Frisch

THE ANALYSIS OF STATISTICAL TIME SERIES

		Page
1.	Introduction	1
2.	The Nature of the Trends of Successive Orders	17
3.	Propositions Concerning the Zeros of a Sum of Two Continuous Functions	33
4.	Upper and Lower Limits for Real Solution of the Linear Differential Equation of the Second Order	36
5.	The Method of Normal Points	47
6.	The Method of Moving Differences	74
7.	Summary	106

THE ANALYSIS OF STATISTICAL TIME SERIES

1. INTRODUCTION

All the known methods of determining secular trend in a statistical time series are total methods in the sense that the secular trend is determined by totalling or averaging in some way or another the data covering a shorter or longer period of time. Therefore the secular trend determined by one of these methods gives an adequate expression for the "normal" movement only so far as the conception of "average" may be taken as synonomous with, or at any rate as a first approximation to the conception of "normal".

Prima facie this substitution of the average point of view for the normal point of view seems inevitable when the datum to be analyzed consists of a given statistical time series and nothing else. In this case a total method seems to be the only way of attack. This accounts, I think, for the development of the various total trend methods.

In my estimation this prima facie conclusion does not hold good. The scope of the present essay is to point out how the conception of normal may be defined in another and more rational way, and to show how a given statistical time series may be decomposed not only in a seasonal fluctuation, a cyclical fluctuation and a long time trend, but in any number of components of successive orders, by studying the differential (as distinguished from the total) properties of the composite curve.

This amounts to saying that the ordinates of the components of successive orders in any point (i.e., at any given moment of time) may be determined solely by the course and shape of the given composite curve in the vicinity of the point considered, without any reference to the course of the curve in the past years.

*

May I be permitted to summarize what in my estimation are the main features of the present stage of time series analysis. We find a variety of methods. Some of them are concerned with seasonal fluctuations, some with long time trend and some with the business cycle itself. The nature of the difficulties encountered in the application of these methods is very different. Let us take a short survey of the difficulties connected particularly with the various trend methods.

The Fourier analysis and the periodogram method is only applicable to the analysis of a closed set of data (no account being taken of new data which successively become available), and even for this purpose it is subject to an inconvenience which in my opinion is fundamental. It does not show the evolution of the length of the period or periods present. Take for instance a sine function whose period is constantly growing, say growing in linear progression. If the interval considered is sufficiently large, the Fourier analysis of this function would show no predominant period. And the co-variation of two phenomena of this sort could not be revealed by Fourier analysis. But co-variation with respect to simultaneous lengthening or shortening of the period of oscillation of two phenomena is just one of the most important objects of time series analysis. If a simultaneous variation in the length of the periods of two phenomena (economic or physical) can be traced, this tells much more about the interrelation of the two phenomena than would do the existence of predominant Fourier coefficients for periods of approximately the same length in the two phenomena.

In determining secular trend by some sort of curve fitting, either the number of constants to be determined is small, as for instance in the case of the straight line or a simple curve of Gomperz or logistic type, but then the fit will generally be too poor if the interval considered is of any length, the secular trend being itself a line that changes curvature from time to time. Or the number of constants is great, as for instance in the case of a high order parabola, but then the curve which should be an expression for the secular trend only, will be influenced by the course of the cyclical fluctuations themselves. With regard to the extension of a curve fitted trend (with any number of constants) to new data which become available, experience has shown that even for the relatively short post-war period it has proved necessary in many cases to adjust calculations by changing the trend previously adopted. From a theoretical point of view it is evidently a very unsatisfactory method that from time to time necessitates such changes based on a subjective judgment.

The moving average method eliminates the difficulty of changing trend, but the condition for the applicability of this method is evidently in the first place that the period of the cycle is relatively constant (the number of items in the average then being chosen so as to cover one period), and in the second place that the oscillations above and below the normal are approximately of the same magnitude. As is well known neither of these conditions ~~are~~ is fulfilled in the case of the actual time series which are usually submitted to statistical analysis.

