quantities p and θ are defined in (3·3) and (3·4). We then have

$$u_t = \xi_0 \epsilon_{t+1} + \xi_1 \epsilon_t + \xi_2 \epsilon_{t-1} + \xi_3 \epsilon_{t-2} + \ldots,$$

$$u_{t+1} = \xi_0 \epsilon_{t+2} + \xi_1 \epsilon_{t+1} + \xi_2 \epsilon_t + \xi_3 \epsilon_{t-1} + \ldots,$$

and hence

$$\begin{aligned} \Delta u_t &= u_{t+1} - u_t \\ &= \xi_0 \epsilon_{t+2} + (\xi_1 - \xi_0)\epsilon_{t+1} + (\xi_2 - \xi_1)\epsilon_t + \ldots. \end{aligned} \tag{5·3}$$

Thus, for the infinite series

$$\frac{\operatorname{var} \Delta u_t}{\operatorname{var} \epsilon} = \xi_0^2 + (\xi_1 - \xi_0)^2 + (\xi_2 - \xi_1)^2 + \dots$$

$$= \xi_0^2 + \sum_{j=0}^{\infty} (\Delta \xi_j)^2. \tag{5·4}$$

In a similar way we find

$$\frac{\operatorname{var} \Delta^2 u_t}{\operatorname{var} \epsilon} = \xi_0^2 + (\xi_1 - 2\xi_0)^2 + \sum_{j=0}^{\infty} (\Delta^2 \xi_j)^2, \tag{5·5}$$

and generally

$$\frac{\operatorname{var} \Delta^k u_t}{\operatorname{var} \epsilon} = \xi_0^2 + (\xi_1 - k\xi_0)^2 + \left\{\xi_2 - k\xi_1 + \binom{k}{2}\xi_0\right\}^2 + \dots$$

$$+ \left\{\xi_{k-1} - \binom{k}{1}\xi_{k-2} + \binom{k}{2}\xi_{k-3} \dots \pm \binom{k}{k-1}\xi_0\right\}^2 + \sum_{j=0}^{\infty} (\Delta^k \xi_j)^2. \tag{5·6}$$

We proceed to evaluate the last term on the right. Consider the first difference of

$$\zeta_t = p^t \sin(\theta t + \alpha). \tag{5·7}$$

We have

$$\Delta \zeta_t = p^t \{p \sin(\theta t + \alpha + \theta) - \sin(\theta t + \alpha)\}$$

$$= p^t \beta \sin(\theta t + \alpha + \phi), \tag{5·8}$$

where

$$\beta \cos \phi = p \cos \theta - 1 = \frac{a}{2} - 1$$

$$\beta \sin \phi = p \sin \theta = \sqrt{\left(b - \frac{a^2}{4}\right)},$$

and hence

$$\beta = \sqrt{(1 + a + b)}, \tag{5·9}$$

$$\tan \phi = -\frac{\sqrt{(4b - a^2)}}{a + 2}. \tag{5·10}$$

It follows by a repetition of the process that

$$\Delta^k \zeta_t = p^t \beta^k \sin(\theta t + k\phi + \alpha). \tag{5·11}$$

Hence

$$\sum_{j=0}^{\infty} (\Delta^k \xi_j)^2 = \frac{4\beta^{2k}}{4b - a^2} \left\{\sum_{j=0}^{\infty} p^{2j} \sin^2(\theta j + k\phi)\right\},$$

which, on summation of the trigonometrical series by usual methods, reduces to

$$\frac{2\beta^{2k}}{4b - a^2}\left\{\frac{1}{1-b} - \frac{\cos 2k\phi - b\cos(2k\phi - 2\theta)}{(1+b)^2 - a^2}\right\} = \frac{2\beta^{2k}}{4b - a^2}\left\{\frac{1}{1-b} + \frac{\sin(2k\phi - \chi)}{\{(1+b)^2 - a^2\}^{\frac{1}{2}}}\right\}, \tag{5·12}$$

where

$$r \sin \chi = 1 - b \cos 2\theta = 1 + b - \tfrac{1}{2}a^2,$$

$$r \cos \chi = b \sin 2\theta = -\frac{a}{2}\sqrt{(4b - a^2)},$$

$$r^2 = (1 + b)^2 - a^2,$$

and hence

$$\tan \chi = -\frac{2(1+b) - a^2}{a\sqrt{(4b - a^2)}}. \tag{5·13}$$

This gives us the required sum in (5·6). The other terms can be evaluated from the lower values of ξ and hence $\text{var}(\Delta^k u_t)$ is ascertainable. For the reduced variance used in variate-difference analysis we then have

$$V_k = \frac{\text{var}(\Delta^k u_t)}{\binom{2k}{k}}. \tag{5·14}$$

5·2. To illustrate the calculations let us work out the values of V_k for series I up to $k = 20$.

In this case $a = -1·1$, $b = 0·5$. Hence, from (5·9),

$$\beta = \sqrt{(1+a+b)} = \sqrt{(0·4)} = 0·632,456. \tag{5·15}$$

Also $$\beta\cos\phi = -\frac{a+2}{2} = -0·45.$$

Hence $$\cos\phi = -0·7115, \quad \phi = 135·35°. \tag{5·16}$$

From (5·13) $$r^2 = 2·25 - 1·21 = 1·04, \quad r = 1·0198,$$

$$r\cos\chi = \frac{1·1}{2}\sqrt{(2-1·21)} = 0·48885,$$

$$\cos\chi = 0·4794, \quad \chi = 61·36°. \tag{5·17}$$

As a check in the first instance, let us find the value of $\text{var}(\Delta^0 u_t)$, i.e. the variance of the series itself. This is known to be [10]

$$\frac{1+b}{1-b\{(1+b)^2-a^2\}}\text{var}\,\epsilon = 2·8846\,\text{var}\,\epsilon. \tag{5·18}$$

From (5·12) with $k = 0$ we find, the other terms in (5·6) vanishing in this case,

$$\frac{V_0}{\text{var}\,\epsilon} = \frac{2}{4b-a^2}\left\{\frac{1}{1-b} + \frac{\sin(-\chi)}{\{(1+b)^2-a^2\}^{\frac{1}{2}}}\right\}$$

$$= \frac{2}{0·79}\left\{2 - \frac{\sin 61·36°}{1·0198}\right\}$$

$$= 2·8846, \tag{5·19}$$

which checks with (5·18).

The ξ's obey the relation

$$\xi_{t+2} + a\xi_{t+1} + b\xi_t = 0, \tag{5·20}$$

with the initial conditions $\xi_0 = 0$, $\xi_1 = 1$, and can be calculated from this formula or from

$$\xi_t = \frac{2}{\sqrt{(4b-a^2)}}p^t \sin\theta t. \tag{5·21}$$

They are shown in the first column of Table 5·1 for $a = -1·1$, $b = 0·5$. By writing zeros for values prior to ξ_0 and taking differences we find the terms

$$\xi_{k-1} - \binom{k}{1}\xi_{k-2} + \dots \pm \binom{k}{k-1}\xi_0, \tag{5·22}$$

appearing in (5·6). The successive differences are shown in the table.

TABLE 5·1. Illustration of the calculation of sums of squares of type (5·21) for series 1

Values of k	(1) $\Delta^0\xi_k$	(2) $\Delta^1\xi_k$	(3) $\Delta^2\xi_k$	(4) $\Delta^3\xi_k$	(5) $\Delta^4\xi_k$
0	0	0	0	0	0
1	1	−1	1	−1	1
2	1·1	−0·1	−0·9	1·9	−2·9
3	0·71	0·39	−0·49	−0·41	2·31
4	0·231	0·479	−0·089	−0·401	−0·009
5	−0·1009	0·3319	0·1471	−0·2361	−0·1649
6	−0·22649	0·12559	0·20631	−0·05921	−0·17689
7	−0·198689	−0·027801	0·153391	0·052919	−0·112129
Sum of squares of top $j-1$ terms in column (j)	0	0	1	4·61	14·746

Values of k	(6) $\Delta^5\xi_k$	(7) $\Delta^6\xi_k$	(8) $\Delta^7\xi_k$	(9) $\Delta^8\xi_k$
0	0	0	0	0
1	−1	1	−1	1
2	3·9	−4·9	5·9	−6·9
3	−5·21	9·11	−14·01	19·91
4	2·319	−7·529	16·639	−30·649
5	0·1559	2·1631	−9·6921	26·3311
6	0·01199	0·14391	2·01919	−11·71129
7	−0·064761	0·076751	0·067159	1·952031
Sum of squares of top $j-1$ terms in column (j)	48·732	169·367	606·960	2218·671

TABLE 5·2. Calculation of theoretical reduced variances of differences of series 1, compared with observed values

(1) k	(2) $\Sigma(\Delta^k\xi_j)^2$ from (5·12)	(3) Sum of squares of columns in Table 5·1	(4) Sum of columns (2) and (3) divided by $\binom{2k}{k}$	(5) Column (4) divided by 2·8846	(6) Observed values for series 1
0	2·8846	0	2·8846	1	1
1	1·5388	0	0·7694	0·2667	0·2383
2	1·1540	1	0·3590	0·1245	0·1100
3	0·4053	4·61	0·2508	0·0869	0·0756
4	0·0754	14·746	0·2117	0·0734	0·0627
5	0·038	48·732	0·1935	0·0671	0·0562
6	0·029	169·367	0·1833	0·0635	0·0523
7	0·011	606·960	0·1769	0·0613	0·0497
8	0·002	2218·671	0·1724	0·0598	0·0478
9	0·001	8222·839	0·1691	0·0586	0·0465
10	0·001	30786·012	0·1666	0·0578	0·0454
11	—	116158·9	0·1647	0·0571	0·0446
12	—	440971·7	0·1631	0·0565	0·0439
13	—	1682373·4	0·1618	0·0561	0·0433
14	—	6444862·4	0·1607	0·0557	0·0429
15	—	24774244	0·1597	0·0554	0·0425
16	—	95512813	0·1589	0·0551	0·0421
17	—	369166246	0·1582	0·0548	0·0416
18	—	1430013030	0·1576	0·0546	0·0412
19	—	5550088284	0·1570	0·0544	0·0407
20	—	21577727939	0·1565	0·0543	0·0403

In Table 5·2 are shown the sums of squares of terms like (5·21), the contribution $\Sigma(\Delta^k\xi_j)^2$ for $k = 0$ to 10, and the ratio giving the reduced variance

$$V_k = \frac{\operatorname{var}(\Delta^k u)}{\binom{2k}{k}}$$

as a fraction of the variance of the series. The observed values are shown in the last column.

