The Analysis of Economic Time-Series—Part I: Prices

By M. G. Kendall

Division of Research Techniques, London School of Economics

(Read before the Royal Statistical Society, December 17th, 1952, Professor A. Bradford Hill, C.B.E., Ex-President, in the Chair)

Introduction and Summary

1. It has been customary to analyse an economic time-series by extracting from it a long-term movement, or trend, for separate study and then scrutinizing the residual portion for short-term oscillatory movements and random fluctuations. The assumption latent in this procedure is that the long-term and short-term movements are due to separate causal influences and therefore that the mathematical process of analysis corresponds more or less roughly to a real distinction of type in the generative system. When it began to appear that the classical method of analyzing stationary economic fluctuations into cycles and residual elements broke down, and that better accounts of such movements were given by autoregressive schemes in which the disturbance element played an integral part, it was an easy step to proceed one stage further, and to enquire whether the so-called trend in a series was in fact separable from the short-term movements, or whether it should be regarded as generated by a set of forces which also gave rise to the short-term movements.

2. The work described in the present paper sprang from this idea. It was necessary in the first place to find suitable material on which to test it and I give below particulars of 22 price-series, ranging from 486 terms at weekly intervals to 2,387 terms at weekly intervals, which were chosen for the purpose. I began to construct models and to fit them to these data and also to other series which are not mentioned in this paper; but before the work had gone very far it appeared that the pattern of events in the price series was much less systematic than is generally believed. This is perhaps the most significant part of the present paper. I shall not have space to deal with the corresponding phenomenon in production data, and shall mention only incidentally the problem of trend fitting.

3. Broadly speaking the results are these:

(*a*) In series of prices which are observed at fairly close intervals the random changes from one term to the next are so large as to swamp any systematic effect which may be present. The data behave almost like wandering series.

(*b*) It is therefore difficult to distinguish by statistical methods between a genuine wandering series and one wherein the systematic element is weak.

(*c*) Until some way has been found of circumventing this difficulty, trend fitting, and perhaps the fitting of any model, is a highly hazardous undertaking. It may be possible for an econometrician to test whether the data agree with a hypothesis suggested by prior analysis, but it may be impossible to discriminate between quite different hypotheses which all fit the data.

(*d*) There is experimental evidence and theoretical support for the belief that aggregative index numbers behave more systematically than their components. This *might* be due to the reduction of the random elements by averaging and the consequent emergence of systematic constituents; but it could equally well be due to chance. If it is, there will appear spurious time-correlations in aggregative series and the use of index-numbers in econometric work needs extensive reconsideration.

(*e*) An analysis of stock-exchange movements revealed little serial correlation within series and little lag correlation between series. Unless individual stocks behave differently from the average of similar stocks, there is no hope of being able to predict movements on the exchange for a week ahead without extraneous information.

The Data

4. For the purposes of this study I used 22 economic series as follows:

Actuaries' Index of Industrial Share Prices

	Period	*Interval*	*Number of Terms*
1. Banks and Discount Companies	1928–1938	Week	486
2. Insurance Companies	,,	,,	,,
3. Investment Trusts	,,	,,	,,
4. Building Materials	,,	,,	,,
5. Coal	,,	,,	,,
6. Cotton	,,	,,	,,
7. Electric Light and Power	,,	,,	,,
8. Gas	,,	,,	,,
9. Iron and Steel	,,	,,	,,
10. Oil	,,	,,	,,
11. Total Industrial Productive	,,	,,	,,
12. Home Rails	,,	,,	,,
13. Shipping	,,	,,	,,
14. Stores and Catering	,,	,,	,,
15. Total Industrial Distributive	,,	,,	,,
16. Breweries and Distilleries	,,	,,	,,
17. Miscellaneous	,,	,,	,,
18. Total Industrial Miscellaneous	,,	,,	,,
19. Industrials (all classes combined)	,,	,,	,,
20. Basic cash wheat at Chicago in cents per bushel	Jan., 1883–Sept., 1934 (excluding 1915–1920 inclusive; one missing value interpolated for 10.3.1933)	,,	2,387
21. Monthly average of preceding series	Jan., 1883–Sept., 1934 (excluding 1915–1920)	Month	548
22. Spot cotton at New York in cents per pound	Aug., 1816–Jan., 1951 (excluding 1861–1866 and 1914–20; one missing value for Oct., 1857, inserted as average of the two neighbouring months)	,,	1,446

Sources

For the actuaries' indices I am indebted to the Institute and Faculty of Actuaries who allowed me to extract the figures from the Institute's records.

Series 20 and 21 are taken from *Wheat Studies of the Food Research Institute*, vol. 11, No. 3, November, 1934 (Stanford University, California).

Series 22 is taken from *Statistics on Cotton and Related Data*, U.S. Department of Agriculture, Bureau of Agricultural Economics, Statistical Bulletin, No. 99, Washington, D.C., June, 1951.

5. The golden rule in publishing work on time-series is to give the original data. To do so in this case would have required about thirty pages of journal. I shall, however, be glad to make available copies of the series or the cards on which they have been punched. A good many of the secondary statistics are presented below but here also the working sheets are available to any responsible person who is interested. So far as possible I have not "edited" any of the series even where there was some case for making the attempt; but in some instances I have omitted war periods and should have been compelled to do so by the lack of data. All such omissions are indicated above.

Analysis of the Price Data

The Chicago Wheat Series (20 *and* 21)

6. Although the wheat-price series was not the first I examined it is useful to begin an expository account by considering it because it is a long series of prices wherein little or no element of aggrega-

tion enters. The Stanford Wheat Studies from which the data were taken give graphs of the series. They are unfortunately too long to reproduce on a page of this size.*

From the primary data the series of 2,379 first differences were constructed and a frequency distribution formed which is given as the marginal column in Table 1. It was here that the first fact of significance emerged, for the resulting distribution had nearly perfect symmetry and an appearance of approximate normality. The bivariate distribution of one difference against the next succeeding difference is even more illuminating. Table 1 is a condensed version and presents on the face of it an excellent picture of independence; on a regression diagram the two lines would almost coincide with the variate axes.

7. The decision to use first differences instead of the original series was not an arbitrary one. When a price is fixed on a free market both parties know what was the price of the previous transactions and use that price as a starting point for negotiation. It is the change, not the absolute value, which constitutes the fundamental element in the price determination. One can think of exceptions, perhaps, but for the commodities I am discussing they are not relevant.

8. At first sight the implications of these results are disturbing. If the series is homogeneous, it seems that the change in price from one week to the next is practically independent of the change from that week to the week after. This alone is enough to show that it is impossible to predict the price from week to week from the series itself. And if the series really is wandering, any systematic movements such as trends or cycles which may be "observed" in such series are illusory. The series looks like a "wandering" one, almost as if once a week the Demon of Chance drew a random number from a symmetrical population of fixed dispersion and added it to the current price to determine the next week's price. And this, we may recall, is not the behaviour in some small backwater market. The data derive from the Chicago wheat market over a period of fifty years during which at least two attempts were made to corner wheat, and one might have expected the wildest irregularities in the figures. To the statistician there is some pleasure in the thought that the symmetrical distribution reared its graceful head undisturbed amid the uproar of the Chicago wheat-pit. The economist, I suspect, or at any rate the trade cyclist, will look for statistical snags before he is convinced of the absence of systematic movements. And he will be very right to do so.

9. One possible source of error is that, in treating as homogeneous a price-series extending from 1883 to 1934, we may be oversimplifying the hypotheses. The series was therefore divided into two sections covering the periods 1883–1914 and 1921–1934; and a bivariate distribution of first differences on the lines of Table 1 constructed for each. I omit the figures to save space, but the story told by each section was the same as for the series taken together.

There are seven widely outlying values in the whole series, and it did not seem to me to be sophisticating the data to omit them from the calculation of moments.† The following constants were computed from a more finely grouped distribution and are corrected by Sheppard's formulae for the grouping effect.

	Period 1883–1914	*Period* 1921–1934	*Both Periods Together*
μ_1'	0	0·0964	0·0289
μ_2	7·7576	22·7789	12·3294
μ_3	12·6787	34·5882	20·2709
μ_4	657·3073	2745·8873	1293·0607
β_1	0·344	0·101	0·219
β_2	10·856	5·292	8·506

The distributions are accordingly rather leptokurtic.

10. The general pattern of independence present in these data makes it very unlikely that there are large serial correlations present but to make sure they were worked out to the tenth order and are given in Table 2. For computing purposes it is easier to work out serial covariance

* The weekly series are, so far as possible, closing quotations on Fridays and are given to the nearest $\frac{1}{8}$ cent. The monthly figures are averages of four or five weekly figures.

† Three values are due to a jump from September 21st to September 28th, 1888, of 53 cents and a recession the following week of 38 cents; four values are due to a similar movement in May, 1898.

TABLE 1

Frequency Distribution of Differences of Series 20, Difference between Weeks t *and* t + 1 *against Difference between Weeks* t + 1 *and* t + 2.

(Condensed to save space from original table for which the interval was one unit. Values falling on the border line of intervals were allotted ½ to each.)

Difference, tth week less $(t + 1)$th week; cents per bushel.	*Difference,* (t + 1)th *Week Less* (t + 2)th *Week; Cents per Bushel, Mid-point of Interval* −21	−19	−17	−15	−13	−11	−9	−7	−5	−3	−1	1	3	5	7	9	11	13	15	17	19	21	*Totals*
−17	—	—	—	—	—	—	—	—	—	—	½	—	—	—	—	—	—	—	—	—	—	—	½
−15	—	—	—	—	—	—	—	—	—	1	½	1	—	—	—	2¼	¼	—	—	—	—	—	5
−13	—	—	—	—	—	—	—	—	1	1	1	2	—	—	½	¾	¼	—	—	1	—	1	8½
−11	—	—	—	—	—	—	—	—	1½	—	½	½	2	1	—	—	1	1	—	—	—	—	7½
−9	—	—	—	1	—	—	—	2	2½	3	3½	4	2	3	2	—	—	—	1	—	—	—	24
−7	—	—	—	—	1	—	3	2½	3	4	10	8	5	2	6	1	—	1	—	—	1	—	47½
−5	—	—	—	1	1	1	5	5½	8½	9¼	23	28¾	16	8½	3½	1	2	—	—	—	—	—	114
−3	—	—	—	½	½	2	5	6½	16	28	64¼	91	27¼	12½	6½	—	1	—	—	1	—	1	263
−1	—	—	—	—	3	2½	4	12½	27½	86¾	232	236¾	78½	19	12½	1½	2	2	—	½	—	—	721
1	½	½	—	—	—	—	3½	10½	22	71	268½	207¼	82¼	29¾	6¼	3	1	—	1½	1	—	—	708½
3	—	—	—	—	—	—	1	5	13¾	30¾	82¼	89¼	47½	8¾	4¼	1½	½	—	—	—	—	—	284½
5	—	—	—	—	1	1	1½	3	7¾	16¼	18½	20½	11½	9	5	2	—	1	1	—	—	1	100
7	—	—	—	½	—	—	1	1	6	7½	11½	12½	6½	2½	—	1	—	—	—	—	—	—	50
9	—	—	—	½	1	—	—	1	2	2	1½	—	2½	1½	1½	2½	—	—	—	—	½	½	17
11	—	—	—	—	—	—	—	—	1	1	3½	—	½	1½	—	½	—	—	—	1	—	—	9
13	—	—	—	—	—	—	—	—	—	—	—	1	1	—	2	—	—	—	1	—	—	—	5
15	—	—	½	½	½	—	—	—	1	—	—	1	1	—	—	—	—	—	—	—	—	—	4½
17	—	—	—	—	½	—	—	—	1	—	—	3	—	—	—	—	—	—	—	—	—	—	4½
19	—	—	—	½	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	1½
21	—	—	—	½	—	1	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	—	3½
Totals	½	½	½	5	8½	7½	24	49½	114½	262½	721	707½	283½	100	50	17	8	5	4½	4½	1½	3½	2,379

Note.—There were also seven outlying pairs with values (−6·75, −53·0), (−53·0, 37·75), (37·75, −1·25), (−6·5, −30·75), (−30·75, 3·75), (−20·0. 53·0) and (53·0, 11·75). The last term of the series was taken with the first so that the total number of pairs is 2,386.