Analogous remarks may be offered relative to the area method.and its variants, e.g., the method of moving integration. Furthermore, in my opinion the very idea of defining the secular trend so as to have the average oscillation (or the area) above and below the normal equal, is mistaken, because one of the problems we have to face is just to make a comparison between the magnitude of the positive and the negative oscillations.

The reason for this great variety in the nature of the difficulties is evidently that there is no logical relation between the various methods. There is no logical relation between the trend methods inter se, or between the trend methods and the seasonal index methods etc. There is no general principle from which these various methods may be derived.

This leads up to a more general consideration regarding the present stage of time series analysis.

The time series analysis has now reached a point of development where the problem can no longer be formulated in this simple way: How to determine the long time movement around which the business cycle is fluctuating. This long time movement has proved to be itself a major cycle(1) (with a duration of about 30 to 50 years) oscillating around a still longer "long time movement". And I think that if our series were long enough, we should certainly find that this "still longer move-

(1) The major cycle has recently been discussed by N.D. Kondratieff, Archiv f. Socialwissenschaft u. Socialpolitik, 1926, pp. 573-609. The problem will also be considered in a forthcoming book by Simon Kuznets. Mr. Kuznets has been kind enough to show me some of his charts.

ment" contains in it a super-cycle with a duration of say between 200 and 500 years. One thing suggesting the existence of a super-cycle is the fact that the normal of the major cycle in some cases is found to ascend and in other cases to descend.

On the other hand, the ordinary business cycle is itself a long time movement with respect to the seasonal fluctuations and these are in turn long time movements with respect to the weekly or daily oscillations observed in some series. It is perhaps a daring generalization but is it not plausible to think that many of those fluctuations which we now consider as accidental, are really due to small cycles, and only appear as accidental because our data are not available at intervals short enough to investigate the manner in which these "accidental" fluctuations are generated as an interference phenomenon composed of small and still smaller cycles?

Our time series then should have to be interpreted as a phenomenon composed of a great number of waves, the waves forming a sort of suite or scale which extends practically indefinitely downwards and upwards. Actually we are capable of tracing only the waves in the middle part of this scale much in the same way as our eye is capable of catching only the waves in the visible part of the spectrum.

Anyhow, whatever opinion one might have regarding such a generalization, I think that the only logical consequence of the facts revealed by the modern intensive study of cycles and trends, is that all the hitherto recognized components of a time series, should be considered from the same point of view: as trends of successive orders. They should therefore be treated by methods which in principle are of the same nature, the difference between the methods being rather a difference in application necessitated by the fact that both the period covered by the available data and the length of the interval between consecutive data have a different significance with respect to trends of low and high order.

Consequently the time series problem must be restated and given this more general form: What is the order of complexity of this particular series? That means, what is the number of trends which can actually be traced in it? What are the periods of these trends? How can the trends be eliminated one after another, and how can the ordinates of each isolated trend be determined?

This is the general problem we have to face in this article and whose solution shall be attempted by the differential method.

If a solution is possible, and I believe such is the case, the solution will be much more flexible than any method of curve fitting or seasonal index computation, etc. It will be essentially a "moving" method. The very fact that the method is differential makes it possible to change the basis of computation continuously from one moment of time to the next. If, for instance, the seasonal fluctuations are considered as the trend of lowest order, it will be possible, at any rate theoretically, to trace how the course of these fluctuations changes from one year to another. The practical carrying out of this procedure has proved possible even with monthly data if the seasonal fluctuations are not too complicated, But of course the method will work better with weekly data.

It is quite clear that unless some sort of hypothesis is adopted as to the nature of the trends of successive orders present in the statistical time series to be analyzed, a differential method cannot be devised. In this case the problem is quite indeterminate. Therefore while the total methods are purely empirical, the differential method must be to a certain extent a priori.