5·3. We note in the first place that for k greater than 5 or 6 the sum $\Sigma(\Delta^k\xi_j)^2$ contributes practically nothing to the result. The other terms grow so rapidly as to swamp this element. For that reason we have not bothered to evaluate it for $k > 10$.

5·4. Secondly, the observed and theoretical values run concurrently, but the former are rather lower than the latter. To some extent this is due to the fact that our observed series has a rather greater value than the theoretical expectation, 2535·110 as against 2355·769. When allowance is made for this factor the agreement is fair. We might well have expected to find it a good deal worse, for the observed serial covariances are much higher than the theoretical values (3·7), and from the formulae expressing the variate-difference functions in terms of serial covariances we might have expected the former to be seriously affected. That they are not influenced more than observation shows is, I think, attributable to the fact that in equation (2·18) the higher-order serial covariances are of less weight than the lower-order quantities.

5·5. An interesting feature of Table 5·2 is the peculiar slow downward creep in the coefficients V after $k = 5$. It occurs in both the theoretical and the observed figures. The first two or three differences reduce the variance substantially, but after that successive differencing has an ever-decreasing effect. Before proceeding to discuss this phenomenon I give some further experimental results.

5·6. Table 5·3 shows, for series 2, 3 and 4, the values of V_k up to $k = 20$. The values for series 2 and 3 are of the same type as those for series 1 and require no separate discussion. In series 4 we have a characteristic variate-difference effect for V_1 to V_6, the very short oscillations in this series (and the negative value of ρ_1) resulting in greater values for V's of higher order than for V_0. Once this is exhausted the values of V creep steadily downwards in the characteristic way.

5·7. Compare the effects for the theoretical series with those of Table 5·4, which I quote from Tintner's book [16] on the variate-difference method. Again we find the characteristic creep of later V's after a sudden fall for the first two or three, in annual wool prices, monthly wool prices, annual raw-silk prices, and annual wheat-flour prices—all the series that Tintner examines. These series, incidentally, contain only 48 terms, except the monthly wool price figures which are presumably 12 times as numerous.

TABLE 5·3. Reduced variances of differences V_k for the experimental series 1, 2, 3 and 4

k	V_k			
	Series 1	Series 2	Series 3	Series 4
0	2535·11	3414·42	3900·94	2001·20
1	604·10	523·87	1472·92	3377·47
2	278·79	239·46	523·63	3958·74
3	191·57	183·95	256·33	4241·73
4	158·83	163·17	172·17	4377·38
5	142·49	152·49	142·17	4425·47
6	132·56	146·06	128·52	4413·85
7	125·91	141·96	121·12	4360·24
8	121·23	139·38	116·55	4279·37
9	117·79	137·74	113·43	4184·33
10	115·14	136·59	111·17	4086·08
11	113·02	135·68	109·45	3992·56
12	111·31	134·90	108·09	3907·91
13	109·89	134·22	106·92	3832·62
14	108·72	133·64	105·84	3764·74
15	107·67	133·19	104·81	3701·48
16	106·65	132·91	103·85	3640·63
17	105·57	132·81	103·00	3581·18
18	104·43	132·86	102·33	3523·33
19	103·26	133·03	101·83	3467·97
20	102·14	133·23	101·51	3415·88

TABLE 5·4. Reduced variances of differences V_k for four series of annual wool prices, monthly wool prices, annual raw-silk prices and annual wheat-flour prices (from Tintner [16]). The period covered is 1890–1937 in each case

k	V_k			
	Annual wool prices	Monthly wool prices	Annual raw-silk prices	Annual wheat-flour prices
0	0·1069	0·1121	2·8914	4·7969
1	0·02769	0·002054	0·3849	0·7020
2	0·02590	0·001237	0·2714	0·4402
3	0·02624	0·001096	0·2317	0·3931
4	0·02638	0·001029	0·2033	0·3767
5	0·02626	0·000979	0·1824	0·3662
6	0·02600	0·000945	0·1662	0·3548
7	0·02577	0·000922	0·1533	0·3501
8	0·02553	0·000905	0·1422	0·3426
9	0·02519	0·000892	0·1324	0·3334
10	0·02487	0·000884	0·1249	0·3243

5·8. Let us return to the experimental series, the only ones of whose generating function we are certain. In Table 5·2 the theoretical reduced variances for successive differences seem to be tending to a limit in the neighbourhood of 5·4 % of the variance of the primary series, and the observed variance to about 4 %. Whether this limit has any reality or not, it certainly appears to have for the finite series under examination.

In applying the variate-difference method to series 1 in the standard way we should be led to the conclusion that it can be *represented* as a systematic effect expressible locally as a polynomial together with a superposed random element whose variance is about 5 % of the whole. Such a scheme of representation, whatever its value for smoothing purposes, is obviously very wide of the truth. In short, the variate-difference method does not give us a clue to the true nature of the series. It provides for the autoregressive scheme an alternative and approximate form of mathematical representation, but seems to draw no distinction between autoregression and a functional scheme with superposed random variations.

5·9. Nor does it appear that the slow decline in reduced variances which we have found in the analysis of autoregressive series is necessarily characteristic of such series. It is well known that successive differences tend to become highly correlated and that the correlation of V_k and V_{k+1} tends to unity for high k. We might therefore expect that a series of values of V_k would be smooth. Experimental evidence in confirmation was provided by Anderson [**1**], who analysed 320 terms of a random series subdivided into series of 32 sets of 10, 16 sets of 20, 8 sets of 40, 4 sets of 80, two sets of 160 and as a whole. His results show a smooth course of values of successive V's, but they do not all decline. Some increase, some oscillate and some decrease.

5·10. The mere regularity in the course of an observed series of V's therefore proves very little, and it certainly does not provide presumptive evidence one way or the other in respect of the autoregressive character of the series. It may be that the *downward* movement of the V's is more typical. One would, for instance, not expect it on every occasion if the residuals were random. If the course of the V's were not downwards we might, in the present state of knowledge, doubt the existence of autoregression or at least require to examine the system further before accepting an autoregressive hypothesis. But if the course is downward all we can say is that the hypothesis is not weakened. As in other cases (e.g. that of mean-intervals between peaks discussed in the next chapter) our difficulty is that several different hypotheses may give the same results and it is difficult to arrive at a crucial experiment with the variate-difference method.

Summary

5·11. (1) Variate-differencing of a linear autoregressive series may yield a set of variances which appear to tend to a limit. This must not be held to suggest that the primary series is necessarily generated by a polynomial element plus a superposed random element.

(2) The successive variances decrease slowly but systematically in a way which may be characteristic of the autoregressive scheme but has not yet been shown to be so.

CHAPTER 6

RUNS AND SEQUENCES IN AUTOREGRESSIVE SERIES

Mean-distance between peaks

6·1. We may consider the generating equation of a simple linear autoregressive series

$$u_{t+2} + au_{t+1} + bu_t = \epsilon_{t+2} \tag{6·1}$$

as a difference equation. Its solution, apart from a complementary function which damps out of existence very rapidly and may be neglected, is, as noted in **5·1**,

$$u_t = \sum_{j=0}^{\infty} \xi_j \epsilon_{t-j+1}, \tag{6·2}$$

where

$$\xi_j = \frac{2}{\sqrt{(4b-a^2)}} p^j \sin \theta j, \tag{6·3}$$

and hence $\xi_0 = 0$, $\xi_1 = 1$. Somewhat similar formulae apply to the more extended linear scheme

$$u_{t+m} + a_1 u_{t+m-1} + \ldots + a_m u_t = \epsilon_{t+m}, \tag{6·4}$$

in the sense that the solution u_t is a sum of terms of type (6·2), the ξ's being more general damped harmonic terms depending on the constants of the series.

6·2. It follows that if ϵ is a random normal variable then u_t itself is normally distributed. In the theoretical part of this chapter I shall assume this to be so. The experimental series are based on a rectangular distribution, but it appears that the theoretical results for normal variation are followed very closely by those for rectangular variation. Part, if not the whole, of the explanation of this concordance is due to the fact that the average of a number of random variables tends to normality under certain general conditions. Those conditions are, in fact, not strictly fulfilled in our present case, but doubtless the effect is somewhat similar.

6·3. If u_2 is a peak of the series

$$u_1 \leqslant u_2 \geqslant u_3. \tag{6·5}$$

Let us take new variables

$$\lambda_1 = u_2 - u_1, \quad \lambda_2 = u_3 - u_2. \tag{6·6}$$

Then λ_1 and λ_2 are normally distributed with variance

$$\operatorname{var} \lambda = 2(1-\rho_1) \operatorname{var} u,$$

ρ_1 being the first autocorrelation of the series. The correlation between λ_1 and λ_2 is easily seen to be τ_1 say, where

$$\tau_1 = \frac{-1+2\rho_1-\rho_2}{2(1-\rho_1)}$$

$$= -\frac{\Delta^2 \rho_0}{2(1-\rho_1)}. \tag{6·7}$$

Now the probability that a pair of normal variables with correlation τ_1 obey the relation $\lambda_1 \geqslant 0, \lambda_2 \leqslant 0$ is the volume of the bivariate normal surface lying in the quadrant $\lambda_1 \geqslant 0, \lambda_2 \leqslant 0$. This, by a theorem due to W. F. Sheppard, is $\frac{1}{2\pi}\cos^{-1}\tau_1$. (A proof of this result which is capable of generalization is given below.)

It follows that the mean-distance between peaks in a normal series whose first two autocorrelations are ρ_1 and ρ_2 is

$$\text{m.d. (peaks)} = \frac{2\pi}{\cos^{-1}\tau_1}, \tag{6·8}$$

where τ_1 is given by (6·7). This is true whether the series is autoregressive or not.

As a check we note that if $\rho_1 = \rho_2 = 0$, $\tau_1 = \frac{2\pi}{3}$, and the m.d. (peaks) is 3, the known result for a random series.