TABLE 2

Serial Correlations of First Differences of Wheat Prices (Series 20)

(Decimal points omitted)

Order of Correlation	*Series* 1883–1914 (n = 1,669)	*Series* 1921–1934 (n = 716)	*Whole Series* 1883–1934 (*Omitting* 1915–1920) (n = 2,385)
1	−078	−063	−071
2	+078	+051	+065
3	−000	+030	+014
4	−075	+042	−019
5	+007	−066	−028
6	−032	−071	−051
7	−016	−039	−027
8	−059	−064	−061
9	−047	+023	−014
10	−038	+034	−004
Large sample standard error	0·025	0·038	0·020

for the primary series and to calculate those of the differences from formulae such as

$$\sum_{i=1}^{n-2}(x_{i+2}-x_{i+1})(x_{i+1}-x_i) = 2\sum_{i=1}^{n-1} x_i x_{i+1} - \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n-2} x_i x_{i+2} + x_1^2 + x_n^2 - x_1 x_2 - x_{n-1}x_n \quad . \qquad (1)$$

For long series when end effects can be neglected this is for all practical purposes equivalent to

$$\tau_k = -\tfrac{1}{2}\Delta^2 \rho_{k-1}/(1-\rho_1) \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

where τ represents the autocorrelations of first differences and ρ those of the primary series. In this case it was not much extra trouble to take end-effects into account and the serial correlations for the constituent series in Table 2 were calculated from the formula

$$r_k = \frac{\sum_{i=1}^{n-k-1} u_i u_{i+k} - \frac{1}{n-k-1}\sum_{i=1}^{n-k-1} u_i \sum_{i=k+1}^{n} u_i}{\left[\left\{\sum_{i=1}^{n-k-1} u_i^2 - \frac{1}{n-k-1}\left(\sum_{i=1}^{n-k-1}\right)^2\right\}\left\{\sum_{i=k+1}^{n} u_i^2 - \frac{1}{n-k-1}\left(\sum_{i=k+1}^{n} u_i\right)^2\right\}\right]^{\frac{1}{2}}} \quad . \qquad (3)$$

where u_i is the difference $x_{i+1} - x_i$.

In calculating the serials for the two series together a weighted average was taken, the formula being

$$r_k = \frac{(n_1-k-1)c_1 + (n_2-k-1)c_2}{\{(n_1-k-1)v_{11} + (n_2-k-1)v_{21}\}^{\frac{1}{2}}\{(n_1-k-1)v_{12} + (n_2-k-1)v_{22}\}^{\frac{1}{2}}} \quad . \qquad (4)$$

where n_1, n_2 are the number of terms, v_{11}, v_{21}, v_{12} and v_{22} refer to the variances of the first $n-k$ terms of the first and second series and the last $n-k$ terms of the first and second series respectively and c_1, c_2 are the covariances of the first and second series. Equation (4) has no very firm theoretical basis but it is obviously reasonable and will certainly not underestimate the magnitude of the serial correlations.

11. A comparison of the variances of the two parts of the series suggests that there has been an increase in variability since World War I. This is what one might perhaps have expected, but it is rather a nuisance from the point of view of analysis because it suggests that the series is not stationary. We have here an interesting and rather unusual case of a time-series for which the mean remains constant but the variance appears to be increasing. It is desirable to have a test of this type of departure from stationarity, distribution-free if possible. The problem is easy enough for a random series but not so easy for an autocorrelated series of unknown character, and as I do not require the test for present purposes I omit a discussion of the point.

12. A glance at Table 2 shows that the serial correlations are very small and, for most practical purposes, negligible. However, the first two serials are, in four cases out of six, more than three times their standard error as calculated from R. L. Anderson's formula

$$\text{var } r_k = \frac{n-1}{(n-2)^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

There are several possible explanations of this effect. Anderson's formula is based on normality and constant variance in the series. From general experience one would not expect the sampling variance to be increased by departure from parent normality in the direction of leptokurtosis, the distribution being symmetrical and n large. Nor, as it seems to me, can we appeal to non-stationarity to account for the difference, because the discrepancies occur most markedly in the first part of the series. I proceed to consider two more practical possibilities, one that the primary series was rounded up (or down) systematically, the other that for hour-to-hour or day-to-day changes in price there were systematic variations which have been lost to view in the weekly figures.

13. Suppose that a series of values $x_1, \ldots x_n$ is systematically rounded up by amounts $\varepsilon_1, \ldots \varepsilon_n$, ε thus having a positive mean but being serially independent. We shall then have for the differences $u_i(= x_{i+1} + \varepsilon_{i+1} - x_i - \varepsilon_i)$, in long series

$$\begin{aligned} \text{corr }(u_i, u_{i+k}) &= \frac{-\Delta^2 \rho_{k-1}}{2(1-\rho_1) + 2 \text{ var } \varepsilon/\text{var } x} \qquad k > 1 \\ &= \frac{-\Delta^2 \rho_{k-1} - \text{var } \varepsilon/\text{var } x}{2(1-\rho_1) + 2 \text{ var } \varepsilon/\text{var } x} \qquad k = 1 \quad . \quad . \quad . \quad (6) \end{aligned}$$

In our present case var x is of the order of 10 and (the series being given to the nearest $\frac{1}{8}$th) var $\varepsilon = \frac{1}{12}\left(\frac{1}{8}\right)^2 = 0{\cdot}0^213$. The effect of factors in var ε/var x, of the order of $0{\cdot}0^31$, in (6) is therefore quite insufficient to account for the observed size of the first serial. In any case it would not account for the sign of the second. Even if we suppose that the rounding off was alternately up and down, which would induce autocorrelation in the ε's, the magnitude of the effect would not be large enough to explain the observations.

14. Consider then the attenuation in serial relationship which may be caused by averaging over time-periods. Suppose a market is open from Monday to Friday, that any Monday may be regarded as following immediately on the preceding Friday and that the scheme of generation from day to day is a simple Markoff process

$$u_{t+1} = \alpha u_t + \varepsilon_{t+1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

It is fairly easy to show that the correlation between the average of the five daily prices in one week and the corresponding average in the succeeding week is given by

$$r = \frac{5\alpha^5 + 4(\alpha^6 + \alpha^4) + 3(\alpha^7 + \alpha^3) + 2(\alpha^8 + \alpha^2) + (\alpha^9 + \alpha)}{5 + 8\alpha + 6\alpha^2 + 4\alpha^3 + 2\alpha^4} \quad . \quad . \quad . \quad (8)$$

For $\alpha = 0{\cdot}8$ $r = 0{\cdot}499$ and the correlation is reduced by 38 per cent. For $\alpha = 0{\cdot}5$ $r = 0{\cdot}178$ and the serial correlation between weekly averages is only a third of that between days. For $\alpha = 0{\cdot}1$ $r = {\cdot}02$. Evidently the averaging over successive terms may obliterate some quite substantial serial correlation in prices. The weekly wheat figures were not, in fact, averages over the days of the week, but it is evident that the choice of one day in the week as the representative point will stilll further attenuate any day-to-day correlations.

A simple Markoff process would still not account for the negative sign of r_1 in Table 3 unless the correlation between successive days was negative, which seems unlikely but is not impossible owing to the tendency of a sensitive market to swing too far and to correct itself. I ought perhaps to mention that the possibility of "significant" serials appearing owing to very occasional exceptional values was dismissed after examination of the series at such values.

15. It may well be, then, that the observed serials, small as they are, have some significance. But evidently the systematic element is slight compared with the random component and a very precise technique is going to be necessary in order to extract it, unless we know what to look for.

Some of my colleagues have pointed out that the operation of first differences could easily reduce quite a large cyclical effect below the threshold of detectability. For example, if there were a ten-year cycle in the data, corresponding to about 500 terms of the series, the differencing of a sine wave of amplitude comparable in magnitude to that of the observed fluctuations would be to reduce the amplitude in the differenced series by a factor of about $2\pi/500$ but to double the variance of random effects. We cannot say, then, that the analysis has ruled out the existence of periodic terms. There may well be traces of a seasonal effect present. What we may fairly say, I think, is that we do not need to invoke them to explain the observations.

16. Series 21 is a derivative of Series 20, being a monthly average instead of a weekly one. A frequency distribution of first differences, which I omit to save space, tells much the same story as Series 20. For the moments I find, with Sheppard's corrections

	Whole Series	*Omitting One Extreme Value at Each End*
$\mu_1' =$	0·0776	0·0504
$\mu_2 =$	44·0256	34·8271
$\mu_3 =$	208·2139	6·4153
$\mu_4 =$	32,721·3887	6,914·5455

For the series on the right

$$\beta_1 = 0{\cdot}00151 \qquad \beta_2 = 5{\cdot}701$$

The distribution is nearer normality, as one might expect under the central limit effect. Some of the serial correlations are given in the reply to the discussion.