This fact does not, I think, indicate a difference between the differential and the total methods in the sense that the former is more problematical because more dependent on restrictive assumptions. The total methods, too, are certainly conditioned by restrictive qualifications. The difference is rather that the restrictive assumptions underlying the differential method are stated more explicitly than those involved in the total methods. This should not be a weak point in the differential method.

Furthermore, I believe that the assumptions underlying the differential method may be found on closer examination to be not less plausible than the implicit or explicit assumptions involved in the usual methods of time series analysis. Therefore, in my estimation, no serious objection can be made regarding the theoretical aspect of the differential method. If any objection is to be made, it must be a practical one, namely that the differential method involves the consideration of successive differential coefficients, approximated by successive differences, and the higher the order of a difference, the more it is influenced by accidental errors.

There is only one way of testing the weight of an eventual objection like this: To apply the differential method to actual statistical data. During the past years I have applied the method of normal points (Section 5.) and parts of the method of moving differences (Section 6.) to various kinds of statistical series, and the irregularity of the material used has not proved too great for a successful determination of the various components. But working in the University of Oslo without organized assistance for tabulating and computing work, I have not been able to carry the numerical tests as far as I should have wished. It is, therefore, my sincere hope that some well equipped American research bureau will consider it worth while to undertake a thorough numerical test of the methods here presented.

*

The first and very simple idea which has served as the starting point of the differential method is this:

Take a curve which is a sum of a sine function and a straight line.

Let

$$(1) \qquad w = y + z = [C \sin\sqrt{c}(t-t_0)] + (at + b)$$

$$y = C \sin\sqrt{c}(t-t_0)$$

$$z = at + b$$

be the equation to the curve.

This curve evidently represents in its most simple form the composition of a "cyclical fluctuation" y with a "secular trend" z.

Now differentiate the given function w twice and change the sign. This gives

$$y = -w''/c$$

In this case therefore:

(a) The cyclical fluctuation is simply proportional to the second differential coefficient of the composite function w, which is supposed to be known.

(b) The cyclical fluctuation passes its normal in the very same points where the second differential coefficient of of the known composite curve vanishes.

When we have to consider actual statistical data it would certainly not be a working hypothesis to assume that the given composite curve is rigorously of the form (1). But it might perhaps be plausible to make a more general assumption from which it might be deduced that one or both of the propositions (a) and (b) hold good approximately. To search for such an assumption, let us first consider separately a single component: The ordinary cycle itself.

It is rather popular by way of analogy to speak of the cycle as a pendulum oscillating back and forth. Such an analogy might be good or bad according to the use made of it. Its value can hardly be proved or disproved by any a priori discussion. The ultimate test must be if it works or not when it is applied to actual data. It should be emphasized that the pendulum analogy is here used merely as an illustration for the sake of suggesting some plausible working hypothesis regarding the differential properties of the curves representing the various components in a time series. It is not considered as an adequate representation of the complexity of economic life, from which all kinds of theoretical explanations might be derived.

If a pendulum is oscillating without friction in a gravitation field of constant intensity and the oscillations are small, then the force driving the pendulum back to its normal (i.e. vertical) position, is in every moment of time of opposite sign and proportional to the actual distance of the pendulum from the normal. Therefore if we plot the curve $y(t)$ which shows how this distance varies with time, the differential equation to the curve will be

$$y'' + cy = 0 \tag{2}$$

c being a positive constant, whose magnitude depends on the length of the pendulum and the intensity of the gravitation which strives to drive the pendulum back as soon as it is displaced from its normal position.

The solution of the equation (2) is just the sine function

$$y = C \sin\sqrt{c}(t-t_0)$$

where C designates the maximum deviation from the normal and t_0 one of the moments where the pendulum passes its normal. The ratio $\pi/\sqrt{c}$ is equal to the constant distance between two consecutive zeros of y. In this simple case the average (or the area) of the positive deviations is equal to that of the negative deviations.