6·4. For the autoregressive scheme

$$u_{t+2} + au_{t+1} + bu_t = \epsilon_{t+2},$$

we have, multiplying by u_{t-k} and summing,

$$\rho_{k+2} + a\rho_{k+1} + b\rho_k = \frac{\Sigma(\epsilon_{t+2}u_{t-k})}{\text{var}\, u}. \tag{6·9}$$

Now from (6·2) it is seen that u_{t-k} does not contain ϵ_{t+2} if k is not less than -1. The sum on the right of (6·9) thus vanishes and we have

$$\rho_{k+2} + a\rho_{k+1} + b\rho_k = 0 \quad (k \geqslant -1). \tag{6·10}$$

In particular, for $k = -1, 0$ we have

$$\rho_1 + a + b\rho_1 = 0, \quad \rho_2 + a\rho_1 + b = 0,$$

whence

$$\rho_1 = -\frac{a}{1+b}, \tag{6·11}$$

$$\rho_2 = \frac{a^2 - b(1+b)}{1+b}. \tag{6·12}$$

Hence, from (6·7), we find

$$\tau_1 = \frac{b^2 - (1+a)^2}{2(1+a+b)}, \tag{6·13}$$

and the m.d. (peaks) may be found from (6·8).

For the four experimental series we have considered above the theoretical and observed values were as follows:

Series	m.d. (peaks) from (6·8)	m.d. (peaks) observed
1	4·96	5·05
2	4·96	5·57
3	5·69	5·52
4	2·60	2·69

The agreement is good except perhaps for series 2, which we have already iced as diverging more from expectation than the others.

6·5. It is rather remarkable that the m.d. (peaks) as given by (6·8) and (6·13) for the simple linear autoregressive series is extraordinarily insensitive to variations in a and b. Some typical values are:

$a = -1\cdot2$	$b = 0\cdot4$	m.d. (peaks)	$= 4\cdot96$
$a = -1\cdot5$	$b = 0\cdot8$	,,	$= 5\cdot69$
$a = -1\cdot0$	$b = 0\cdot6$	,,	$= 4\cdot96$
$a = -0\cdot8$	$b = 0\cdot8$	,,	$= 5\cdot13$

It appears, in fact that, for a wide range of values of ρ_1 and ρ_2 we shall find a m.d. (peaks) of between 4 and 6 units. The appearance of such a 'period' therefore throws very little light on the nature of the series, except perhaps that it may be held to confirm the existence of random disturbances.

6·6. Although I do not want to lay myself open to the charge of indulging in the sort of numerology which I am always criticizing in others, I think it may be of interest to observe that a table quoted by Davis ([6], p. 548) gives the distribution of 166 'cycles' in 17 countries, presumably calculated by distances between peaks or troughs, the mean value of which is 5·2 years. In my mind, this does not by any means imply that there is any kind of rhythmic influence at work generating business oscillations with a mean period of about 5 years. It would be quite consistent with the observations to suppose that the economic structure in the various countries was capable of representation by autoregressive models with perhaps quite different constants in the different countries.

6·7. Sir William Beveridge [4] has recently published some interesting series of British industrial activity going back to 1785. He deduces from his main series* a period of about 8 years by counting the occurrence of what he regards as the principal peaks; but to reach this result he has to ignore a number of minor ones. Had he included them all he would have found a 'period' of 4 years, precisely as the foregoing comments lead us to expect.

6·8. Now is it legitimate to exercise an individual judgment in the rejection of minor peaks in this way? Table 6·1 shows the distribution of values of the 30 peaks of the series. Sir William has rejected exactly half the peaks (and hence doubled the mean-distance between peaks). The line of division is far from clear for Sir William has accepted one peak in the range 100, but rejected seven with greater values. The maximum number of acceptances, in the range 104–108 is also the region of maximum rejections. The general run of values in the final column shows how difficult it is to decide, subjectively or objectively, on acceptance by relation to the peak values alone.

* Sir William's index for all industries runs from 1785 to 1938. I have worked on the series from 1785 to 1913 as there is a gap of 5 years, 1914–19, for which no figures are available.

6·9. The difficulty obviously becomes intensified if we allow ourselves to be influenced by other factors such as the existence of neighbouring peaks in the series. There will be a tendency to regard two peaks which are close together as corresponding to a single oscillatory maximum and hence to reject one of them. There is no justification for such a course, so far as I can see; and if it is pursued to any extent there will result too few short intervals and an excessively long m.d. (peaks), which is just the kind of thing we do observe in many inquiries.

TABLE 6·1. Numbers of peaks with given values in the Beveridge series of industrial activity in Britain, 1785–1913. (Data from reference [**4**], pp. 310–313.)

Values of series	No. of peaks counted by Sir William Beveridge	Other peaks	Total peaks
94–	—	1	1
96–	—	2	2
98–	—	3	3
100–	1	2	3
102–	—	—	—
104–	2	4	6
106–	5	3	8
108–	1	—	1
110–	2	—	2
112–	2	—	2
114–	1	—	1
...	—	—	—
132–	1	—	1
Totals	15	15	30

Mean-distance between upcrosses

6·10. By a similar line of reasoning to that used above in paragraph **6·3** it will be seen that the mean-distance between upcrosses in any normal series is given by

$$\text{m.d. (upcrosses)} = \frac{2\pi}{\cos^{-1}\rho_1}, \tag{6·14}$$

where ρ_1, the first autocorrelation, is given for the autoregressive case by (6·11). In general we expect that m.d. (upcrosses) is greater than m.d. (peaks) because of fluctuations which may take place below or above the zero axis, providing peaks but adding nothing to the upcrosses. For our four experimental series we have:

Series	m.d. (upcrosses) from (6·14)	m.d. (upcrosses) observed	m.d. (peaks) observed
1	8·40	8·30	5·05
2	11·61	12·39	5·57
3	6·87	6·38	5·52
4	2·73	2·76	2·69

The agreement appears to be satisfactory.

6·11. In the Beveridge series of Table 1·6 the m.d. (peaks) is 4·67 years and the m.d. (upcrosses) 7·08 years. This is consistent with a scheme of autoregression.

6·12. I do not propose on this occasion to discuss which is the 'best' measure of oscillatory mean-period in an autoregressive scheme, but it is worth while pointing out from the observed differences between the m.d. (upcrosses) and m.d. (peaks), together with the differences of both from the autoregressive period, how necessary it is to be exact in defining one's usage. It may also be added that if an observed series has a trend present and that trend is removed by fitting polynomials, it is quite possible that part of the oscillatory effect will be incorporated into the trend effect, which will shorten the m.d. (upcrosses) but may not affect the m.d. (peaks) so greatly. I shall discuss the question of trend elimination on a future occasion; but I mention the point now as it provides a possible explanation of the failure on the part of most writers to stress the difference between the two measures.

Distribution of runs and intervals

6·13. The mean-distances are only central values of distributions, and, as has been seen, are insensitive to changes in the nature of the generating process. It is, however, possible that a more sensitive discriminator between different types of series, or series of the same type with different constants, could be found in the *distribution* of intervals. The method has not been used hitherto because the distributions were not known except in the case when the series is random. (For an account of what is known of the random case, with a list of references, see Wallis and Moore [17].) I have obtained expressions which provide a mathematical solution of the distribution problem, and give them below; but unless they can be simplified arithmetical application would be rather tedious.

6·14. Consider in the first place 'upruns' in the series, that is to say, sequences in which each value is not less than the preceding value. We note that if

$$\lambda_j = u_{j+1} - u_j = \Delta u_j \tag{6·15}$$

the sequence $u_1 \dots u_{k+1}$ is an uprun if all the λ's are not negative. Further, all the λ's have variance $2(1-\rho_1)\,\mathrm{var}\,u$ and their serial correlations are given by

$$\begin{aligned} r_j &= \frac{-\rho_{j-1} + 2\rho_j - \rho_{j+1}}{2(1-\rho_1)} \\ &= \frac{-\Delta^2 \rho_{j-1}}{2(1-\rho_1)}. \end{aligned} \tag{6·16}$$

The probability that a given sequence of $k+1$ terms is an uprun is therefore equivalent to the relative frequency in the joint distribution of $\lambda_1 \dots \lambda_k$ in the non-negative part of the domain. Our problem then reduces to that of finding the content of a k-dimensional normal hypersurface contained in the positive hyperquadrant $\lambda_1 \geqslant 0, \ldots, \lambda_k \geqslant 0$.

I have given the solution of an analogous problem in a previous note [7] in the form of an expansion in ascending powers of the correlations. The expansion is, in fact, a generalization of the tetrachoric series.

For instance, if $k = 2$ we have the tetrachoric series ([13], p. 356)

$$\int_h^\infty \int_k^\infty dF = \sum_{j=0}^{\infty} \{H_{j-1}(h)\,\alpha(h)\,H_{j-1}(k)\,\alpha(k)\}\frac{\tau_1^j}{j!}, \tag{6·17}$$

where

$$\alpha(x) = \frac{1}{\sqrt{(2\pi)}}\exp\left(-\tfrac{1}{2}x^2\right), \tag{6·18}$$

$$\alpha(x)\,H_r(x) = \left(-\frac{d}{dx}\right)^r \alpha(x), \tag{6·19}$$

and by convention the first term is

$$\frac{1}{2\pi}\int_h^\infty \exp\left(-\tfrac{1}{2}x^2\right)dx \int_k^\infty \exp\left(-\tfrac{1}{2}y^2\right)dy. \tag{6·20}$$

In our case the limits h and k are zero and since

$$\left.\begin{aligned} H_r(0) &= 0, \quad r \text{ odd} \\ H_r(0) &= \frac{(-1)^j\,(2j)!}{2^j\,j!}, \quad r \text{ even} = 2j \end{aligned}\right\} \tag{6·21}$$

the probability required, say P_2, reduces to

$$P_2 = \frac{1}{4} + \frac{1}{2\pi}\sum_{j=1}^{\infty}\left\{\frac{\tau_1^{2j+1}(2j)!}{(2j+1)\,2^{2j}(j!)^2}\right\} \tag{6·22}$$

$$= \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\tau_1$$

$$= \frac{1}{2} - \frac{1}{2\pi}\cos^{-1}\tau_1. \tag{6·23}$$

This checks with the result of Sheppard already used above.