17. A comparison of the variances and fourth moments of Series 20 and 21 raises no reason to suspect their wandering nature. If $u_1, \ldots u_{4n}$ is a series of $4n$ weekly terms, the series of monthly terms is $\frac{1}{4}(u_1 + u_2 + u_3 + u_4)$, $\frac{1}{4}(u_5 + u_6 + u_7 + u_8)$, etc. For the differences of these typified by η, we have (means being approximately zero)

$$\text{var}\,\eta = \frac{1}{16(n-1)} \sum_{j=0}^{n-2} (u_{4j+1} + u_{4j+2} + u_{4j+3} + u_{4j+4} - u_{4j+5} - u_{4j+6} - u_{4j+7} - u_{4j+8})^2$$

or, for large n,

$$\text{var}\,\eta = \frac{1}{16n} \sum_{=0}^{n-2} \{(u_{4j+1} - u_{4j+2}) + 2(u_{4j+2} - u_{4j+3}) + 3(u_{4j+3} - u_{4j+4}) + 4(u_{4j+4} - u_{4j+5}) + 3(u_{4j+5} - u_{4j+6}) + 2(u_{4j+6} - u_{4j+7}) + (u_{4j+7} - u_{4j+8})\}^2.$$

If the differences of u, typified by ε, are independent we then have

$$\text{var}\,\eta = \frac{1}{16}\,\text{var}\,\varepsilon\{1^2 + 2^2 + 3^2 + 4^2 + 3^2 + 2^2 + 1^2\}$$

$$= \frac{11}{4}\,\text{var}\,\varepsilon \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (9)$$

The ratio of variances of the monthly to the weekly series is $34{\cdot}8271/12{\cdot}3294 = 2{\cdot}82$ against the theoretical value of $2{\cdot}75$. It may be shown similarly that $\kappa(\eta) = 113/64\ \kappa(\varepsilon)$. For the whole series the ratio $\kappa_4(\eta)/\kappa_4(\varepsilon)$ is $3{\cdot}12$ and for the series omitting extreme values $2{\cdot}12$ against the theoretical value $1{\cdot}77$.

18. There seems nothing to be gained by taking averages of the monthly figures to obtain an annual figure. Under the central limit effect the resulting series would be nearly normal; and successive values would almost certainly be nearly independent.

19. In the past I have often wondered whether annual figures were much use in the study of economic phenomena, except of course in relation to intermittent variables like crop-yields. So much can happen in a year that one feels the underlying causational system to require examination under a finer structure than an annual interval. But if this wheat-series is any guide, it seems that what we gain by observing the phenomenon at short intervals is lost, or at any rate deeply obscured, by the stochastic discontinuity of the process. It may be that the motion is genuinely random and that what looks like a purposive movement over a long period is merely a kind of economic Brownian motion. But economists—and I cannot help sympathizing with them—will doubtless resist any such conclusion very strongly. We can at this point suggest only a few conclusions:

(*a*) the interval of observation may be very important;

(*b*) it seems a waste of time to try to isolate a trend in data such as these;

(*c*) prediction in such a series, from internal behaviour alone, is subject to a wide margin of error and the best estimate of the change in price between now and next week is that there is no change.

British Industrial Share Prices (*Series* 1–19)

20. Series 1 to 19 are weekly index numbers based on 1930 = 100. (The price taken is that of Tuesday of each week, not an average of the week's quotations.) They are not all independently compiled, series 11 (Total industrial production) being an average of series 12–14, series 18 (Total industrial miscellaneous) being an average of series 16 and 17, and series 19 being an average of all classes. There are thus 15 independent series. During the period covered by the figures there were, I believe, some changes such as substitutions for quotations which dropped out. But this hardly affects my argument. Taken together, these are as complete and reliable a set of figures describing the inter-war movement of the U.K. industrial share market as one is likely to find.

21. The first 29 serial correlations were computed for each series and are given in Table 3. The method used was similar to that for series 20 except that corrections for means were omitted as being negligible and the variance of the whole series used in the denominator, the formula accordingly being

$$r_k = \frac{\dfrac{1}{n-k}\sum_{i=1}^{n-k} u_i u_{i+k}}{\dfrac{1}{n}\sum_{i=1}^{n} u_i^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

It may be shown that for these series this approximation can affect the third decimal place at the most, except perhaps for series 3. To resolve any doubt the serials were worked out exactly for this series and the exact values were found to differ from the approximate ones only in the fourth decimal place.

For some of the series a bivariate distribution on the lines of Table 1 was constructed. In some cases the general picture of independence was presented, but in others there were signs of dependence as reflected in the serial correlations. I have space to reproduce only one, that for Series 3, where the dependence seems to be the highest (Table 4).

22. Such serial correlation as is present in these series is so weak as to dispose at once of any possibility of being able to use them for prediction. The Stock Exchange, it would appear, has a memory lasting less than a week except perhaps for Investment Trusts (Series 3), Stores and Catering (Series 14) and all series together. It may be, of course, that series of *individual* share prices would behave differently; the point remains open for inquiry. But the aggregates are very slightly correlated and some of them are virtually wandering. Investors can, perhaps, make money on the Stock Exchange, but not, apparently by watching price-moments and coming in on what looks like a good thing. Such success as investors have seems to be due (*a*) to chance,

TABLE 3

Serial Correlations of Series 1–19, *Lag* 1 *to* 29

(Decimal points omitted)

Number of Series (*see Para.* 4)

Lag	1	2	3	4	5	6	7	8	9	10
1 . .	058	052	301	125	148	087	181	096	088	−013
2 . .	061	−016	356	−001	075	121	055	185	−055	045
3 . .	014	082	158	−025	061	061	053	049	011	−024
4 . .	−050	−019	164	−079	−056	015	−016	086	043	−022
5 . .	−087	−082	066	−025	−055	−030	−084	−005	−015	044
6 . .	−028	−093	−101	−034	−056	−125	−085	−022	−091	−034
7 . .	−015	−006	−030	001	−054	−044	−092	−036	−076	−029
8 . .	−011	−042	−042	−013	−001	−058	−055	001	−020	004
9 . .	004	−025	−030	−041	−084	−118	−067	056	−022	005
10 . .	015	008	−033	010	−063	−067	−012	027	−015	−044
11 . .	045	−050	−013	−022	042	−037	064	025	007	052
12 . .	002	000	−094	048	040	004	000	043	046	004
13 . .	−014	037	−052	045	021	048	040	049	094	009
14 . .	−017	004	009	008	−037	−056	081	067	−060	−025
15 . .	−062	032	027	050	−006	062	047	046	−039	−019
16 . .	030	006	008	118	076	015	039	049	069	002
17 . .	072	079	109	030	068	153	041	119	076	043
18 . .	097	049	086	045	030	070	018	028	−028	036
19 . .	098	−016	080	−016	031	−017	−009	−055	−036	−015
20 . .	−037	−077	020	−026	−046	072	−023	016	−033	037
21 . .	009	060	031	074	−031	−078	−056	052	004	−021
22 . .	−027	011	−056	044	−067	−094	−029	074	−023	076
23 . .	−044	−151	−047	056	−004	−084	−029	−041	−091	−048
24 . .	034	−005	−086	−023	−020	−017	−001	043	−049	017
25 . .	104	017	031	069	−039	015	016	133	076	−009
26 . .	021	080	−048	038	−001	−012	029	067	044	−032
27 . .	000	004	011	−017	082	−046	094	044	032	−052
28 . .	045	130	053	071	097	−011	036	036	040	087
29 . .	−011	−040	050	−014	060	−014	033	037	061	031

Number of Series

Lag	11	12	13	14	15	16	17	18	19
1 . .	195	010	053	230	237	034	200	177	234
2 . .	061	−057	029	054	076	087	103	106	105
3 . .	053	044	−044	018	054	003	053	058	066
4 . .	−034	039	−014	001	−040	−056	−013	−044	−046
5 . .	−051	−015	−034	−084	−062	067	−080	−044	−055
6 . .	−102	−081	058	−146	−093	−002	−115	−079	−094
7 . .	−072	018	−013	−144	−072	019	−104	−087	−093
8 . .	−043	139	036	−057	−044	−114	004	−033	−041
9 . .	−047	033	037	−038	165	−018	−003	−003	−008
10 . .	−047	−044	021	042	021	035	012	−091	−023
11 . .	−003	079	046	−006	013	085	−014	013	005
12 . .	039	069	044	086	089	017	024	142	050
13 . .	082	−009	−009	−047	−006	077	037	064	045
14 . .	−029	−041	−107	004	−008	−036	−037	−029	−025
15 . .	036	−129	096	067	018	−084	045	001	033
16 . .	069	030	090	064	077	055	158	122	114
17 . .	145	005	168	029	086	001	124	120	148
18 . .	047	072	063	047	082	029	089	079	084
19 . .	−010	013	042	−001	057	047	010	022	038
20 . .	007	005	008	035	014	−078	−001	−045	−017
21 . .	−018	−012	−030	090	043	041	009	031	008
22 . .	−035	009	−027	032	013	064	−015	033	−006
23 . .	−082	−056	022	−039	−082	−005	−053	−045	−070
24 . .	−041	−032	−037	−034	−021	028	−008	−008	−028
25 . .	051	036	−000	002	−001	058	−001	017	026
26 . .	025	037	086	−038	−005	046	036	049	036
27 . .	047	055	092	−032	052	−019	030	012	038
28 . .	168	048	016	−037	007	023	072	063	075
29 . .	−069	039	−054	−004	032	−014	018	001	014

TABLE 4

Bivariate Distribution of First Differences of Series 3, *Difference between Weeks* t *and* t + 1 *against Difference between Weeks* t + 1 *and* t + 2; *Units, One-tenth of a Point*

Difference, tth week, less $(t+1)$th week	Difference t + 1th Week Less t + 2th Week: < −12	−12, −11	−10, −9	−8, −7	−6, −5	−4, −3	−2, −1	0, 1	2, 3	4, 5	6, 7	8, 9	10, 11	⩾ 12	*Totals*
< −12	5	2	4	5	1	2	—	—	1	—	—	—	—	1	21
−12, −11	4	1	1	—	3	1	3	—	—	—	—	—	—	—	13
−10, −9	1	2	2	—	2	2	4	—	—	1	—	1	—	—	15
−8, −7	3	2	—	4	4	5	4	—	1	—	1	—	—	1	25
−6, −5	4	3	2	4	7	11	9	3	1	1	—	—	—	1	46
−4, −3	1	2	2	6	7	17	13	14	2	—	1	1	—	—	66
−2, −1	1	1	1	2	9	19	14	18	5	5	2	2	1	1	81
0, 1	—	—	—	1	6	8	22	26	13	9	3	2	1	—	91
2, 3	—	—	2	1	1	—	5	15	9	4	3	3	—	1	44
4, 5	1	—	—	1	3	—	3	7	7	5	1	1	2	1	32
6, 7	—	—	—	—	2	1	—	3	1	2	—	2	1	2	14
8, 9	—	—	1	—	1	—	1	2	1	3	2	—	—	4	15
10, 11	—	—	—	—	—	—	1	2	2	1	—	1	—	1	8
⩾ 12	—	—	—	—	—	—	1	1	1	1	—	2	3	1	14
Totals	21	13	15	25	46	66	81	91	44	32	14	15	8	14	485

(*b*) to the fact that at certain times all prices rise together so that they can't go wrong, (*c*) to having inside information so that they can anticipate a movement, (*d*) to their being able to act very quickly, (*e*) to their being able to operate on such a scale that profits are not expended in brokers' fees and stamp duties. But it is unlikely that anything I say or demonstrate will destroy the illusion that the outside investor can make money by playing the markets, so let us leave him to his own devices.