Now suppose that the intensity of the field which strives to drive the pendulum back to its normal is no longer constant but a function of time $F(t)$. The constant length of the pendulum may be considered as included in the function $F(t)$ as a feature characterizing the intensity, so that the equation to the curve will be

(3) $$y'' + Fy = 0$$

This means that at a given moment of time (i.e. when the intensity of the gravitation has a given magnitude) the force acting on the pendulum will still be the stronger the more the pendulum is displaced from its normal, but if we compare the situations at two different moments, the force will no longer be proportional to the distance from the normal. The equation (3) is the canonical form of the differential equation of the second order, any equation of the second order being reducible to this form. Our only restrictive assumption is that F shall be regular and essentially positive over the whole t-range considered.

In this case the pendulum will still be oscillating around its normal, but otherwise the oscillations may take any form compatible with the condition that the force acting on the pendulum is always directed toward the normal position (because F is supposed positive).

For instance, the movement away from the normal may be sharp and quick, the return slow, eventually speeding up at moments more or less suddenly. The duration of an oscillation to one side, or the maximum deviation to one side, may be greater than to the other side, and so forth. If we add that the pendulum may be exposed to small accidental pushes during its movement, I think we shall have a fairly good picture of a kind of cyclical fluctuation where positive and negative deviations alternate much in the same manner as in the cycles revealed by the study of actual statistical data.

Furthermore not only the ordinary cycle, but also the shorter oscillations: the seasonal fluctuations etc., as well as the longer fluctuations: major cycle etc., may be illustrated by the pendulum movement. For the average length of the period of oscillation may be varied by varying the average magnitude of F. A great F means a strong force driving the pendulum towards the normal and at the same time a short period of oscillation (i.e. seasonal fluctuations), and inversely (i.e. major cycle). If $F = 0$ we have as a limiting case the straight line.

These considerations will justify the hypothesis on which the differential trend method is based. We assume that each of the trends which make up the composite time curve, is a solution of a differential equation of the type (3).

F may be called the <u>gravitation</u> of the trend considered. And the reciprocal of the average value of F or rather of $\sqrt{F}$ may be taken as a preliminary expression for the order of the trend. We shall later discuss the question of trend order in more detail.

As to the manner in which the successive trends make up the composite curve w, various assumptions are possible.

If $y_n (n = 0, 1 \ldots N)$ are the trends of successive orders, each y_n being a solution of an equation of the form:

(4a) $$y_n'' + F_n y_n = 0$$

then we may, for instance, consider additive trends:

(4b) $$w = y_0 + y_1 + \ldots + y_N = \sum y_n$$

or multiplicative trends $w = y_0 y_1 \ldots y_N$, (i.e. $\log w = \sum \log y_n$). Or we may consider any combination of additive trends and multiplicative trends. In this article I shall only consider additive trends, but it can be shown that under certain additional assumptions, some of the results hold good approximately also for multiplicative trends.

If we plot the solution y_n of (4a), we shall have a curve which may take nearly any shape provided it has a finite differential coefficient and always turns the concave side to the t-axis. This property of the curve y_n is only another expression for the fact that the force acting on the pendulum (and therefore also the acceleration) is always directed toward its normal position. This is an essential feature of our assumption. It entails an important consequence regarding the conception of the normal of a variable which oscillates as a function of time. The normal or equilibrium position is defined as the position toward which the force tends to push the variable. If we do not want to introduce the conception of force in the definition, but only want to take account of the shape of the curve which represents how the phenomenon varies with time, then the normal will be defined as some position lying to the side toward which the curvature is directed.

The maximum deviation of the curve y_n to one side of the normal will no longer as in the simple case of the sine function be equal to the maximum deviation to the other side. Nor will the average (or the area) of the positive deviations be equal to that of the negative deviations. This "total" property of the curve is only an incidental property which is present in the special case where the gravitation is constant. The essential property which characterize the points in which the normal is passed, is a differential property, namely that these points are the very points where the second differential coefficient vanishes. This property holds good exactly not only for the sine function but also for the general kind of functions here considered.