For $k = 3$ we have

$$P_3 = \int_0^\infty \int_0^\infty \int_0^\infty dF$$

$$= \Sigma\left[(-1)^{j+k+l}\frac{\tau_1^j \tau_2^k \tau_1^l}{j!\,k!\,l!}\{H_{j+l-1}(0)\,H_{j+k-1}(0)\,H_{k+l-1}(0)\}\,\alpha^3(0)\right]. \tag{6·24}$$

Since of $j+l-1$ and the two similar terms one at least must be odd, the only surviving terms are those for which two of j, k, l are zero and the other one is odd. We then find

$$P_3 = \frac{1}{8} + \frac{1}{4\pi}\left(2\sin^{-1}\tau_1 + \sin^{-1}\tau_2\right)$$

$$= \frac{1}{4} - \frac{1}{4\pi}\left(2\cos^{-1}\tau_1 + \cos^{-1}\tau_2 - \pi\right). \tag{6·25}$$

6·15. Pausing for a moment, we may observe that (6·23) and (6·25) may be obtained by geometrical considerations. In fact, P_2 is $\frac{1}{2}$ less the proportional volume between two hyperplanes of angle $\cos^{-1}\tau_1$; P_3 is $\frac{1}{4}$ less the area of a spherical triangle cut off on the unit sphere in three dimensions by planes whose angles are $\cos^{-1}\tau_1$, $\cos^{-1}\tau_2$, $\cos^{-1}\tau_1$. At one time I had hopes of generalizing this approach by finding the content of a region cut off on the hypersphere in k dimensions by k hyperplanes with given angles of intersection, but the direct geometrical approach baffled me; and from the complexity of the results for $k > 3$ I have now some doubts whether any simple solution exists. It may very well be, however, that recurrence relations exist for expressing the values for k in terms of those for $k-1$ and lower orders.

6·16. For general k and a multivariate normal form with correlations between the jth and kth variate of χ_{jk} the proportional volume in the positive quadrant is

$$\frac{1}{(2\pi)^{k}}\int_0^\infty dx_1 \ldots \int_0^\infty dx_k \int_{-\infty}^\infty \exp\left(-\tfrac{1}{2}t_1^2 - it_1x_1\right) \ldots \int_{-\infty}^\infty \exp\left(-\tfrac{1}{2}t_k^2 - it_kx_k\right)$$
$$\sum_{j=0}^\infty \frac{(-1)^j}{j!}(\chi_{12}t_1t_2 + \chi_{13}t_1t_3 + \ldots + \chi_{jk}t_jt_k + \ldots)^j\, dt_1 \ldots dt_k, \quad (6·26)$$

the expressions in the second lot of integrals being expressible in terms of the Tchebycheff-Hermite polynomials. For instance, with $k = 4$, writing $\tau_1 = \chi_{j,j+1}$, etc., we have

$$P_4 = \Sigma\left[\frac{\tau_1^j\tau_2^k\tau_3^l\tau_1^m\tau_2^n\tau_1^p}{j!\,k!\,l!\,m!\,n!\,p!}H_{j+k+l-1}(0)\,H_{j+m+n-1}(0)\,H_{k+m+p-1}(0)\,H_{l+n+p-1}(0)\,\alpha^4(0)\right], \quad (6·27)$$

$$= \frac{1}{4\pi^2}\Sigma\frac{\tau_1^{j+m+p}\tau_2^{k+n}\tau_3^l(j+k+l-1)!\,(j+m+n-1)!\,(k+m+p-1)!\,(l+n+p-1)!}{j!\,k!\,l!\,m!\,n!\,p!\,2^{j+k+l+m+n+p-2}\left(\frac{j+k+l-1}{2}\right)!\left(\frac{j+m+n-1}{2}\right)!\left(\frac{k+m+p-1}{2}\right)!\left(\frac{l+n+p-1}{2}\right)!}, \quad (6·28)$$

subject only to the condition that the fractions in the denominator must be integral.

6·17. These expressions are hardly as difficult as they look. For a damped autoregressive series the expressions for the P's converge fairly quickly and are, I think, amenable to calculation. At any rate, I have not been able to simplify them to any considerable extent. Tables of tetrachoric functions can, of course, be used in the evaluation of the separate terms.

6·18. From the values of P_k we can easily derive the distribution of upruns. In fact if the relative frequency of a complete uprun of exactly k

intervals is f_k, the relative frequency of upruns of 1 is

$$\sum_{j=1}^{\infty}(jf_j) = P_1, \tag{6·29}$$

that of upruns of 2 is $$\sum_{j=2}^{\infty}\{(j-1)f_j\} = P_2, \tag{6·30}$$

and so on. Thus

$$P_{k+2} - 2P_{k+1} + P_k = \sum_{j=k+2}^{\infty}\{(j-k-1)f_j\} + \sum_{j=k}^{\infty}\{(j-k+1)f_j\} - 2\sum_{j=k+1}^{\infty}\{(j-k)f_j\}$$
$$= f_k,$$

or $$f_k = \Delta^2 P_k. \tag{6·31}$$

6·19. For the purpose of comparing theory with experiment in a given series it would probably be simpler to work with upruns and downruns rather than with intervals between peaks, but the distribution for peaks may, if so desired, be derived from that of runs. In fact, the probability that an interval of extent k is an interval between peaks is the sum of the separate probabilities (1) that the first two terms are a complete downrun and the next a complete uprun, (2) that the first three are a complete downrun and the next $k-2$ a complete uprun, and so on. The required distribution is then obtainable by summing appropriate terms of the distribution of runs.

Distribution of intervals between crosses

6·20. The same methods will yield the distributions of intervals from upcross to downcross and from upcross to upcross. The probability that a series of $k+2$ terms is an interval from upcross to downcross is the probability that $u_1 \leqslant 0$, $u_2 \geqslant 0$, ..., $u_{k+1} \geqslant 0$, $u_{k+2} \leqslant 0$, and may be evaluated from the content of the appropriate quadrant of the multidimensional distribution of the u's themselves; and the probability that an interval is one from upcross to upcross follows as in the previous paragraph.

Summary

6·21. (1) Expressions are given for m.d. (upcrosses) and m.d. (peaks) in a series for which the terms are normally distributed. These agree well with the experimental series, notwithstanding that the latter are based on rectangular variation.

(2) In linear autoregressive series the m.d. (peaks) is very insensitive to changes in the constants of the series and therefore provides a poor discriminator between different series.

(3) It is bad statistical practice to exercise any personal judgment in the selection of peaks as of greater importance than others.

(4) Expressions are given for the determination of distributions of runs and intervals in non-random normal series. It may be that such distributions would provide better discriminators than the mean-distances themselves, but further work on this point is necessary.

CHAPTER 7

SUMMARY AND CONCLUSIONS

7·1. In this chapter I lay aside the statistical and mathematical tools which are necessary for a complete discussion of the behaviour of time-series and attempt to explain the main features of the foregoing work in plain English. This will involve some sacrifice of accuracy in expression, but it has been suggested to me that economists who are interested in time-series without being statisticians would wish to have a summary of the basic ideas in language with which they are more familiar.

7·2. The type of time-series encountered in economic work usually possesses oscillatory features which are undoubtedly not due to chance, by which I mean the fluctuations cannot be regarded as happening entirely haphazardly. If we draw values at random out of a hat in some temporal order, the series so obtained will fluctuate; but each value will bear no relation to the next. This is not what we observe, for in general successive values are dependent to a greater or less extent. In short, there does exist something to be explained as a systematic element of the series.

7·3. The existence of this systematic effect can often be seen at a glance from a graph of the series, but if necessary exact tests of randomness can be applied. Assuming the effect to have been demonstrated, our primary statistical problem is to examine its nature and if possible to formulate the laws of its behaviour. The explanation of those laws in terms of economics is more a matter for the economist than the statistician, though the latter is by no means uninterested. The first stage of his work, however, is to find out precisely what is to be explained.

7·4. The classical method of approach to the problem of formulating the laws of oscillatory series was based on harmonic analysis. The typical oscillator was regarded as the kind of wave which is constantly encountered in the physical sciences and is expressible as a sine or cosine of the time variable. Any ordinary oscillatory movement can be represented as a sum of harmonic elements of this kind, and the object of the analysis was to isolate the constituent elements whose sum composed the series.

7·5. This method has never been found very successful in economic work (or, I may add, in meteorology and geophysics). In fact, it depends fundamentally on one assumption which was noticed in **1·22**, and which is neither plausible in theory nor consonant with observation. It had to be recognized from the outset that one could not expect to fit a sum of harmonics to an observed series *exactly* unless one was to take a fantastic number of them. There thus appeared discrepancies between the mathematical model and the observed series, *and these discrepancies were regarded as comparable to* *errors of observation*. That is, they were supposed to have an instantaneous existence, to occur by some sort of random process and then to disappear without exerting any effect on the future motion of the system.

7·6. Now it seems quite clear that if a disturbance does occur in an economic system its effects do endure to some extent. Sometimes at least they endure for a long time. Disturbances may or may not be random in the sense that they occur according to the laws of chance; but once they have occurred they become just as integral a part of the history of the system as any other constituent element.

7·7. A possible model of a series behaving in this way (though not the only one and perhaps not even the best one) is provided by the autoregressive series proposed by Udny Yule. This allows for the incorporation of a stream of disturbances into the system, and there is a considerable volume of evidence to show that it does provide at least an approximation to observed phenomena in economics, meteorology and geophysics. The object of the work described in the foregoing pages was to discover what sort of results were obtained by applying current techniques for the analysis of time series to the autoregressive series. In order that there should be no doubt about the nature of the series under investigation a number of artificial autoregressive series of considerable length were constructed for the experiment.

7·8. The principal result (Chapter 4) was that periodogram analysis (the classical method of detecting harmonics) broke down completely on the autoregressive series considered. Not only did it fail to indicate the true character of the oscillations but it gave a number of 'significant' periods which were quite spurious. My conclusion is that if there is any reason to suppose that the series under examination is autoregressive periodogram analysis is dangerously misleading and is not worth the arithmetical work involved.

7·9. Secondly, it has been shown (Chapter 6) that the method of counting intervals between peaks or upcrosses in the series may also be misleading. Apart from the tendency on the part of some investigators to 'edit' the series and reject certain effects on their own judgment as unimportant it appears that for many types of simple autoregressive series encountered in practice the mean-distances between peaks (and to a smaller extent between upcrosses) is much the same whatever the constants of the series may be. In short, we are always liable to find these 'periods', and their existence throws very little light on the true nature of the generating scheme. I do not mean, of course, that such effects are on that account unimportant and unworthy of study. But I would infer that comparisons of different series among themselves, arguments purporting to explain why some have longer periods than others, attempts to bolster up an economic hypothesis merely on the existence of such effects, have very little validity, simply because the effects are quite consistent with so many hypotheses.