23. There are several factors of these series of correlations requiring explanation.

(*a*) If we arrange them according to the magnitude of the first serial correlation, which seems a fair summary measure of internal correlation, we get the following order:

Series	3	Investment Trusts.
	15	Total Distribution.
	19	All classes.
	14	Stores, etc.
	17	Miscellaneous.
	11	Total Production.
	7	Electric Light.
	18	Total Miscellaneous.
	5	Coal.
	4	Building Materials.
	8	Gas.
	9	Iron and Steel.
	6	Cotton.
	1	Banks.
	13	Shipping.
	2	Insurance.
	16	Breweries.
	12	Home rails.
	10	Oil.

We can understand why this list should be headed by investment trusts, which are almost an aggregative index in themselves, and why it should be tailed by shipping, breweries, home-rails and oil. But it is not clear why banks and insurance companies come so low.

(*b*) Apart from investment trusts and stores, the series showing the greatest internal correlations are the aggregative series 11, 15, 18 and 19. We get greater serial correlation in the averages than in the constituent series, which at first sight seems absurd and in any case is very misleading. One possibility is that this effect is generated by the method of construction of the series, e.g., the use of geometric means, the substitution of new quotations and so on. With the help of Mr.

S. T. David and Mr. Haycocks I went into this explanation fairly thoroughly but it does not seem to account for the phenomenon. The use of geometric means could raise or lower the serial correlations but need not bias them upwards; substitution of new quotations is carried out only in December and its effect, even if systematic, would be so diluted in the correlation of weekly figures as not to explain the magnitude of the aggregation effects. A more likely explanation lay in the existence of lag correlations between the series and I revert to the point below in paragraph 28.

24. Whatever the reason, the existence of these serial correlations in averaged series is rather disturbing. If the effect is a general one, it means that we must be very chary of drawing inferences from series of index numbers of the aggregative type, which is most unfortunate because such series are, in many instances, all that the statistician or the economist has to work on. The so-called "cycles" appearing in such series may not be due to endogenous elements or structural features at all, but to the correlations between disturbances acting on the constituent parts of the aggregative series.

25. I am led to infer that wherever possible the econometrician must study individual series rather than aggregates, just as he must study closely neighbouring points of time rather than observations at long intervals. This may be part of the reason why attempts to fit simple models to a whole economy have usually failed. It is as though a physicist set out to investigate the properties of light without being able to isolate a pure colour and with a diffraction grating ruled at intervals of an inch apart. Such conclusions would apply to price data, or to any nundinal data where there is a rapid adjustment of forces in a fluid market. They may not apply with equal force to production data.

Cross-correlation of the Share-prices

26. To study the series further we have to examine the correlations between them. It is possible to pick out pairs from 19 individuals in 171 ways and even with modern computing facilities it was too laborious to work out a number of lag correlations for them all. I chose 28 pairs, taking those for which the analysis seemed likely to be most rewarding, as follows:

(1, 4), (1, 5), (1, 6), (1, 9), (1, 12), (1, 13)
(4, 5), (4, 6), (4, 8), (4, 12), (4, 13)
(5, 6), (5, 8), (5, 9), (5, 12)
(6, 8), (6, 9), (6, 11), (6, 12)
(8, 11), (8, 18)
(9, 11), (9, 12), (9, 13)
(11, 12), (11, 13), (11, 18)
(12, 13)

The numbers correspond to the series set out in section 4. The lags even worked out as far as the sixth each way, e.g., (1, 4) means that lag correlations were worked out between first differences of series 1 and 4 (Banks, etc., and Building Materials) for lags of -6 to 6, thirteen coefficients in all. The results are given in Table 5.

27. The main feature of this set of correlations is the smallness of the magnitudes for the lags. There are traces of correlation in some instances but, however real, they are very slight. Thus no series acts as a "leader" for the others. Not only is it impossible to predict a series from its own internal behaviour but it seems equally impossible to predict it from the behaviour of the other price series.

28. We can now revert to the aggregation effect mentioned in paragraph 23 and discuss it in simplified terms. Suppose we have a series of consecutive pairs of terms $(\varepsilon_1, \eta_1)(\varepsilon_2, \eta_2) \ldots (\varepsilon_n, \eta_n)$ with zero means. Consider the correlation of sums $E \equiv \sum_{i=1}^{n} \varepsilon_i$ and $F \equiv \sum_{i=1}^{n} \eta_i$. If the variance of ε_i and η_i is σ_i^2 we find

$$\operatorname{corr}(E, F) = \frac{\sum_{i,j} \sigma_i \sigma_j \operatorname{corr}(\varepsilon_i \eta_j)}{\sum \sigma_i \sigma_j \operatorname{corr}(\varepsilon_i, \varepsilon_j)} \qquad i, j = 1 \ldots n \quad . \quad . \quad . \quad (11)$$

TABLE 5

Lag Correlations of the Differences of Certain Pairs of Series from Series 1–19

	Lag												
Series	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6
(1, 4)	−008	−066	−077	−021	036	028	414	100	015	008	023	−080	−125
(1, 5)	−060	−134	−137	−068	015	030	255	077	102	−011	015	011	−066
(1, 6)	−080	−037	−074	−087	019	029	228	126	163	010	062	−010	−118
(1, 9)	−023	−102	−139	−014	−027	−021	295	053	050	028	−020	027	−028
(1, 12)	−037	−023	−046	−029	017	−044	275	154	073	0003	−001	022	−118
(1, 13)	−033	−068	−060	−124	113	051	258	121	030	−015	004	−063	−070
(4, 5)	−064	−108	−102	002	034	070	425	200	035	052	−044	−019	016
(4, 6)	−059	−036	−047	−003	059	058	424	201	076	054	−027	−003	−089
(4, 8)	−037	−067	−015	−024	049	085	310	−002	023	030	006	−027	−046
(4, 12)	022	−071	−049	026	−009	−006	367	187	−006	059	031	018	−066
(4, 13)	048	−052	−033	054	046	111	320	096	019	002	−068	001	010
(5, 6)	−080	009	051	083	048	182	454	161	081	054	−043	−108	−134
(5, 8)	−013	−013	026	00001	044	009	272	004	025	035	−056	−042	007
(5, 9)	−075	−016	−065	070	013	096	661	226	033	024	015	−065	−053
(5, 12)	026	−014	−036	074	037	063	392	167	024	−013	025	−048	−072
(6, 8)	−040	−033	037	−004	064	024	194	014	−031	012	−003	032	−072
(6, 9)	−122	083	004	078	038	085	482	228	030	013	071	024	−115
(6, 11)	−127	−073	004	092	104	156	750	140	080	033	018	011	−124
(6, 12)	−095	−034	−031	110	014	089	368	194	−027	011	−024	033	−114
(8, 11)	−020	−049	−027	069	002	006	389	066	075	−002	023	−046	−030
(8, 18)	−033	−064	−032	055	−005	−046	399	079	125	002	010	−051	−028
(9, 11)	−087	−004	042	022	022	228	795	071	−032	033	−038	−060	−101
(9, 12)	−056	−025	041	039	008	016	469	110	−070	060	042	−048	−149
(9, 13)	007	−031	026	133	092	144	279	102	−037	047	−111	−029	−002
(11, 12)	−053	−035	−060	072	010	054	522	224	−024	059	036	−004	−120
(11, 13)	066	−076	038	082	042	199	407	140	041	011	−056	−067	021
(11, 18)	−051	−004	007	036	094	196	824	166	069	065	−108	−084	−101
(12, 13)	−039	−022	044	041	039	146	281	045	031	001	−085	−008	−031

Note.—Decimal points omitted. Lag counted as interval of first series after second, e.g., lag 2 for series (1, 4) is correlation of series 1 at time $t + 2$ and series 4 at time t.

or, if all the variances are equal,

$$\text{corr}\,(E, F) = \frac{\text{mean corr}\,(\varepsilon_i, \eta_j)}{\text{mean corr}\,(\varepsilon_i, \varepsilon_j)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

For my series I do not possess all the lag correlations but as an example of the kind of thing which may happen let us suppose that all the correlations $(\varepsilon_i, \varepsilon_j)$ have a mean value of $0{\cdot}4$ (this being about the mean of those values we have in Table 5) and that the mean correlation of (ε_i, η_j) is $0{\cdot}1$. Then we find

$$\text{corr}\,(E, F) = \frac{225\,(0{\cdot}1)}{15 + 210\,(0{\cdot}4)} = 0{\cdot}23$$

so that the first serial of the aggregate series is more than twice the mean serials of its constituents. It is a piece of good luck, I suppose, that the first serial of series 19 turns out to be $0{\cdot}234$.

29. The cross correlations of order zero suggest that there is a good deal of *simultaneous* sympathetic movement between the series; (simultaneous, that is, within the compass of a single week). They are more or less as one would expect but not always so big as a simple model of the economy would suggest. If we regard these random changes from week to week as due to exogenous elements, we must suppose that the elements are substantially different from one series to another, albeit not independent.

30. One could clearly take this analysis further, for instance by working out multiple regressions or attempting a component analysis to see whether there were any common factors. I have refrained from doing so at this stage for two reasons; one is that further experiments on other

price series and for different intervals of time are desirable before we refine the analysis; the other is that the existence of the aggregative effect would probably vitiate the findings and I think we shall have to go back to primary series before attempting to isolate components.

New York Cotton Prices

31. The New York spot cotton prices of series 22 are a salutary warning against undue generalization. They are monthly, not weekly, series and cover about 125 years. One might have expected them to behave like the wheat series, but they do not. Table 6 gives the serial correlations up to the sixteenth order. As there were obvious breaks in continuity during the

TABLE 6

Serial Correlations of Cotton Prices (Series 22)

(Decimal points omitted)

Order of Correlation	*Period* 1816–1860	*Period* 1868–1914	*Period* 1921–1950	*Total Series* (1816–1950)
1	346	393	227	313
2	096	046	−012	039
3	−049	−096	−027	−053
4	−045	−153	−172	−125
5	−075	−060	024	−032
6	−053	−075	095	−000
7	−013	−068	031	−010
8	015	−064	135	041
9	−037	−028	021	−012
10	001	071	122	068
11	029	088	−000	033
12	−051	080	−106	−038
13	−093	−006	024	−014
14	−143	032	047	012
15	−105	−011	025	016
16	−022	−050	042	008

American Civil War and during World War I, I have calculated the serials separately for the periods 1816–1860, 1868–1914 and 1921–1950.

The bivariate distributions of the type of Table 1 are regular, looking very much like a normal form, but the correlations are significant. We now find a pattern of behaviour rather like that of a simple Markoff series. The price-change from month t to month $t + 1$ is now clearly correlated with that from month $t + 1$ to month $t + 2$, but we do not require a memory of more than a month in the market to account for the serial correlation.

There are several possible explanations of this effect, although it is difficult to be sure whether any of them is the right one. Wheat is a commodity which is grown in both hemispheres and a continual supply on the market may keep the price more fluid than for cotton, a product of the northern hemisphere only. Moreover, crop forecasts are better for wheat than for cotton and there is a tendency for the latter to be underestimated at the beginning of the season; the slow improvement in estimates as the growing season proceeds might induce some correlation in the prices. The slower rate of progress of cotton through the manufacturing process may also have some effect. But whatever the reason, it seems that we are not entitled to generalize from one agricultural commodity to another and a systematic survey of the major raw materials of commerce is necessary before we can lay down any general rules.