If y_0 is the trend of lowest order, the second differential coefficient y_0'' of y_0 will no longer as in the simple case (1) be rigorously equal to the second differential coefficient w'' of the composite curve w. But under certain conditions regarding the relative difference in order between successive trends, y_0'' will be the predominant term in w'' in points not lying in the vicinity of the zeros of y_0, so that the points where w'' vanishes will approximately indicate the location of the zeros of y_0'', and hence of y_0. This generalization of the above proposition (b)√ (p.9) is the basis of the "method of normal points" developed in Section 5.

The ordinate y_n of a trend will not in general be proportional to its second differential coefficient. But it might be so approximately in the vicinity of a point (not lying close to a zero of y_n), if the proportional variation of the corresponding gravitation F_n is small.

Furthermore if not only the first order proportional variation of the various gravitations are small, but also a certain number of the higher order proportional variations (which will be defined more precisely in the following Section), then the ordinates of the successive trends may be expressed in terms of the differential coefficients of even order of the composite curve. This generalization of the above proposition (a)√ (p.9) is the basis of the "method of moving differences" (method of instantaneous approximation) developed in Section 6.

Note on the relationship of successive trends.

The pendulum movement might further be applied as an illustration of the relationship between the successive components in a time series, i.e. the way in which a fluctuation of low order might be said to generate those of higher order or vice versa. As this problem does not fall within the scope of the present article, I shall only give a short indication of an illustration which might perhaps be found suggestive.

Suppose we have a pendulum of considerable length and mass. To this we attach a pendulum, much shorter and with much smaller mass; to this a third pendulum, still much shorter and with still much smaller mass, etc. Suppose the chain of pendula is at rest in a vertical position. Now set the upper pendulum in movement without touching the rest of the chain. The very fact that the upper pendulum gets into oscillation will entail the oscillation of the next one, etc. When the system is in movement and the mass of each pendulum is very small in proportion to the mass of the next higher, then the oscillation of each pendulum referred to its normal, i.e. the vertical through the next higher pendulum, will approximately be a sine function, and the duration of its oscillation will be approximately proportional to the square root of its length. The fluctuation of the lowest (smallest) pendulum referred to the vertical through the point of suspension of the whole system, will be a composite fluctuation, where the oscillations of the higher pendula (referred to their respective normals), will represent " trends of successive orders".

Inversely: if the lowest pendulum is set into oscillation, this will entail the oscillation of the next higher, much in the same way as a sudden draught will give impetus to the oscillation of some freely suspended body, say the cord of the window shade. The mean amplitude of the oscillation of the cord will depend on the intensity of the impulse, but the mean duration of an oscillation will not. In the same manner the movement of the second lowest pendulum of our chain will propagate itself to the higher pendula until the whole system is in oscillation. And again the oscillations of the various pendula will represent "trends of successive orders". The same result will follow if one of the intermediate pendula is set in motion.

If friction is present, a stream of new impulses will be necessary to keep the system going. It may be impulses of a more or less accidental character distributed more or less irregularly over the interval of time considered, and over the various pendula. In this case it would have no meaning to say that the movement of any particular pendulum generates the movements of the other. The movements of the various pendula might rather be said to be in part independent and in part generated by each other.

2. THE NATURE OF THE TRENDS OF SUCCESSIVE ORDERS

The theory to be set forth in the present article is essentially concerned with approximations. If the theory is to be something more than a pure statement of rules of thumb, it is of fundamental importance that the nature of the approximations involved should be thoroughly investigated. Especially I consider it important that it should be made quite clear on which properties of the components the closeness of the approximations really depends.

The object of the present Section is to give the precise definition of those properties of the trends, with which the argument of the following sectionsis concerned. In order not to break up the introductory matter of the first sections, some propositions are here assumed without proof, the proofs being given in Sections 3 and 4.

The properties of the trends present in a time series may be classified under two headings: properties which characterize the relation between successive trends, and properties which characterize the nature of a trend as such. Let us take these properties up in order.