7·10. Thirdly, it has been shown (Chapter 5) that the variate-difference method is also unreliable in providing an estimate of random variation in an autoregressive series. The method was not, of course, intended for that purpose; but the significance of the result is that for such series it gives an answer which appears to be meaningful but in fact may not be so. Like periodogram analysis, it provides a mathematical representation of the data which may be quite foreign to the nature of the generating process.

7·11. All this, I fear, is rather destructive. If the conclusions are correct, much of the work which has been done on the analysis of oscillatory series will have to be reconsidered. Nevertheless, it is essential to get at the truth in these matters, and if only we can recognize the failure of hitherto accredited techniques something will have been done to clear the way for better methods.

7·12. The fourth principal conclusion (Chapter 3) is that the correlogram, if carefully interpreted, will give a reliable guide to the nature of the series; but that for finite series it fails to damp out according to the expectation for a series of infinite length. This is in accordance with what I believe to be generally true of time-series, that large differences between observation and expectation are more the rule than the exception in series of the length which arises in practice.

7·13. A disadvantage of the correlogram appears to be its insensitivity to more complicated schemes of autoregression. Departures from expectation, so to speak, submerge the finer variations which we should expect, for an infinite series, to permit of the dissection of different autoregressive elements in the primary series. From this viewpoint the correlogram provides a kind of lower limit to the oscillatory effects, but may not exhibit with sufficient clarity the different elements which compose them.

7·14. How far these conclusions will be accepted by statisticians and economists I cannot foresee; but I hope that on one point at least there will be no dissent; the necessity for a great deal of further research. Although the work described herein has taken up more of my time than I care to think about I have done no more than scratch the surface of this, perhaps the most difficult and certainly one of the most important of statistical problems.

REFERENCES

[1] ANDERSON, O. (1926/1927). Über die Anwendung der Differenzen-Methode (Variate-Difference Method) bei Reihenausgleichungen, Stabilitätsuntersuchungen und Korrelationsmessungen. *Biometrika*, **18**, 293 and **19**, 53.

[2] BEVERIDGE, SIR WILLIAM H. (1921). Weather and harvest cycles. *Econ. J.* **31**, 429.

[3] BEVERIDGE, SIR WILLIAM H. (1922). Crop yields and rainfall in western Europe. *J.R. Statist. Soc.* **85**, 412.

[4] Beveridge, Sir William H. (1944). *Full Employment in a Free Society.* London: Allen and Unwin.

[5] Burger, H. C. and van Cittert, P. H. (1933). Wahre und scheinbare Intensitätsverteilung in Spektrallinien. *Z. Phys.* **79**, 722 and **81**, 428.

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

[7] Kendall, M. G. (1941). Proof of relations connected with the tetrachoric series and its generalisation. *Biometrika,* **32**, 196.

[8] Kendall, M. G. (1943). *The Advanced Theory of Statistics,* **1**. London: Charles Griffin and Co.

[9] Kendall, M. G. (1944). Oscillatory movements in English agriculture. *J.R. Statist. Soc.* **106**, 91.

[10] Kendall, M. G. (1944). On autoregressive time series. *Biometrika,* **33**, 105.

[11] Kendall, M. G. (1945). On the analysis of oscillatory time series. *J.R. Statist. Soc.* (in the Press).

[12] Kendall, M. G. and Babington Smith, B. (1939). *Tables of Random Sampling Numbers.* Tracts for Computers, no. 24. Cambridge University Press.

[13] Martindale, J. (1941). A correlation periodograph for the measurement of periods in disturbed wave forms. *Shirley Inst. Mem.* **18**, no. iv.

[14] Schumann, T. E. W. and Hofmeyer, W. L. (1942). The problem of autocorrelation of meteorological time series, illustrated by examples from some South African Stations. *Quart. J. Roy. Met. Soc.* **68**, 177.

[15] Schuster, Sir A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terr. Mag.* **3**, 13.

[16] Tintner, G. (1941). *The Variate Difference Method.* Indiana: Bloomington Press.

[17] Wallis, W. A. and Moore, G. H. (1941). A significance test for time series. *Tech. Pap. Nat. Bur. Econ. Res., Wash.*, no. 1.

[18] Walker, Sir Gilbert (1931). On periodicity in series of related terms. *Proc. Roy. Soc.* A, **131**, 518.

[19] Whittaker, E. T. and Robinson, G. (1941). *The Calculus of Observations,* 3rd ed. Blackie and Sons.

[20] Wold, H. (1938). *A Study in the Analysis of Stationary Time Series.* Uppsala.

[21] Yule, G. Udny (1921). On the time correlation problem. *J.R. Statist. Soc.* **84**, 497.

[22] Yule, G. Udny (1926). Why do we sometimes get nonsense correlations between time series? *J.R. Statist. Soc.* **89**, 1.

[23] Yule, G. Udny (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Philos. Trans.* A, **226**, 267.

APPENDIX

Tables for periodogram and harmonic analysis

In carrying out the analysis described in Chapter 4 I found that existing tables were insufficient and had to construct my own. These have subsequently been extended to cover trial periods up to and including 100 and are given below. Notwithstanding the conclusion in the text that periodogram analysis is misleading there may be occasions when it will be required, e.g. for experimental purposes; and in any case there are fields where harmonic analysis is still necessary. I hope, therefore, that the tables may be useful.

The tables give values of $\sin\dfrac{2k\pi}{p}$ and $\cos\dfrac{2k\pi}{p}$ for integral values of p from 5 to 100 and integral values of k from 1 to $\frac{1}{4}(p-4)$ when p is a multiple of 4, from 1 to $\frac{1}{4}(p-2)$ when p is a multiple of 2 but not of 4, and from 1 to $\frac{1}{2}(p-1)$ when p is odd. The tables are to four places, and both decimal points and signs are omitted. The signs are best dealt with under a systematization of the calculations described below.

For periodogram analysis the sums U of the Buys-Ballot table of (1·1) are first ascertained. The next stage is the calculation of the sums A and B of equations (1·2) and (1·3); and it is at this point that the tables are required. If p is a multiple of 4 the terms $\sin\dfrac{2\pi k}{p}$ repeat themselves four times owing to the symmetry of the sine function, and, moreover, the cosine terms are the same as the sine terms of complementary angles, so that it is unnecessary to tabulate both sine and cosine. If p is a multiple of 2 but not of 4, or is odd, separate columns are required for sine and cosine. In the former case, owing to symmetry, only the values in the first quadrant need be tabulated; in the latter case those in the first two quadrants are required. A few examples will illustrate the arithmetical process.

(*a*) *p even.* The following are the values of the U's for series 2 and $p = 16$: 44, 7, −124, −99, −14, −226, −256, −333, −337, −235, −18, −93, −104, −34, −158, −179. These sums are based on 224 terms. We write them down in four columns a, b, c, d, as follows:

a	b	c	d	$a+b-c-d$	$a-b-c+d$
44	—	−337	—	—	381
7	−333	−235	−179	88	396
−124	−256	−18	−158	−204	−8
−99	−226	−93	−34	−198	186
—	−14	—	−104	90	—

The method of writing down will be clear on examination. The first column runs downwards; we then cross to the second column, start a line lower down and write upwards, stopping at the second row; cross again to the third column and write downwards, and again to the fourth column writing upwards.

We then form the two columns on the right. In the first row we have a blank in the first column and $a-c$ in the second; in the last row $b-d$ in the first and blank in the second; and in intermediate rows $a+b-c-d$ in the first and $a-b-c+d$ in the second.

In conjunction with the values for $p = 16$ in Table 1, we then find A' and B' (the sums $\Sigma U_{j+1} \cos \dfrac{2\pi j}{p}$ and $\Sigma U_{j+1} \sin \dfrac{2\pi j}{p}$, i.e. the function A and B except for the constant $2/mp$) as

$$\begin{aligned} A' &= 381+396(\cdot 9239)-8(\cdot 7071)+186(\cdot 3827) \\ &= 812\cdot 390, \\ B' &= 88(\cdot 3827)-204(\cdot 7071)-198(\cdot 9239)+90 \\ &= -203\cdot 503, \end{aligned}$$

giving $$A'^2+B'^2 = 701{,}390\cdot 983.$$

Hence, the value of $2/(mp)$ being $2/224$ we have

$$I = \left(\frac{2}{mp}\right)^2 (A'^2+B'^2) = 55\cdot 9145.$$

It is left to the reader to verify that this process gives the required intensity. To reduce to the standardized intensity E we divide by twice the variance of series 2, namely, 6828·844, obtaining $E(16) = 0\cdot 0082$ as shown in Table 4·4.

(*b*) *p even but not a multiple of* 4. The following are the sums U for series 2 and $p = 18$: -25, -84, -38, -157, -61, -168, -19, -37, -125, -96, -189, -316, -469, -275, -213, -247, -69, 72, the figures being based on 216 terms. We now write the values in four columns as follows:

a	b	c	d	$a+b-c-d$	$a-b-c-d$
−25	—	−96	—	—	71
−84	−125	−189	72	−92	302
−38	−37	−316	−69	310	246
−157	−19	−469	−247	540	84
−61	−168	−275	−213	259	169

The difference from case (*a*) lies in the direct crossover at the foot of columns a and c.

From the values for $p = 18$ in Table 2 we find A' and B' as

$$\begin{aligned} A' &= 71+302(\cdot 9397)+246(\cdot 7660)+84(\cdot 5000)+169(\cdot 1736) \\ &= 614\cdot 564, \\ B' &= -92(\cdot 3420)+310(\cdot 6428)+540(\cdot 8660)+259(\cdot 9848) \\ &= 890\cdot 507. \end{aligned}$$

Hence

$$I = \left(\frac{2}{mp}\right)^2 (A'^2 + B'^2) = \left(\frac{2}{216}\right)^2 (1{,}170{,}691{\cdot}627)$$

$$= 100{\cdot}3679,$$

and $$E = I/(2 \operatorname{var} u) = 0{\cdot}0147.$$

(*c*) *p odd*. The following are the sums U for $p = 17$ in series 2, the number of terms being 238: -409, -395, -432, -403, -387, -266, 53, 269, 261, 240, 50, -24, -138, -194, -117, -25, -93. We write them down in *two* columns:

a	b	$a-b$	$\pm(a+b)$
−409	—	—	−409
−395	− 93	−302	−488
−432	− 25	−407	−457
−403	−117	−286	−520
−387	−194	−193	−581
−266	−138	−128	404
53	− 24	77	− 29
269	50	219	−319
261	240	21	−501

The third column gives the values of $a-b$. The fourth column gives $a+b$ in the first five rows and $-(a+b)$ in the next four. For general odd p, we take $a+b$ in the first $[\frac{1}{4}p]+1$ rows and $-(a+b)$ in the remainder where $[\frac{1}{4}p]$ is the integral part of $\frac{1}{4}p$.