First Experiments in Trend Fitting

32. The typical Yule autoregressive scheme which has been fitted to economic data with a certain amount of success may be written

$$\alpha u_{t+2} + \beta u_{t+1} + \gamma u_t = \varepsilon_{t+2} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

It seems natural to generalize this by considering the case when the constants α are themselves slowly moving through time as the economy changes. I therefore examined the scheme

$$(\alpha_0 + \alpha_1 t + \alpha_2 t^2)\, u_{t+2} + (\beta_0 + \beta_1 t + \beta_2 t^2)\, u_{t+1} + (\gamma_0 + \gamma_1 + \gamma_2 t^2)\, u_t = \varepsilon_{t+2} \quad . \qquad (14)$$

A second-order Yule scheme was chosen because general experience so far has indicated that nothing much is to be gained by taking higher orders into account; and second-degree polynomials were chosen because the general run of the series suggested that they would be sufficient and would keep the arithmetic within reasonable bounds. But a decision on these points is rather arbitrary and more extended schemes might be necessary in some cases. The fitting was carried out on the primary series not on the differences.

33. In equation (14), it will be noted, I have allowed the coefficient of u_{t+2} to vary with time. It would be more consonant with the idea of auto*regression* to take the coefficient as a constant, say unity. However, the more general expression has one advantage, namely that it allows for movements in the variance of the disturbance term ε, which may well vary in time, especially for price data. If the coefficients are small compared with the maximum observed value of t^2 equation (14) may be written

$$u_{t+2} + (\beta_0' + \beta_1' t + \beta_2' t^2)\, u_{t+1} + (\gamma_0' + \gamma_1' + \gamma_2' t^2) = (\delta_0 + \delta_1 + \delta_2 t^2)\, \varepsilon_{t+2} \quad . \qquad (15)$$

which exhibits the scheme as an approximation to a generalized Yule process with trend in the variance of the disturbance.

34. It will also be noticed that (14) is not a stationary process. There is no reason why it should be. No economic system yet observed has been stationary over long periods. It is true that schemes of type (14) may be explosive, which all economic systems are not; but that only implies that in representing the slow movements by polynomials we are approximating to some non-explosive function, a familiar procedure to which no exception is taken in other fields of application.

35. I fitted a scheme of type (14) by least squares, that is to say, by minimising the square of the expression on the left for variations in the nine parameters. This leads to nine equations with coefficients of the type $\Sigma\, t^2 u^2_{t+2}$, $\Sigma\, t\, u_{t+1} u_{t+2}$, etc. They are troublesome to compute but there is no theoretical difficulty. For the experiment series 1, 6 and 8 (Banks, etc., Cotton and Gas) were taken. The following are the results. The origin in each case is at the 244th term so that t ranges from -243 to 242.

Series 1

$$(1 + 0\cdot021{,}279{,}208t - 0\cdot0^4 31{,}214t^2)\, u_{t+2} + (-1\cdot236{,}892{,}571 - 0\cdot0^2 2{,}127{,}249t$$
$$+ 0\cdot0^4 40{,}297t^2)\, u_{t+1} + (0\cdot236{,}578{,}860 + 0\cdot0^3{,}848{,}707t - 0\cdot0^5 9{,}021t^2)\, u_t = \varepsilon_{t+2} \;. \qquad (16)$$

Series 6

$$(1 - 0\cdot0^2 3{,}435{,}032t + 0\cdot0^3{,}135{,}958t^2)\, u_{t+2} + (-0\cdot086{,}729{,}101 - 0\cdot0^4 95{,}266t$$
$$- 0\cdot0^4 54{,}730t^2)\, u_{t+1} + (-0\cdot921{,}767{,}317 + 0\cdot0^2 3{,}562{,}777t - 0\cdot0^4 81{,}816t^2)$$
$$= \varepsilon_{t+2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (17)$$

Series 8

$$(1 + 0\cdot0^2 1{,}273{,}056t - 0\cdot0^4 37{,}456t^2)\, u_{t+2} + (-1\cdot421{,}129{,}272 - 0\cdot0^2 2{,}055{,}170t$$
$$+ 0\cdot0^4 56{,}754t^2)\, u_{t+1} + (0\cdot422{,}070{,}596 + 0\cdot0^3 822{,}663t - 0\cdot0^4 19{,}549)\, u_t = \varepsilon_{t+2} \;. \qquad (18)$$

36. These equations were worked out before it had dawned on me that the parent series were behaving in a wandering way, and at that stage it was a surprise when Mr. K. H. Medin, working in my department, pointed out that there were identities in the coefficients. In fact, very approximately

$$\left.\begin{aligned} \alpha_0 + \beta_0 + \gamma_0 &= 0 \\ \alpha_1 + \beta_1 + \gamma_1 &= 0 \\ \alpha_2 + \beta_2 + \gamma_2 &= 0 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (19)$$

For instance in (16)

$$1 - 1 \cdot 236{,}892{,}571 + 0 \cdot 236{,}578{,}860 = 0 \cdot 0^3 313{,}911$$

and in (18)

$$- 0 \cdot 0^4 37{,}456 + 0 \cdot 0^4 56{,}754 - 0 \cdot 0^4 19{,}549 = - 0 \cdot 0^6 251$$

It follows, of course, that we can write (12) as

$$(\alpha_0 + \alpha_1 t + \alpha_2 t^2)(u_{t+2} - u_{t+1}) - (\gamma_0 + \gamma_1 t + \gamma_2 t^2)(u_{t+1} - u_t) = \varepsilon_{t+2} \quad . \quad . \quad (20)$$

so that first differences obey a generalized Markoff scheme. The ratio

$$(\gamma_0 + \gamma_1 t + \gamma_2 t^2)/(\alpha_0 + \alpha_1 t + \alpha_2 t^2)$$

varies to some extent over the range of t; for Series 1 from about $0 \cdot 3$ at one extreme to $0 \cdot 2$ at the other, being about $0 \cdot 2$ over the middle range; for Series 6 from about $- 0 \cdot 7$ to $- 0 \cdot 5$, about minus unity over the middle range; for Series 8 about $0 \cdot 4$ over the middle range.

On the whole I regard this experiment as a failure. It seems asking too much of quadratic polynomials to expect them to give a good representation over series of 500 terms; and in any case it is doubtful whether the series are sufficiently systematic to react significantly to this kind of treatment. The only reason I refer to the trend-fitting problem at all at this stage is to call attention to its difficulty in the hope that others may be led to study it.

37. I am well aware that this paper raises more difficulties than it resolves. Most papers on time-series do. The results concerning the time-interval of observation, the aggregative-effects and trend-fitting seemed to me, however, to be important enough to justify publication, if only to prevent my fellow-workers from spending time on profitless inquiries. More positive results must await further exploration of individual series, not only of prices but of production, and attempts to improve available statistical techniques for analysing them.

Acknowledgments

38. The computations involved have been severe, particularly in connection with the lag correlations of series 1 to 19. I am indebted to Mrs. Joan Humphries and Miss Julia Grahame for a great deal of the work, which they carried out with zeal and gratifying cheerfulness. I am also indebted to Mr. E. C. Fieller and the National Physical Laboratory for allowing me to use the electronic A.C.E. in determining serial covariances. My colleagues in the Division of Research Techniques at the London School of Economics, particularly Mr. S. T. David and Mr. K. H. Medin, have been very helpful in exploring lines of inquiry, many of which are referred to only very briefly or are not mentioned at all, that had to be followed up to decide incidental points; and my colleagues among the economists have endured much in discussions about the econometric implications of the results.

Discussion on Professor Kendall's Paper

Professor R. G. D. Allen: In moving this vote of thanks to Professor Kendall I should like to express the indebtedness of the Society not only to him but also to the Division of Research Techniques at the London School of Economics of which he is the Director. The Division, only a few years old, has been generously supported by the Nuffield Foundation. The publications of the Division are numerous, massive, and indicative of a long life for the Division. The work has not been at all concentrated; it has ranged over a surprisingly wide field. Some of the work on survey techniques has already been presented to this Society. Other work has been in the econometric field, as represented by the paper Professor Kendall has read tonight. This contribution is a generally understandable one. Not all econometric work is as intelligible to the layman or even to the expert. I hope that all economists as well as statisticians will read the paper and study it.

I want to take the point of view of the economist, not the statistician, who approaches the problem of handling time-series rather hopefully and at the same time apprehensively. Economists have learned much on this subject during the two decades or more since Yule's famous paper was read to the Society. They have come to look upon time-series with considerable suspicion. They have been told to beware of a fearsome devil—serial correlation—and to look to statisticians to cast the devil out for them. It seems now that the more statisticians work on the casting out of devils, the more they cast out everything else, including what the economists want.

If the economist is to design a model to relate to actual data, he must deal in averages and aggregates, e.g., such broad aggregates as national consumption, income, savings, investment. Moreover, the testing of models in the past has usually been in the terms of annual data. The question is how to break up aggregates and how to shorten the intervals of recording in the time-series. Short annual series of data are usually not good enough to test any model whatever, and what Professor Kendall now appears to show is that, with more frequent data, many different models fit the data very well. This is a serious dilemma.

In para. 19 of the paper which indicates Professor Kendall's conclusion on wheat prices he points out that what we gain by observing the phenomena at short intervals is lost or obscured by "the stochastic discontinuity of the process", that it may be that the motion is genuinely random, and that what looks like a purposive movement over a long period is "merely a kind of economic Brownian motion". That is a very depressing kind of conclusion to the economist. The conclusions on cotton prices (para. 31) are rather more encouraging, a "memory" of a month. Even so, this may arise because of defects in the cotton market as opposed to the wheat market. Cotton is practically confined to the northern hemisphere and the making of crop estimates is a rather peculiar process. Finally, on stock prices, apart from the question of how to "play the market", we would expect to have dependable aggregate series of prices of all stocks together. Paragraphs 24 and 25, however, indicate that the problem is not a simple one.

In the present state of knowledge the economist must work with simple models relating to broad aggregates. If he ventures further the variables and relations—and the problems—crowd in on him. Professor Kendall's conclusion is that any attempt to fit simple models to data must inevitably fail. The econometrician is far from attaining the position in which he can hope to determine the parameters in which he is interested. This paper must be regarded as the first dividend on a notable enterprise. Some "shareholders" may feel disappointed that the dividend is not larger than it is, but we hope to hear more from Professor Kendall and to have further, and larger, declarations of dividends.

Professor Champernowne (in seconding the vote of thanks): Professor Kendall is a pioneer in the study of time series by examining serial correlation coefficients, and a paper from such an authority gives us hope for further revolutionary advances. Every hint that old techniques are faulty and every suggestion of new methods of analysis will eagerly be followed by those who have time and resources to experiment with new ideas. Therefore the more conservative amongst us, especially amongst the economists, may be excused if we endeavour to pick holes in the argument before it is allowed to become gospel. As seconder of the vote of thanks I feel an obligation in this respect, and shall devote my attention almost entirely to the points on which I disagree.