If an illustration was to be given for the purpose of visualizing the manner in which a composite time series is made up of trends of successive orders, one would probably start by drawing the first component in the shape of some sort of cycle curve, where the zero distance, i.e. the distance between two consecutive zeros, did not vary widely. The relative difference between one zero distance and the next following would always be less than say 100%. The next component would be of the same kind, a sort of major cycle, where the average zero distance would be far greater than the average zero distance in the first cycle, say 7 or 10 times as great. The very conception of the second component as a trend of higher order than the first, implies that the second component does not change its curvatures as often as the first component does. If the average zero distance in the second component should be nearly as small as that of the first component, the curve representing the sum of the two components would certainly be a composite curve, but it would not be plausible to consider the two components as trends of different orders. Similarly the average zero distance in a third component would be much greater than that of the second component, and therefore still greater, say 50 or 100 times as great as that of the first component, etc. Let us state this fact in a precise form.

We introduce the following notations

t_{nk} $(n= 0,1,..N;\ k=1,2...)$, zeros of y_n

(5a) $D_{nk} = t_{n,k+1} - t_{nk}$, zero distances in y_n

(5b) i_{nk} = interval $(t_{nk}, t_{n,k+1})$

If t is any point, the corresponding intervals i_{oj} and i_{nk} may be defined without ambiguity by

(5c) $$t_{oj} \leq t < t_{o,j+1}$$

(5d) $$t_{nk} \leq t < t_{n,k+1}$$

Consider the function $D_n(t) = D_{nk}$, where k is defined by (5d).

If y_o, y_1... are trends of **different** orders, the value of the function $A_n(t) = D_o/D_n$ would always be a small fraction, and the smaller the greater n is, $A_n(t)$ would oscillate (discontinuously) between say $(1/7)^n$ and $(1/10)^n$. The upper limit A_n of $A_n(t)$ for the whole t-range considered would also be a small fraction, say $(1/7)^n$. Let A be the greatest of the numbers $A_1, A_2 \dots$. The magnitude of the positive numbers A_n $(n = 1,2 \dots N)$, and their upper limit A is one of the features which characterize the relative difference in order between the successive trends.

If t is any point and j and k are defined by (5cd), then

(6a) $$D_{oj}/D_{nk} \leqq A_n \leqq A$$

Now consider the ratio H_{nk}/D_{nk} , where H_{nk} designates the value (regardless of sign) of the extremum of y_n between t_{nk} and $t_{n,k+1}$ (one and only one such extremum exists because y_n always turns its concave side to the t-axis). The ratio H_{nk}/D_{nk} gives an idea of the intensity (the distinctness) of the wave in y_n between t_{nk} and $t_{n,k+1}$. The wave is sharp if H_{nk}/D_{nk} is great, otherwise it is flat.

The conception of trends of different orders does not imply that the average value of H_{nk}/D_{nk} shall decrease say in the same proportion as the numbers A_n when the order of the trend is rising. It would indeed be quite possible to imagine a time series where, for instance, the major cycle was more distinct than the ordinary cycle. But still I believe that in most practical cases the ratio H_{nk}/D_{nk} will be decreasing or anyhow not increasing as n increases.

Let

$$B_n(t) = \frac{H_{nk}/D_{nk}}{H_{oj}/D_{oj}}$$

where j and k are defined by (5cd).

And let the positive number B_n be the upper limit of the function $B_n(t)$ for the whole t-range. If $y_o y_1 \ldots$ are trends of different orders, then B_n would probably in most cases be less than unity. And in many cases it would probably be a small fraction say not greater than $(\frac{1}{2})^n$.

The magnitude of the positive numbers B_n ($n = 1,2 \ldots N$) is the second feature which characterize the relative difference in order between the successive trends.

If t is any point and j and k are defined by (5cd) then

(6b) $$H_{nk}/D_{nk} \leqq B_n \cdot H_{oj}/D_{oj}$$

It should be pointed out that the definition of the numbers A and B does not in itself involve any assumption as to the magnitude of these numbers. The mathematical deduction of the following sections is therefore general. It is only when the ultimate formulae are to be interpreted that we come back to the assumptions regarding A and B.