From Table 3 for $p = 17$ we then find:

$$A' = -409 - 488({\cdot}9325) - 457({\cdot}7390) - 520({\cdot}4457) - 581({\cdot}0923)$$
$$+ 404({\cdot}2737) 4 - 29({\cdot}6026) - 319({\cdot}8502) - 501({\cdot}9830)$$
$$= -2157{\cdot}771,$$
$$B' = -302({\cdot}3612) - 407({\cdot}6737) - 286({\cdot}8952) - 193({\cdot}9957)$$
$$-128({\cdot}9618) + 77({\cdot}7980) + 219({\cdot}5264) + 21({\cdot}1837)$$
$$= -774{\cdot}001.$$

Thus $$E = \frac{1}{2 \operatorname{var} u}\left(\frac{2}{mp}\right)^2 (A'^2 + B'^2)$$
$$= 0{\cdot}0543.$$

TABLE 1. Harmonic analysis for p a multiple of 4

Values of p

k	8	12	16	20	24	28	32	36	40	44	48	52
1	7071	5000	3827	3090	2588	2225	1951	1736	1564	1423	1305	1205
2	—	8660	7071	5878	5000	4339	3827	3420	3090	2817	2588	2393
3	—	—	9239	8090	7071	6235	5556	5000	4540	4154	3827	3546
4	—	—	—	9511	8660	7818	7071	6428	5878	5406	5000	4647
5	—	—	—	—	9659	9010	8315	7660	7071	6549	6088	5681
6	—	—	—	—	—	9749	9239	8660	8090	7557	7071	6631
7	—	—	—	—	—	—	9808	9397	8910	8413	7934	7485
8	—	—	—	—	—	—	—	9848	9511	9096	8660	8230
9	—	—	—	—	—	—	—	—	9877	9595	9239	8855
10	—	—	—	—	—	—	—	—	—	9898	9659	9350
11	—	—	—	—	—	—	—	—	—	—	9914	9709
12	—	—	—	—	—	—	—	—	—	—	—	9927

Table 1 (*continued*)

Values of p

k	56	60	64	68	72	76	80	84	88	92	96	100
1	1120	1045	0980	0923	0872	0826	0785	0747	0713	0682	0654	0628
2	2225	2079	1951	1837	1736	1646	1564	1490	1423	1362	1305	1253
3	3303	3090	2903	2737	2588	2455	2334	2225	2126	2035	1951	1874
4	4339	4067	3827	3612	3420	3247	3090	2948	2817	2698	2588	2487
5	5320	5000	4714	4457	4226	4017	3827	3653	3495	3349	3214	3090
6	6235	5878	5556	5264	5000	4759	4540	4339	4154	3984	3827	3681
7	7071	6691	6344	6026	5736	5469	5225	5000	4792	4601	4423	4258
8	7818	7431	7071	6737	6428	6142	5878	5633	5406	5196	5000	4818
9	8467	8090	7730	7390	7071	6773	6494	6235	5993	5767	5556	5358
10	9010	8660	8315	7980	7660	7357	7071	6802	6549	6311	6088	5878
11	9439	9135	8819	8502	8192	7891	7604	7331	7071	6826	6593	6374
12	9749	9511	9239	8952	8660	8372	8090	7818	7557	7308	7071	6845
13	9937	9781	9569	9325	9063	8795	8526	8262	8005	7757	7518	7290
14	—	9945	9808	9618	9397	9158	8910	8660	8413	8170	7934	7705
15	—	—	9952	9830	9659	9458	9239	9010	8777	8544	8315	8090
16	—	—	—	9957	9848	9694	9511	9309	9096	8879	8660	8443
17	—	—	—	—	9962	9864	9724	9556	9369	9172	8969	8763
18	—	—	—	—	—	9966	9877	9749	9595	9423	9239	9048
19	—	—	—	—	—	—	9969	9888	9771	9629	9469	9298
20	—	—	—	—	—	—	—	9972	9898	9791	9659	9511
21	—	—	—	—	—	—	—	—	9975	9907	9808	9686
22	—	—	—	—	—	—	—	—	—	9977	9914	9823
23	—	—	—	—	—	—	—	—	—	—	9979	9921
24	—	—	—	—	—	—	—	—	—	—	—	9980

Table 2. Harmonic analysis for p even but not a multiple of 4

Values of p

k	6		10		14		18		22		26	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	8660	5000	5878	8090	4339	9010	3420	9397	2817	9595	2393	9709
2	—	—	9511	3090	7818	6235	6428	7660	5406	8413	4647	8855
3	—	—	—	—	9749	2225	8660	5000	7557	6549	6631	7485
4	—	—	—	—	—	—	9848	1736	9096	4154	8230	5681
5	—	—	—	—	—	—	—	—	9898	1423	9350	3546
6	—	—	—	—	—	—	—	—	—	—	9927	1205

Values of p

k	30		34		38		42		46		50	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	2079	9781	1837	9830	1646	9864	1490	9888	1362	9907	1253	9921
2	4067	9135	3612	9325	3247	9458	2948	9556	2698	9629	2487	9686
3	5878	8090	5264	8502	4759	8795	4339	9010	3984	9172	3681	9298
4	7431	6691	6737	7390	6142	7891	5633	8262	5196	8544	4818	8763
5	8660	5000	7980	6026	7357	6773	6802	7331	6311	7757	5878	8090
6	9511	3090	8952	4457	8372	5469	7818	6235	7308	6826	6845	7290
7	9945	1045	9618	2737	9158	4017	8660	5000	8170	5767	7705	6374
8	—	—	9957	0923	9694	2435	9309	3653	8879	4601	8443	5358
9	—	—	—	—	9966	0826	9749	2225	9423	3349	9048	4258
10	—	—	—	—	—	—	9972	0747	9791	2035	9511	3090
11	—	—	—	—	—	—	—	—	9977	0682	9823	1874
12	—	—	—	—	—	—	—	—	—	—	9980	0628

Table 2 (*continued*)

Values of p

k	54		58		62		66		70		74	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	1161	9932	1081	9941	1012	9949	0951	9955	0896	9960	0848	9964
2	2306	9730	2150	9766	2013	9795	1893	9819	1786	9839	1690	9856
3	3420	9397	3193	9477	2994	9541	2817	9595	2660	9640	2520	9677
4	4488	8936	4199	9076	3944	9190	3717	9284	3514	9362	3331	9429
5	5495	8355	5156	8569	4853	8743	4582	8888	4339	9010	4119	9112
6	6428	7660	6052	7961	5713	8208	5406	8413	5129	8584	4877	8730
7	7274	6862	6877	7260	6514	7588	6182	7861	5878	8090	5600	8285
8	8021	5972	7622	6474	7248	6890	6901	7237	6579	7531	6282	7780
9	8660	5000	8277	5612	7908	6121	7557	6549	7228	6911	6919	7220
10	9182	3961	8835	4684	8486	5290	8146	5801	7818	6235	7507	6607
11	9580	2868	9290	3701	8978	4404	8660	5000	8346	5509	8040	5946
12	9848	1736	9635	2675	9378	3473	9096	4154	8806	4739	8515	5243
13	9983	0581	9868	1618	9681	2507	9450	3271	9195	3930	8929	4502
14	—	—	9985	0541	9885	1514	9718	2358	9511	3090	9279	3729
15	—	—	—	—	9987	0506	9898	1423	9749	2225	9562	2928
16	—	—	—	—	—	—	9989	0476	9909	1342	9776	2107
17	—	—	—	—	—	—	—	—	9990	0449	9919	1270
18	—	—	—	—	—	—	—	—	—	—	9991	0424

Values of p

k	78		82		86		90		94		98	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	0805	9968	0765	9971	0730	9973	0698	9976	0668	9978	0641	9979
2	1604	9871	1526	9883	1456	9893	1392	9903	1333	9911	1279	9918
3	2393	9709	2279	9737	2174	9761	2079	9781	1992	9800	1912	9816
4	3167	9485	3017	9534	2881	9576	2756	9613	2642	9645	2537	9673
5	3920	9200	3738	9275	3572	9340	3420	9397	3280	9447	3151	9491
6	4647	8855	4437	8962	4245	9054	4067	9135	3904	9206	3753	9269
7	5345	8452	5110	8596	4894	8720	4695	8829	4510	8925	4339	9010
8	6007	7994	5753	8179	5518	8340	5299	8480	5096	8604	4907	8713
9	6631	7485	6362	7715	6112	7915	5878	8090	5659	8244	5455	8381
10	7212	6927	6934	7205	6673	7448	6428	7660	6197	7848	5981	8014
11	7746	6324	7466	6653	7199	6941	6947	7193	6708	7417	6482	7614
12	8230	5681	7953	6062	7687	6397	7431	6691	7188	6952	6957	7183
13	8660	5000	8394	5436	8133	5819	7880	6157	7637	6456	7403	6723
14	9035	4287	8785	4777	8536	5209	8290	5592	8051	5932	7818	6235
15	9350	3546	9125	4091	8893	4572	8660	5000	8429	5381	8202	5721
16	9605	2782	9411	3380	9203	3911	8988	4384	8770	4806	8551	5184
17	9798	2000	9643	2650	9464	3229	9272	3746	9071	4209	8866	4625
18	9927	1205	9817	1904	9675	2529	9511	3090	9332	3594	9144	4048
19	9992	0403	9934	1147	9834	1816	9703	2419	9551	2963	9385	3454
20	—	—	9993	0383	9940	1094	9848	1736	9728	2318	9587	2845
21	—	—	—	—	9993	0365	9945	1045	9861	1663	9749	2225
22	—	—	—	—	—	—	9994	0349	9950	1001	9872	1596
23	—	—	—	—	—	—	—	—	9994	0334	9954	0960
24	—	—	—	—	—	—	—	—	—	—	9995	0321