Turning, however, to the list of conclusions given in para. 3, I do not find that I disagree violently with anything except perhaps the last sentence of (*d*), that the more systematic behaviour of aggregative index numbers may be due to chance, and that if so there will appear spurious time-correlations in aggregative series, and the use of index numbers in econometric work will need extensive reconsideration. I hope that those responsible for using or criticizing or constructing index numbers will not be led by these conclusions to suppose that there is something wrong about index numbers, and that they should use individual quotations or do something more drastic instead. Unless I misunderstand Professor Kendall's arguments, he is drawing a red herring across any discussion of the merits of index numbers for the purposes for which they are usually intended.

He has shown that if in a set of price-series the intercorrelations lagged by one time-unit are nearly all positive, and if the unlagged correlations are not large, then the serial correlation coefficient with unit lag for the index number may be much larger than for any individual series. If the phenomenon is due to a cancelling-out of random elements he and all of us would agree that this merely shows that the index number expresses the systematic part of the price-movements better than any individual series. But Professor Kendall suggests that the effect may be entirely spurious and due merely to chance. This is possible, but it seems to me that pure chance alone will only on very rare occasions lead to the intercorrelations of lag 1, nearly all having positive sign and an average nearly as high as that of $0 \cdot 1$ found in Table 5. This does not look like chance; it indicates that there is a tendency towards a positive correlation between one term in one series and the following term in another. If so, the higher serial correlation of the index number reflects its advantageous quality and is not spurious.

With reference to conclusion (*e*), I may perhaps be excused for agreeing with the moral drawn there. The professionals have to keep a sharp eye on the movement of prices, on the one hand stifling large shifts of price that are likely to be immediately reversed, and on the other anticipating any movements of prices that would otherwise continue over a period of weeks; and the low serial correlation coefficients found in this particular series may reflect the success with which the professionals are doing their job. It is reassuring that statisticians will not be able to earn vast sums by scientific gambling on the Stock Exchange, and will still find it worth while to continue their own work. I cannot resist saying that the analysis of prices of less speculative markets will have even greater interest for the economist.

Turning to conclusion (*c*), again it is difficult to disagree. It may be impossible to distinguish between different hypotheses which all fit the data, and the choice of possible hypotheses by prior analysis is, in the present state of knowledge, as essential as the testing and comparing of hypotheses by statistical methods. But there is a danger that people reading this conclusion will think that because Professor Kendall finds certain speculative price series have correlograms like those of wandering series, *ergo* it is futile to compare the trend, of say, real wages with the trend of other variables which, according to economic theory, determine real wages in the long run. Leaving aside the question whether seasonal and cyclical variations of prices have any precise meaning, there can be little doubt that the graphs of many price series exhibit such movements. If one is merely analysing the series by taking weekly changes in such prices and drawing up tables like Table 3 of the paper, I doubt whether the results would be so different from those of that table, my doubt being based on Professor Kendall's own admirable argument in the second part of para. 15, where he says that one does not need to invoke seasonal or cyclical changes to explain the observations. The danger is that this might be interpreted by some as meaning that cyclical and seasonal influences are not an important part of the movement of many economic time series. The fact that the correlograms of many economic series are like those of wandering series need not disturb us. If this effect is widespread it should give greater confidence in applying regression analysis to first differences of series. For example, even if the wholesale price and the retail price of a commodity, each regarded separately, behave in a manner indistinguishable from the wandering series, it might still be demonstrated that the relation between the two could not be best explained on the basis of chance. Here again I wish to defend Professor Kendall from those who might possibly misinterpret his own conclusions.

I fear that, after all, in what I have said I have failed to attack and have merely rushed in to defend the author against his more enthusiastic admirers, who may overstate some of his more spectacular claims and suggest that we should all study the correlograms of individual time series instead of the equations connecting them. This cannot be Professor Kendall's intention: indeed he has foreshadowed a further paper on the relations between price series and production series, to which we shall look forward with keen interest.

Mr. Quenouille: Professor Kendall's paper is, as other speakers have indicated, a provocative one. Professor Kendall has dealt with such a wide diversity of subjects that, upon first reading this paper, I was at rather a loss as to where to start any investigation of his analysis.

As a first step, I decided to see what happened if the traditional autoregressive series were taken at a somewhat shorter interval than usual. I therefore assumed that the damping factor was $1-\varepsilon$ and that the period was θ, where both ε and θ are small. The first two serial correlations of the differences of this series would then be expected to be of the order

$$-\frac{\varepsilon}{2}+\frac{\theta^2}{2\varepsilon+\theta^2-2\varepsilon\theta}.$$

It is obvious from this formula that if ε and θ are both small, we should expect to have small serial correlations.

My next step was to substitute values for ε and θ in this formula. The first serial correlation coefficient for Kendall's series No. 21 calculated using the figures for the first week of every year was 0·66, and, since this was not far from the corresponding correlation coefficient of Kendall's artificial series $u_n - 1{\cdot}1\, u_{n-1} + 0{\cdot}5\, u_{n-2} = \varepsilon_n$, I decided to use the latter series in the further analysis. If this series were observed at monthly intervals, i.e., at one-twelfth the interval normally used, the damping factor would be 0·97 and θ would be 0·055 roughly. With these two values, the first serial correlation works out at 0·04. For narrower spacing, at weekly intervals, it works out at 0·01.

These results come out much as might be expected. Basically, the use of first differences of the original series gives rise to quantities which are proportional to the second differences of the correlogram. If the interval is small, these differences are correspondingly small, and so are the serial correlations. It is impossible to distinguish between different types of correlogram when we are working only with second differences of values from near the origin. It is not surprising that Professor Kendall concludes that the data behave almost like wandering series.

The above formula also throws some light on what happens when autoregressive schemes with the same random element are added together. It is fairly easy to demonstrate that this gives rise to an autoregressive scheme of higher order. Thus, the addition of two Markoff processes gives rise to a non-oscillatory second-order process (one for which θ^2 is negative). These higher processes will usually have constants such as θ, which will tend to increase the serial correlation coefficients of their differences. To put it simply, the increased complexity of the correlogram will increase the second differences. These results seem to indicate that Professor Kendall's conclusions are similar to those that would be reached by taking the usual autoregressive schemes at narrower intervals.

Concerning Professor Kendall's method of taking first differences, I remember when I first started at Rothamsted being shown a set of data of catches of insects in a light trap which was said to have been analysed by Professor Fisher. The method of analysis involved the use of differences between successive catches. It is interesting that this method is coming into use in the economic field to an increasing extent. Dr. Orcutt has been advocating it for some time, and he deserves some acknowledgment for his foresight. It is a very useful method, but it is necessary to beware of its limitations.

Obviously, it is restricted to series for which the serial correlations are high, so that if we carry out too complicated an analysis with the first differences there is a danger that we shall get to a position where we are, in effect, dealing with differences at a wider spacing, and our analysis may consequently be invalid. In general it will be valid, although there will be a limitation placed upon the narrowness of spacing by the errors of observation (in the simplest case by the errors of rounding-off) which will, if the observations are too close, swamp the normal random element. Also we shall run into problems of lag correlation for narrower intervals of spacing, for while it may be reasonable to assume that two events are simultaneous when our observations are yearly averages, it will often be altogether unreasonable when they are daily or weekly figures.

As to methods of dealing with trends in time series, I agree that it is difficult to distinguish between time trends and highly-correlated stochastic elements, but I do not think the position is as bad as Professor Kendall indicates. There are several rough and ready methods for use in this connection. One of these fits polynomials, examines the residuals in the light of the fitting, re-examines the fitting, and so on. A second method involves the use of moving averages (and includes the use of first differences as a special case) and the correction, if necessary, of the residuals to allow for the effect of the moving average. This is rather difficult, since initially it is desirable to maintain an open mind upon the form of average required to eliminate the trend. Yet another method employs the use of the spectrum of a series to discover its stochastic nature. This can be done, since trend is removed with the components of low periodicity.

We have had a stimulating paper, and the subject has benefited from the careful investigation carried out by Professor Kendall.

Mr. S. J. Prais: I wish to make three points. The first is that the reader must be grateful for the opportunity of examining analyses of really long economic time-series in which large sample theory can be used without the usual qualms. The use of an electronic computer to carry out the heavy numerical work is also noteworthy, and I wish to add my appreciation to that of earlier speakers.

Secondly, there is the question of the relevance of the type of autoregressive equation, the parameters of which Professor Kendall has been trying to estimate, to the type of dynamic theory economists usually deal with.

A dynamic economic system consists of a number of equations expressing relationships (generally demand or supply relationships) between economic variables, some of which at least involve time in an essential way in the form of time derivatives or time integrals. Since observations of economic phenomena are well separated in time, these relationships may without any considerable loss of generality be expressed in the form of difference equations. In a matrix notation such a complete set of equations may be written as

$$\mathbf{B}\,\mathbf{y} = \boldsymbol{\epsilon}, \qquad (1)$$

where $\mathbf{y}$ is a vector of n variables and $\mathbf{B}$ is a square matrix, the elements of which are polynomials in the delaying operator E^{-i}. The relationships are not supposed to be exact, so there is written on the right-hand side of the equation a vector of random variables ε with means zero. It is this type of economic relationship that economists usually investigate.

Now, in the usual way, if

$$\boldsymbol{\epsilon} = \mathbf{0}, \qquad (2)$$

the path in time of every economic variable in the system would be given by the same characteristic equation,

$$|\,\mathbf{B}\,|\;y = o, \qquad (3)$$

where $\mid \mathbf{B} \mid$ is obtained by expanding the coefficients of the system (1) as a determinant, and (3) may be written in the form of a difference equation

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots a_p y_{t-p}. \qquad (4)$$

Since in these circumstances the same difference equation will apply to all variables it is possible to investigate the dynamic properties of the system by examining the fluctuations of any one variable over time, no matter which variable is chosen. It is assumed that an argument of this kind lies at the base of Professor Kendall's investigations.

If, however, the equations of the system (1) are not exact, the autoregressive equation of the system will also not be exact. Further, each variable will have a different path over time; for (1) may then be "solved" for each variable in the form

$$\mid \mathbf{B} \mid \; y_i = \mathbf{B}'_i \boldsymbol{\epsilon} = \eta_i, \qquad (5)$$

say, where $\mathbf{B}_i$ is the vector of co-factors of the i^{th} column of $\mathbf{B}$. While the variances of the residuals of the behaviour equations (1) may be small so that good dynamic theory is possible, the variance of the residual of (5) may be very large; for, if the variance matrix of the $\boldsymbol{\epsilon}$ is $\sigma^2 I$, the variance of η_i is given by

$$V(\eta_i) = \mathscr{E}\{\eta_i^{\,2}\} = \mathscr{E}\{\mathbf{B}'_i \, \boldsymbol{\epsilon} \, \boldsymbol{\epsilon}' \, \mathbf{B}_i\} = \sigma^2(\mathbf{B}'_i \, \mathbf{B}_i), \qquad (6)$$

and the increase in variance depends on the size of the cofactors $\mathbf{B}_i$ or, generally, on the degree of ill-conditioning of the system.

As a simple example a Keynesian-type system was considered:

$$C = \alpha E^{-1} Y + \varepsilon_1 \qquad (7a)$$

$$I = \beta E^{-1} Y + \gamma E^{-2} Y + \varepsilon_2 \qquad (7b)$$

$$Y = C + I, \qquad (7c)$$

where C, I and Y are consumption, investment and income; and α, β and γ are parameters. The autoregressive equation for consumption is

$$[1 - (\alpha + \beta)\, E^{-1} - \gamma E^{-2}]\, C = [1 - (\beta E^{-1} + \gamma E^{-2})]\, \varepsilon_1 + \alpha E^{-1}\, \varepsilon_2 \,. \qquad (8)$$

The variance of the expression on the right-hand side of (8) on the assumptions that $\mathscr{E}(\varepsilon_1^{\,2}) = \mathscr{E}(\varepsilon_2^{\,2}) = \sigma^2$ and $\mathscr{E}(\varepsilon_1 \varepsilon_2) = 0$ is

$$V(\eta) = (1 + \alpha^2 + \beta^2 + \gamma^2)\, \sigma^2, \qquad (9)$$

which is always greater, and may be much greater, than σ^2. It may be noted that the values of η are not serially independent and not independent of lagged values of consumption.

It is, of course, true that the auto-correlations of an economic variable may be of negligible size if they are not generated by a dynamic system. But it follows from the above argument that, even if the dynamic behaviour relationships are well established, the auto-correlations of the variables will be weak on account of the increase in the residual variance of the transformed system. It may therefore be concluded that Professor Kendall's investigations of auto-correlations cannot in principle throw any light on the possibility of estimating the kind of dynamic economic relationships in which economists are usually interested.

The argument so far has been concerned with formal matters. My third and final point is concerned with the material nature of the markets investigated. These are share and commodity markets which are the best examples of markets that are dynamically perfect. That is, any expected future changes in the demand or supply conditions are already taken into account by the price ruling in the market as a result of the activities of hedgers and speculators. There is, therefore, no reason to expect changes in prices this week to be correlated with changes next week; the only reason why prices ever change is in response to unexpected changes in the rest of the economy; but when prices do change, as shown by Professor Kendall's calculations in Table 5, they all tend to change together.

In formal terms this could be expressed by saying that the system of behaviour relationships linking the stock market to the rest of the economy was ill-conditioned, and from the point of view of investigating the dynamic properties of the system it is therefore particularly unfortunate that Professor Kendall found it necessary to choose these markets for his investigations.

Professor F. Paish: I speak with some trepidation in such an assembly, but there are two things I should like to say. I am extremely grateful to Professor Kendall for his conclusion.

As to his arguments I am not competent to judge. Many people in this world have spent many years trying to forecast what the stock markets are going to do on the basis of what they have already done. I always thought it futile, and am glad Professor Kendall has confirmed my opinion. It seems inevitable that where prices are based on expectation markets are as likely to go down as up. If the markets thought they were more likely to go up they would have gone up already. For this reason when there is option dealing, one always finds that the price of a double option is twice that of a single. Nevertheless, I should have thought that there was a possibility that some rather more complicated patterns might be discernible. The market depends on two different sorts of people—speculators and investors—or, in produce markets, hedgers. The former come in quickly but cannot stay in for long. The latter come in more slowly but stay. If the investors do not come in or do not come in as much as expected, the speculators get tired and have to sell. I should have expected some slight negative correlation after a lapse of some weeks. When the market has gone up sharply the initial rise is often followed by a fall as the speculators get tired of waiting for the investors to come in and take over. There are possibly patterns of that sort which could be distinguished in the data by somebody competent to do so.

Professor BARTLETT: I think that all I wanted to say has effectively already been said, but I should like to reiterate what seems to me to be the fundamental point. Professor Kendall in stressing the apparent randomness of his series has raised a paradox, but perhaps it may largely be resolved as a matter of interpretation. Incidentally, even if something is *apparently* random it does not necessarily mean that it *is* so; it may depend on something of which one is not aware. I have heard, for example, of a cotton firm in Manchester which in the good old days used to base its prices of cotton goods largely on the previous year's rainfall in Morocco, because they had a big export to Morocco, and if the rainfall there had been good the people would have a good harvest and could afford to pay high prices. If the annual rainfall were random, so would the annual prices appear so.

But the main point that one asks is: what are the chief characteristics of the data which Professor Kendall has been discussing? One is really not in a position to say from his paper because, as he has said, he is not able to quote the whole of the data, and if one works merely with differences one may be missing much of importance. It is true that the particular problem of predicting prices on the Stock Exchange can be discussed on the basis of the correlations of weekly differences, but these do not complete the summary of the variability and characteristics of the data, as previous speakers, and Professor Kendall himself, are aware. Mr. Quenouille has referred to second-order autoregressive series, so I will content myself with even simpler examples. Thus the price might be rising steadily, and in one sense be perfectly predictable; or we might have a stationary Markoff series in which the random residual became a negligible fraction of the total variance. (I do not think the total variance of Professor Kendall's series is given.)

I had the impression from the size of the cross-correlations in the paper that there may be more of a pattern than appears from the autocorrelations. It may be worth noting that if two autoregressive series are cross-correlated through a cross-correlation between the contemporaneous residuals, this cross-correlation will be preserved between the differences, even if the auto-correlations have been drastically reduced by differencing.

On a minor point, I was not altogether clear whether the trend-fitting at the end was on first differences, or on the original series.

Professor KENDALL: They were on the original series.

Professor BARTLETT: Yes, but the same notation was, I think, earlier used for first differences. If so, this might be made clearer.

Dr. K. S. RAO: Professor Kendall's criticism of the economist and especially the trade-cyclist is perhaps not of very wide generality. Large Sample methods of treatment of economic time-series, however attractive from the point of view of statistical theory, are difficult to justify from the point of view of economic theory, which such methods are intended to help.

A frequency distribution of a time-series of large length and the smallness of the first few serial correlation coefficients may not necessarily rule out systematic models, as can be seen by the following example, which is possibly one of a set that can be constructed.

Let m be an odd prime number and l an integer not divisible by m. Let n be a "large" integer. Consider a "time-series" of $N = n(m - 1)$ terms defined as follows:
Let

$$\theta = 2\pi\, l/m \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and

$$u_t = \frac{a}{\sqrt{m}} \sin \theta \left[\frac{n+t}{n}\right] + a \sin \theta \left[\frac{n+t}{n}\right] t, \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

for $t = 1, 2, \ldots N$, where $[x]$ denotes the greatest integer in x and a any given constant.

The Series in (2) is formed by using only the sine function. It can be noticed that, when we form a relative frequency table, we get a symmetric distribution with zero mean. Purely for purposes of simpler algebra, if we take n to be divisible by m, then

$$\bar{u} = \frac{1}{N} \sum_{t=1}^{N} u_t = 0. \qquad (3)$$

$$\text{Var } u_t = \frac{1}{N} \sum_{t=1}^{N} u_t^2 = \frac{a^2}{2} \frac{m}{m-1}. \qquad (4)$$

If P_k, for $1 \leqslant k < m$, denotes the k^{th} serial product of the series defined by the expression

$$P_k = \sum_{t=1}^{N-k} u_t u_{t+k}, \qquad (5)$$

after some simplification we get

$$P_k = na^2 \left\{ \sum_{s=1}^{m-1} \left(\frac{1}{m} \sin^2 \theta s + \frac{1}{2} \cos \theta s k \right) \right\} + O(1), \qquad (6)$$

where the constant implied in O depends only on m, a^2 and k. It can be seen that the first term on the right-hand side of (6) vanishes.

Hence, if r_k is the k^{th} serial correlation coefficient of the series in (2), then, for finite m, from (3), (4) and (6), it follows that

$$r_k = O\left(\frac{1}{N}\right) \text{ for } k = 1, 2, \ldots (m-1) \qquad (7)$$

Professor Jevons: I am not competent to deal with the mathematics of Professor Kendall's paper, but I feel that the economist is entitled to say something about the assumptions which are subjected to mathematical processes. At one time I paid some attention to time-series, and what became perfectly plain to me was that variations of the price of an individual commodity were perfectly random, and this was more particularly the case with highly organized markets such as the Chicago wheat market or the London Stock Exchange. If the day-to-day fluctuations of these markets are examined they will be found to be due to all sorts of causes—rumours of wars, changes of weather, forecasts from agricultural departments, or again the operations of bulls and bears; so I am not in the least surprised that Professor Kendall did not claim any very significant results from his analysis. I would suggest that if he had taken the time series of, say, agricultural commodities in some backward country, where the market is not properly organized, as, for instance, India or Burma, he would have found many more interesting things in the way of correlation.

Professor Kendall mentioned trade cycles in his paper. I did a little work myself on that subject from 1907 on. I was anxious to see whether statistics would support my father's theory that commercial crises depended on causes coinciding with the sunspot cycle. I found that there are two trade cycles—one a short period of about 3½ years, and the other a longer period which is usually ten years, or may be seven or eleven years. In order to put the idea to proof it seemed to me that we must take, not individual series of prices, but aggregate production—that is to say, if fluctuations in harvests produce any effect upon the activities of trade in general, it is not the result of fluctuations in one or more commodities, but of all the important commercial products. Therefore I took the production figures of the United States as far as available, and took weighted averages of these from year to year—weighted according to their average prices. I found quite remarkable evidence with regard to the shorter trade cycle. I later found, working on long series of statistics, both of prices and trade production, that the true period for the short cycle is 3·4 or 3·3 years. This period, practically one-third of ten years, has since been definitely established and accepted in the United States—the 40-month period, as they call it; and it is pretty obvious that the trade cycle is composed of two or three, or sometimes four, of these short periods, which are themselves fluctuating, probably, on the whole, with greater intensity than the longer periods.

This is a big subject. I should like to suggest that if you aggregate your figures (or sometimes you may take them individually, it depends on your purpose) you will find some interesting results. I only hope that Professor Kendall will continue his investigations along these lines.

The following contributions were received in writing after the meeting:

Mr. H. S. HOUTHAKKER: We are all indebted to the author for his work in the development and exposition of statistical methods, and his desire to apply some of these methods himself has much to commend it. Such applications, if based on a sound understanding of the material considered, can be of great benefit both to the subject under investigation and to the mathematical statistics. The qualification is important, for even the most refined methods, in fact especially those, will not lead to valid and interesting results unless they are applied to suitable data within an acceptable theoretical framework.