TABLE 3. Harmonic analysis for *p* odd

Values of p

k	5		7		9		11		13		15	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	9511	3090	7818	6235	6428	7660	5406	8413	4647	8855	4067	9135
2	5878	8090	9749	2225	9848	1736	9096	4154	8230	5681	7431	6691
3	—	—	4339	9010	8660	5000	9898	1423	9927	1205	9511	3090
4	—	—	—	—	3420	9397	7557	6549	9350	3546	9945	1045
5	—	—	—	—	—	—	2817	9595	6631	7485	8660	5000
6	—	—	—	—	—	—	—	—	2393	9709	5878	8090
7	—	—	—	—	—	—	—	—	—	—	2079	9781

Values of p

k	17		19		21		23		25		27	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	3612	9325	3247	9458	2948	9556	2698	9629	2487	9686	2306	9730
2	6737	7390	6142	7891	5633	8262	5196	8544	4818	8763	4488	8936
3	8952	4457	8372	5469	7818	6235	7308	6826	6845	7290	6428	7660
4	9957	0923	9694	2455	9309	3653	8879	4601	8443	5358	8021	5972
5	9618	2737	9966	0826	9972	0747	9791	2035	9511	3090	9182	3961
6	7980	6026	9158	4017	9749	2225	9977	0682	9980	0628	9848	1736
7	5264	8502	7357	6773	8660	5000	9423	3349	9823	1874	9983	0581
8	1837	9830	4759	8795	6802	7331	8170	5767	9048	4258	9580	2868
9	—	—	1646	9864	4339	9010	6311	7757	7705	6374	8660	5000
10	—	—	—	—	1490	9888	3984	9172	5878	8090	7274	6862
11	—	—	—	—	—	—	1362	9907	3681	9298	5495	8355
12	—	—	—	—	—	—	—	—	1253	9921	3420	9397
13	—	—	—	—	—	—	—	—	—	—	1161	9932

Values of p

k	29		31		33		35		37		39	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	2150	9766	2013	9795	1893	9819	1786	9839	1690	9856	1604	9871
2	4199	9076	3944	9190	3717	9284	3514	9362	3331	9429	3167	9485
3	6052	7961	5713	8208	5406	8413	5129	8584	4877	8730	4647	8855
4	7622	6474	7248	6890	6901	7237	6579	7531	6282	7780	6007	7994
5	8835	4684	8486	5290	8146	5801	7818	6235	7507	6607	7212	6927
6	9635	2675	9378	3473	9096	4154	8806	4739	8515	5243	8230	5681
7	9985	0541	9885	1514	9718	2358	9511	3090	9279	3729	9035	4287
8	9868	1618	9987	0506	9989	0476	9909	1342	9776	2107	9605	2782
9	9290	3701	9681	2507	9898	1423	9990	0449	9991	0424	9927	1205
10	8277	5612	8978	4404	9450	3271	9749	2225	9919	1270	9992	0403
11	6877	7260	7908	6121	8660	5000	9195	3930	9562	2928	9798	2000
12	5156	8569	6514	7588	7557	6549	8346	5509	8929	4502	9350	3546
13	3193	9477	4853	8743	6182	7861	7228	6911	8040	5946	8660	5000
14	1081	9941	2994	9541	4582	8888	5878	8090	6919	7220	7746	6324
15	—	—	1012	9949	2817	9595	4339	9010	5600	8285	6631	7485
16	—	—	—	—	0951	9955	2660	9640	4119	9112	5345	8452
17	—	—	—	—	—	—	0896	9960	2520	9677	3920	9200
18	—	—	—	—	—	—	—	—	0848	9964	2393	9709
19	—	—	—	—	—	—	—	—	—	—	0805	9968

TABLE 3 (*continued*)

Values of p

k	41		43		45		47		49		51	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	1526	9883	1456	9893	1392	9903	1333	9911	1279	9918	1229	9924
2	3017	9534	2881	9576	2756	9613	2642	9645	2537	9673	2439	9698
3	4437	8962	4245	9054	4067	9135	3904	9206	3753	9269	3612	9325
4	5753	8179	5518	8340	5299	8480	5096	8604	4907	8713	4731	8810
5	6934	7205	6673	7448	6428	7660	6197	7848	5981	8014	5778	8162
6	7953	6062	7687	6397	7431	6691	7188	6952	6957	7183	6737	7390
7	8785	4777	8536	5209	8290	5592	8051	5932	7818	6235	7594	6506
8	9411	3380	9203	3911	8988	4384	8770	4806	8551	5184	8336	5524
9	9817	1904	9675	2529	9511	3090	9332	3594	9144	4048	8952	4457
10	9993	0383	9940	1094	9848	1736	9728	2318	9587	2845	9432	3324
11	9934	1147	9993	0365	9994	0349	9950	1001	9872	1596	9768	2139
12	9643	2650	9834	1816	9945	1045	9994	0334	9995	0321	9957	0923
13	9125	4091	9464	3229	9703	2419	9861	1663	9954	0960	9995	0308
14	8394	5436	8893	4572	9272	3746	9551	2963	9749	2225	9882	1534
15	7466	6653	8133	5819	8660	5000	9071	4209	9385	3454	9618	2737
16	6362	7715	7199	6941	7880	6157	8429	5381	8866	4625	9209	3898
17	5110	8596	6112	7915	6947	7193	7637	6456	8202	5721	8660	5000
18	3738	9275	4894	8720	5878	8090	6708	7417	7403	6723	7980	6026
19	2279	9737	3572	9340	4695	8829	5659	8244	6482	7614	7179	6961
20	0765	9971	2174	9761	3420	9397	4510	8925	5455	8381	6269	7791
21	—	—	0730	9973	2079	9781	3280	9447	4339	9010	5264	8502
22	—	—	—	—	0698	9976	1992	9800	3151	9491	4180	9085
23	—	—	—	—	—	—	0668	9978	1912	9816	3032	9529
24	—	—	—	—	—	—	—	—	0641	9979	1837	9830
25	—	—	—	—	—	—	—	—	—	—	0616	9981

Values of p

k	53		55		57		59		61		63	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	1183	9930	1140	9935	1100	9939	1063	9943	1028	9947	0996	9950
2	2349	9720	2265	9740	2187	9758	2114	9774	2046	9789	1981	9802
3	3482	9374	3360	9418	3247	9458	3141	9494	3041	9526	2948	9556
4	4566	8897	4412	8974	4268	9044	4132	9106	4005	9163	3884	9215
5	5586	8294	5406	8413	5237	8519	5077	8616	4925	8703	4783	8782
6	6528	7575	6330	7741	6142	7891	5964	8027	5794	8150	5633	8262
7	7378	6750	7171	6969	6973	7168	6783	7348	6602	7511	6428	7660
8	8125	5830	7919	6106	7719	6357	7526	6585	7339	6793	7159	6982
9	8757	4828	8563	5164	8372	5469	8183	5748	7998	6002	7818	6235
10	9267	3758	9096	4154	8923	4515	8748	4846	8573	5148	8400	5425
11	9646	2636	9511	3090	9365	3506	9213	3888	9057	4239	8899	4562
12	9890	1476	9801	1986	9694	2455	9574	2887	9445	3285	9309	3653
13	9996	0296	9963	0856	9905	1374	9827	1853	9733	2297	9626	2708
14	9960	0888	9996	0286	9996	0276	9968	0798	9917	1284	9848	1736
15	9786	2060	9898	1423	9966	0826	9996	0266	9997	0257	9972	0747
16	9473	3203	9671	2542	9815	1917	9912	1327	9970	0772	9997	0249
17	9028	4301	9319	3628	9544	2985	9714	2373	9838	1793	9922	1243
18	8456	5338	8844	4667	9158	4017	9407	3392	9601	2795	9749	2225
19	7765	6301	8255	5644	8660	5000	8993	4373	9263	3767	9479	3185
20	6966	7175	7557	6549	8058	5922	8477	5304	8827	4700	9115	4113
21	6068	7949	6762	7367	7357	6773	7866	6175	8297	5582	8660	5000
22	5085	8610	5878	8090	6568	7541	7165	6976	7679	6406	8119	5837
23	4031	9151	4917	8707	5698	8218	6382	7698	6979	7162	7498	6617
24	2921	9564	3893	9211	4759	8795	5528	8333	6206	7841	6802	7331
25	1769	9842	2817	9595	3763	9265	4611	8874	5367	8438	6038	7971
26	0592	9982	1705	9854	2721	9623	3642	9313	4471	8945	5214	8533
27	—	—	0571	9984	1646	9864	2631	9648	3528	9357	4339	9010
28	—	—	—	—	0551	9985	1591	9873	2547	9670	3420	9397
29	—	—	—	—	—	—	0532	9986	1539	9881	2468	9691

TABLE 3 (*continued*)

Values of p

k	65		67		69		71		73		75	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	0965	9953	0936	9956	0909	9959	0884	9961	0860	9963	0837	9965
2	1921	9814	1865	9825	1811	9835	1761	9844	1713	9852	1668	9860
3	2859	9582	2776	9607	2698	9629	2624	9650	2554	9668	2487	9686
4	3771	9262	3664	9305	3562	9344	3466	9380	3375	9413	3289	9444
5	4647	8855	4519	8921	4397	8981	4282	9037	4172	9088	4067	9135
6	5480	8365	5334	8458	5196	8544	5064	8623	4938	8696	4818	8763
7	6262	7797	6103	7922	5951	8036	5806	8142	5667	8239	5534	8329
8	6985	7156	6818	7315	6657	7462	6503	7597	6354	7722	6211	7837
9	7643	6448	7473	6645	7308	6826	7149	6992	6995	7147	6845	7290
10	8230	5681	8063	5916	7899	6133	7739	6333	7583	6519	7431	6691
11	8740	4860	8581	5135	8424	5389	8268	5624	8115	5843	7965	6046
12	9168	3994	9024	4309	8879	4601	8733	4872	8588	5124	8443	5358
13	9511	3090	9388	3445	9260	3774	9129	4081	8996	4367	8862	4633
14	9764	2158	9669	2550	9565	2916	9454	3258	9338	3577	9219	3875
15	9927	1205	9866	1634	9791	2035	9705	2410	9611	2761	9511	3090
16	9997	0242	9975	0703	9935	1136	9880	1542	9813	1925	9736	2284
17	9974	0724	9997	0234	9997	0228	9978	0663	9942	1074	9893	1461
18	9857	1684	9931	1170	9977	0682	9998	0221	9998	0215	9980	0628
19	9649	2627	9778	2094	9873	1587	9939	1104	9979	0645	9998	0209
20	9350	3546	9539	3001	9688	2478	9802	1978	9887	1501	9945	1045
21	8964	4432	9216	3881	9423	3349	9589	2837	9721	2345	9823	1874
22	8495	5276	8812	4727	9079	4192	9301	3673	9484	3172	9632	2689
23	7946	6072	8331	5531	8660	5000	8940	4481	9176	3975	9373	3486
24	7323	6810	7776	6287	8170	5767	8509	5253	8800	4750	9048	4258
25	6631	7485	7153	6988	7611	6486	8011	5985	8359	5488	8666	5000
26	5878	8090	6468	7627	6990	7151	7451	6669	7856	6187	8211	5707
27	5070	8620	5725	8199	6311	7757	6833	7302	7296	6839	7705	6374
28	4214	9069	4932	8699	5579	8299	6160	7877	6681	7441	7145	6997
29	3319	9433	4096	9123	4802	8772	5440	8391	6016	7988	6534	7570
30	2393	9709	3224	9466	3984	9172	4677	8839	5307	8475	5878	8090
31	1445	9895	2323	9726	3133	9496	3878	9217	4559	8900	5180	8554
32	0483	9988	1402	9901	2257	9742	3048	9524	3777	9259	4446	8957
33	—	—	0469	9989	1362	9907	2194	9756	2967	9550	3681	9298
34	—	—	—	—	0455	9990	1324	9912	2135	9769	2890	9573
35	—	—	—	—	—	—	0442	9990	1287	9917	2079	9781
36	—	—	—	—	—	—	—	—	0430	9991	1253	9921
37	—	—	—	—	—	—	—	—	—	—	0419	9991