The evaluation of Professor Kendall's paper by the latter criterion is made difficult by the fact that there is no reference to a theoretical framework anywhere, nor indeed to work of others which the author may have had in mind. In his section 1, which is the only indication of the author's motives in undertaking this study, it is claimed that "it has been customary to analyse an economic time series by extracting from it a . . . trend . . . and then scrutinizing the residual portion for short term oscillatory movements and random fluctuations." This sentence would be correct if it began "it was customary twenty or thirty years ago." From about 1930 the development of econometrics has led to a general conviction among economists that little purpose is served by studying individual time series separately, and that they become interesting only when linked to each other in economically meaningful relations. The models thus constructed may well be of an autoregressive type, but trends or short-term oscillatory movements occupy only a minor place in them. A time series like the one for first differences of wheat prices may appear purely random if considered in isolation, yet regression analysis will soon reveal that the means of these apparently random differences are functions of supply and demand factors. And can there be any doubt that the movements of share prices are connected with changes in dividends and the rate of interest? To ignore these determinants is no more sensible or rewarding than to investigate the statistical properties of the entries in a railway time-table without recognizing that they refer to the arrivals and departures of trains.

It would nevertheless be wrong to dismiss this inquiry as merely the backwash of the Yule tide. Professor Kendall's iconoclastic efforts, even though directed against straw men, point to an interesting conclusion which is almost exactly the opposite of his own conclusion (*c*) in section 3. The series considered all refer to variables which in most econometric models would be regarded as endogenous. In recent statistical work much attention has been given to the complicated nature of the errors in these endogenous variables; it would make our work much easier if they were in fact random, for they could then be treated as if they were predetermined and the model accordingly simplified. However, the incompleteness of the author's analysis makes this conclusion so tentative that it would not yet seem safe to act on it.

Dr. D. R. COX: The serial correlations in Professor Kendall's Table 3 are all small, but nevertheless the correlograms for the different series behave in a remarkably similar fashion. This can be seen in an elementary way by considering the number of negative coefficients:

Lag	1	2	3	4	5	6	7	8
Number of negative coefficients out of 19 .	1	4	3	13	16	18	16	14

Lag	9	10	11	12	13	14	15	16	17	18 . . .
Number of negative coefficients out of 19	13	10	7	1	6	13	6	0	0	1 . . .

Similarly for the cross-correlations in Table 5 we find

Lag	1	2	3	4	5	6
Number of negative coefficients out of 56	5	10	13	31	45	47

In both cases there is very strong evidence for the existence of a non-random component, presumably of approximately the same structure for all or most of the 19 series. This, of course, does not affect Professor Kendall's point that the correlations are small, and hence useless for short-term prediction.

It is worth pointing out explicitly that the presence of certain types of non-randomness is shown, not by the occurrence of high serial correlations, but by other features of the correlogram, such as the existence of a run of coefficients all with the same sign. For example if the successive differences of price were completely random, price itself would undergo a simple random walk and would not be stationary. If there were in addition some tendency to return to a stable level of price, the correlations of differences would be small and negative (see, for example, Table 2, last column.)

As another example, Mr. W. L. Smith and I, in some recent work to appear in *Biometrika*, have discussed a special time-series arising in neuro-physiology, and have shown that a certain simple completely deterministic time-series, relevant to the neurological problem, is indistinguishable from a random series by the conventional correlogram method.

The problem, raised by Professor Kendall in Section 1, of dealing with long-term and short-term movements in one model arises in an acute form in investigations of irregularity in textile processing, notably in drawing and spinning. Drawing is a process of successive attenuation, each stage introducing an irregularity. In the final time-series the irregularities from the last stage have a characteristic length of a few cm., while the irregularities from the initial stages spread over several hundreds of yards. A method that is useful here is analysis of variance, in which the total variation is divided into variation between and within lengths l, and the coefficient of variation between the means of different sections of length l is plotted against l. This method was suggested by Tippett, and later by Yule under the name of the lambdagram. It is roughly equivalent to integrating the correlogram twice, and so is sensitive to runs of correlations all of the same sign. A related method has been useful in the neurological problem mentioned above.

Mr. A. S. C. Ehrenberg: Whilst it is difficult to describe time-series fully, they seem to possess features of some relevance apart from the first differences discussed by Professor Kendall. For example, the two wheat price series (§§6–19) are measured in cents on a scale with a more or less absolute minimum, and the mean, the scatter and any general trends of the primary data might be of interest. "End-effects" would be expected if the mean is low, and "scaling effects" generally. Thus the threefold increase in variance of the first differences of the weekly wheat series (§§9–11) might be connected with an upward trend of the *primary* data. (The constant mean referred to in §11 is presumably that of the first differences.) Since heteroscedasticity greatly reduces the utility of correlation coefficients, any mathematical transformation to counteract it might simplify the description of the data. Supposing that a transformation suggested by the trend (if any) could be found which gives homoscedasticity, it would of course not necessarily be useful to think of the original change in variance as just a scaling effect. But this might be so if the same transformation were to simplify many different price series, or perhaps if it were closely related to the economists' ideas on "real" prices. Information on such points could, one feels, be given simply enough. The scaling argument, for example, could not be countenanced at all in the *absence* of any upward trend in the primary series.

Some of the relevant information is admittedly contained in the description of the first differences given in the paper, though extracting it may raise more problems than it settles. Thus, there does appear to be a trend, of 69 cents, in the primary weekly wheat prices—entirely after the Great War—whilst in the same data treated as *monthly* instead of *weekly* figures (series 21, §16) the trend is only 28 cents, omitting one value at each extreme (or 43 cents for the whole series). No other deductions can be made about the variability of the primary series, at least not without some wild assumptions about their serial correlations. For example, if one supposed that in fact only the linear trend was striking, and that the second order serial correlation was no more than two-thirds of the first order correlation, the standard deviation would seem to be about five cents or less (and the first order correlation already as high as about ·75). In as far as such a standard deviation would then be anywhere near the truth—this would be likely in view of the rejection of widely outlying values (footnote, §9) of the order of 40 cents—these trends and discrepancies seem striking enough at least to deserve being explained away.

Turning to Professor Kendall's lack of sympathy with aggregative series and their "spurious" serial correlations (e.g., §§3, 23, 28), it may avoid confusion to mention that the correlations are *not* spurious in the sense that they *do* describe such series. Difficulties arise when one wishes to compare different aggregative series, let alone propound any so-called "causal" arguments. In Professor Kendall's illustrative example (§28), the correlation between the aggregative series E and F, each made up of n primaries, depends not only on the intercorrelations, but also on the *number*, of primaries. (Equation 12 may be a little misleading here since the n cor $(\varepsilon_i\varepsilon_i)$ in the denominator are unity.) Thus the correlation between E and F varies from ·1 ($n = 1$) through ·14 ($n = 2$), ·17 ($n = 3$), ·23 ($n = 15$—Kendall's example) to a limit of ·25. The "spurious" element arises because by averaging similar readings the strength of the relationship is affected, and not only the accuracy of its determination. This phenomenon is of course one of the reasons why correlation coefficients are used less and less frequently elsewhere—nothing like it seems to occur with most models of the analysis of variance kind, for example.

It should perhaps be remembered that in aggregating time-series like this, we are not necessarily averaging similar readings in the usual sense (e.g., sampling from an infinite population), since the supply of relevant time-series satisfying the original requirements (§28), that cor $(\varepsilon_i\varepsilon_j) = \cdot 4\ (i \neq j)$

and cor $(\varepsilon_i \eta_j) = \cdot 1$, may be strictly limited. There can, however, be little doubt that when the primary series are available, their serial correlations will be more useful than those of some aggregative series. For example, suppose that one wishes to compare the behaviour of two sets of data and that one of the primary series is missing, or different, in one of the sets of data. Then if the two sets of serial correlations for the aggregative series are identical, the two sets of data cannot be the same, as judged by the serial correlations of the primaries, whilst if these latter are identical, the correlations of the aggregative series would not show it.

Professor KENDALL, in reply: I should like to thank the proposer and seconder for their remarks and all the contributors to a helpful discussion. I made this paper as little mathematical as possible in the expectation that the mathematicians would all agree with me and the economists perhaps disagree. But as it has turned out, the mathematicians are doubtful, and the economists in their conclusions are inclined to agree.

Professor KENDALL subsequently wrote as follows:

With much of the discussion I agree and will confine myself to points of difficulty or disagreement.

(*a*) As Professor Champernowne points out, I am not attacking the use of index numbers for customary purposes, and I agree with points made by him, Professor Bartlett and Mr. Ehrenberg that "spurious" is not a good word to describe the serial correlations found in series of index-numbers. At the same time, it seems to me necessary to proceed very cautiously in the interpretation of serials in aggregative series, especially if they are used to estimate autoregressive periods. The problem is similar to that of measuring correlation in divisible units, and one has to be sure that one is measuring a property of the system itself, not merely a property of aggregation.

(*b*) In his observations on rough and ready methods of trend-determination Mr. Quenouille has, I think, failed to see my point of view. I do not deny that there are methods. What troubles me is whether they have any purpose except a purely descriptive one.

(*c*) Dr. Rao is quite correct in pointing out that one can construct non-random series for which a set of serial correlations vanish. But if we find in practice a set of small serial correlations we are, I think, on safe ground in supposing that successive values are nearly independent, as against the alternative that the generating system consists of a set of harmonic oscillations with very special periods.

(*d*) I look forward to seeing Dr. Cox's work and agree with him except on one point. He calls attention to patterns of signs in the serial correlations which indicate that, although small, the correlations are not haphazard. I think he is probably right, but I do not regard the point as settled beyond all doubt. Certain runs of signs might occur from bias in the estimations or from the fact that they may be serially correlated (by chance) even in random series. I propose to settle this question by constructing correlograms of a series of normal independent variates.

(*e*) Professor Allen, Professor Champernowne, Mr. Prais and particularly Mr. Houthakker all raise the question how far can one profitably analyse time-series without a theoretical model in the background? The general question of model-building in statistics and economics is one which I propose to discuss at length on another occasion. All I need say at this stage is that I think it is futile to lay down any rules as to whether fact should precede theory or theory fact. Scientific discovery proceeds by an alternation of the two. I feel entitled, however, to protest at some of Mr. Houthakker's comments. I could perhaps overlook his description of my work as to some extent Yulean backwash because of the mordancy of the phrase. But his claim that "even the most refined methods . . . will not lead to valid and interesting results unless they are applied to suitable data *within an acceptable theoretical framework*" seems to me the negation of the scientific spirit. I have tried to elicit certain facts about economic series. They may be wrong, but if they are correct they are facts, irrespective of any theoretical framework.

As a result of the ballot taken during the meeting, the candidates named below were elected Fellows of the Society:

Harry Brierley.	Denys William Humphries.	Thomas Alma Perkins.
Ronald Edgar Brook.	Michael Harold Kreps.	Robert Brian Sanderson.
Thomas Peter Browell.	John Kenneth Lambert.	Om Chandra Sharma.
Peter Richard Browning.	George Gordon Lilley.	Robert John Taylor.
Charles Frederick Evinton.	Denis Walter McAllister.	Alan Arthur Walters.
Derek John Finch.	Ronald Macer.	John Wrigley.

Corporate Representative

Peter Kenneth Wiggs, *representing* The Morgan Crucible Co. Ltd.