Table 3 (*continued*)

Values of p

k	77		79		81		83		85		87	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	0815	9967	0795	9968	0775	9970	0756	9971	0739	9973	0722	9974
2	1625	9867	1584	9874	1545	9880	1508	9886	1473	9891	1439	9896
3	2424	9702	2363	9717	2306	9730	2252	9743	2199	9755	2150	9766
4	3206	9472	3128	9498	3053	9522	2982	9545	2914	9566	2849	9586
5	3968	9179	3873	9220	3782	9257	3695	9292	3612	9325	3533	9355
6	4703	8825	4593	8883	4488	8936	4387	8986	4291	9032	4199	9076
7	5406	8413	5284	8490	5167	8562	5054	8629	4947	8691	4843	8749
8	6074	7944	5942	8043	5815	8136	5693	8222	5575	8302	5462	8377
9	6701	7422	6562	7546	6428	7660	6298	7767	6173	7867	6052	7961
10	7284	6851	7141	7000	7002	7139	6868	7269	6737	7390	6610	7503
11	7818	6235	7675	6411	7534	6575	7398	6729	7264	6872	7135	7007
12	8301	5577	8160	5781	8021	5972	7885	6150	7752	6317	7622	6474
13	8727	4882	8593	5114	8460	5332	8328	5536	8197	5727	8069	5907
14	9096	4154	8972	4415	8848	4660	8723	4890	8598	5106	8474	5309
15	9405	3399	9295	3689	9182	3961	9068	4217	8952	4457	8835	4684
16	9650	2621	9559	2939	9461	3237	9360	3519	9256	3784	9150	4034
17	9832	1826	9762	2170	9684	2494	9600	2801	9511	3090	9417	3364
18	9948	1018	9903	1387	9848	1736	9784	2067	9713	2379	9635	2675
19	9998	0204	9982	0596	9953	0968	9912	1321	9862	1656	9803	1973
20	9981	0612	9998	0199	9998	0194	9984	0567	9957	0923	9920	1261
21	9898	1423	9951	0993	9983	0581	9998	0189	9998	0185	9985	0541
22	9749	2225	9840	1780	9908	1353	9955	0945	9985	0554	9998	0181
23	9535	3012	9668	2556	9773	2117	9855	1695	9916	1290	9959	0902
24	9258	3780	9434	3316	9580	2868	9699	2436	9794	2019	9868	1618
25	8919	4522	9141	4055	9329	3602	9487	3162	9618	2737	9726	2326
26	8521	5234	8790	4769	9022	4314	9221	3871	9390	3439	9533	3021
27	8066	5911	8383	5452	8660	5000	8901	4557	9110	4124	9290	3701
28	7557	6549	7923	6101	8247	5656	8531	5217	8781	4785	8998	4362
29	6999	7143	7414	6711	7784	6278	8112	5847	8403	5421	8660	5000
30	6393	7690	6857	7279	7274	6862	7647	6444	7980	6026	8277	5612
31	5745	8185	6257	7801	6720	7405	7138	7004	7513	6599	7850	6194
32	5059	8626	5618	8273	6126	7904	6588	7524	7005	7136	7383	6745
33	4339	9010	4942	8693	5495	8355	6000	8000	6459	7634	6877	7260
34	3590	9333	4236	9058	4831	8756	5377	8431	5878	8090	6335	7737
35	2817	9595	3503	9366	4138	9104	4724	8814	5264	8502	5760	8174
36	2026	9793	2748	9615	3420	9397	4044	9146	4622	8868	5156	8569
37	1221	9925	1975	9803	2682	9634	3341	9425	3955	9185	4524	8918
38	0408	9992	1190	9929	1927	9813	2619	9651	3265	9452	3868	9221
39	—	—	0398	9992	1161	9932	1881	9821	2558	9667	3193	9477
40	—	—	—	—	0388	9992	1133	9936	1837	9830	2501	9682
41	—	—	—	—	—	—	0378	9993	1107	9939	1796	9837
42	—	—	—	—	—	—	—	—	0370	9993	1081	9941
43	—	—	—	—	—	—	—	—	—	—	0361	9993

TABLE 3 (*continued*)

Values of p

k	89		91		93		95		97		99	
	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine	sine	cosine
1	0705	9975	0690	9976	0675	9977	0661	9978	0647	9979	0634	9980
2	1407	9900	1377	9905	1347	9909	1319	9913	1292	9916	1266	9920
3	2102	9777	2057	9786	2013	9795	1971	9804	1931	9812	1893	9819
4	2787	9604	2727	9621	2670	9637	2615	9652	2562	9666	2511	9679
5	3457	9383	3384	9410	3314	9435	3247	9458	3182	9480	3120	9501
6	4110	9116	4025	9154	3944	9190	3865	9223	3789	9254	3717	9284
7	4743	8804	4647	8855	4555	8902	4466	8947	4380	8990	4298	9029
8	5352	8447	5247	8513	5146	8575	5048	8633	4953	8687	4862	8738
9	5935	8048	5822	8131	5713	8208	5607	8280	5505	8348	5406	8413
10	6488	7610	6369	7710	6254	7803	6142	7891	6034	7974	5929	8053
11	7008	7133	6886	7252	6766	7363	6650	7468	6537	7567	6428	7660
12	7494	6621	7370	6759	7248	6890	7129	7012	7014	7128	6901	7237
13	7942	6076	7818	6235	7697	6385	7577	6526	7460	6659	7346	6785
14	8351	5501	8230	5681	8110	5850	7992	6011	7876	6162	7761	6306
15	8718	4898	8602	5099	8486	5290	8372	5469	8258	5640	8146	5801
16	9042	4271	8933	4494	8824	4705	8715	4904	8606	5093	8497	5272
17	9321	3622	9222	3867	9122	4098	9020	4318	8917	4525	8815	4723
18	9553	2956	9467	3221	9378	3473	9286	3712	9192	3939	9096	4154
19	9738	2274	9667	2560	9591	2832	9511	3090	9427	3336	9341	3569
20	9874	1582	9820	1887	9760	2178	9694	2455	9623	2718	9549	2969
21	9961	0881	9927	1205	9885	1514	9835	1809	9779	2090	9718	2358
22	9998	0177	9987	0518	9964	0843	9933	1155	9894	1452	9848	1736
23	9986	0529	9999	0173	9999	0169	9988	0496	9967	0809	9938	1108
24	9924	1232	9963	0862	9987	0506	9999	0165	9999	0162	9989	0476
25	9812	1929	9880	1547	9930	1180	9966	0826	9988	0486	9999	0159
26	9652	2617	9749	2225	9828	1847	9889	1483	9936	1131	9969	0793
27	9443	3291	9573	2893	9681	2507	9770	2133	9842	1772	9898	1423
28	9187	3949	9350	3546	9489	3154	9608	2774	9706	2405	9788	2048
29	8886	4587	9083	4183	9255	3788	9403	3403	9530	3028	9638	2665
30	8540	5202	8773	4799	8978	4404	9158	4017	9314	3639	9450	3271
31	8152	5792	8421	5393	8660	5000	8872	4613	9059	4234	9224	3863
32	7723	6352	8029	5961	8303	5573	8548	5190	8766	4812	8960	4441
33	7256	6881	7598	6501	7908	6121	8186	5743	8436	5369	8660	5000
34	6752	7376	7132	7010	7476	6641	7789	6272	8071	5904	8326	5539
35	6215	7834	6631	7485	7011	7131	7357	6773	7672	6414	7958	6056
36	5647	8253	6099	7925	6514	7588	6893	7244	7241	6897	7557	6549
37	5051	8631	5538	8327	5987	8010	6400	7684	6779	7351	7127	7015
38	4430	8965	4950	8689	5432	8396	5878	8090	6289	7775	6668	7453
39	3786	9256	4339	9010	4853	8743	5330	8461	5773	8166	6182	7861
40	3124	9500	3707	9288	4252	9051	4759	8795	5232	8522	5671	8237
41	2446	9696	3057	9521	3631	9318	4168	9090	4669	8843	5137	8580
42	1756	9845	2393	9709	2994	9541	3558	9346	4087	9127	4582	8888
43	1057	9944	1718	9851	2342	9722	2933	9560	3488	9372	4009	9161
44	0353	9994	1034	9946	1681	9858	2294	9733	2874	9578	3420	9397
45	—	—	0345	9994	1012	9949	1646	9864	2248	9744	2817	9595
46	—	—	—	—	0338	9994	0990	9951	1612	9869	2203	9754
47	—	—	—	—	—	—	0331	9995	0970	9953	1580	9874
48	—	—	—	—	—	—	—	—	0324	9995	0951	9955
49	—	—	—	—	—	—	—	—	—	—	0317	9995