I now proceed to the properties which characterize a single trend. These properties will be concerned with the average value and the fluctuation of the gravitation F_n of the trend y_n.

Let i_{nk} designate the interval defined in (5b). If F_n is considered as an arbitrary function, any of the various average definitions may be applied to it. Consequently the average value of F_n in i_{nk} might turn out differently. However, if F_n is considered in its relation to y_n, only one average definition is possible for i_{nk}. The average of F_n in i_{nk} must be put equal to

(7) $$c = (\pi/D_{nk})^2$$

where D_{nk} is defined by (5a).

For in conformity with the conception of an average, the average value of F_n in i_{nk} must evidently be such a value that if the various actual values of F_n in the interval are put equal to the adopted average, the effect produced should be the same, namely, $y_n(t_{nk}) = y_n(t_{n,k+1}) = 0$ (with $|y_n| > 0$ between t_{nk} and $t_{n,k+1}$). And by (Ia) of Section 4 no other value than c can satisfy this condition.

This consideration is perfectly similar to the consideration which leads to the definition of the average (proportional) increase per unit of time in a population between two censuses or to the definition of the average rate of interest in the period between two moments of time at which the magnitude of capital is given.

The question might arise if F_n really assumes the value c at least once in the interval i_{nk}. That this is really so will be shown in (IIIb) of Section 4. From (IIId) of the same section will further follow that the average-definition for F_n satisfies the condition that if all the values of F_n are multiplied by some constant quantity, the average too will be multiplied by this same quantity.

Without further explanation the notion of F_n as the "gravitation" of the trend y_n might perhaps be somewhat vague. I believe that this notion is rendered more palpable by the fact that the average value of F_n in i_{nk} is equal to $(\pi/D_{nk})^2$. This fact also shows, I think, that the notion of gravitation deserves to be given some attention in time series analysis.

In connection with the gravitation we shall consider what might be called the fictive (as distinguished from the actual) zero distance of y_n for a given moment of time t. By this is meant the length of the half period (i.e., distance between two consecutive zeros of y_n) which would be realized if the gravitation F_n kept constant and equal to the actual value it has in the point t. The fictive zero distance of the trend y_n in the point t is therefore equal to .

(8a) $$d_n(t) = \pi/\sqrt{F_n}$$

If F_n is constant i.e. if y_n is rigorously a sine function, then the fictive and actual zero distances will be equal in any point. If F_n is not constant, there will still be a certain relation between the magnitudes of the fictive and the actual zero distances. And the relation will be all the closer the less the proportional variation in F_n.

In fact the average value of $1/\sqrt{F_n}$ for a certain interval may be defined as $1/\sqrt{A}$ where A is the average of F_n for the interval. If this definition is adopted, we see that the average value of the fictive zero distance, taken throughout the interval i_{nk}, is just equal to the length of this interval, i.e., equal to the actual zero distance. This is another way of rendering the notion of the fictive zero distance more intuitive.

Now considering the fictive zero distance instead of the actual, we may define two sets of numbers a_n and b_n analogous to the numbers A_n and B_n previously defined. In view of the relation existing between the actual and the fictive zero distance, the ratio of the fictive zero distance in the trend y_n to the fictive zero distance in the trend y_n

(8b) $$a_n(t) = d_o/d_n$$

will be a function fluctuating (continually) between limits not very different from the limits of the previously defined ratio $A_n(t)$. If a_n is the upper limit of $a_n(t)$ for the whole t-range, then the numbers $a_1 a_2 \ldots a_N$ will form a suite of approximately the same kind as the suite $A_1 A_2 \ldots A_N$.

In any point t we evidently have

(9a) $$\sqrt{F_n} \leqslant a_n \sqrt{F_o}$$

Similarly we might consider the ratio

$$b_n(t) = \frac{H_{nk}/d_{nk}(t)}{H_{oj}/d_{oj}(t)}$$

which is analagous to the previously defined ratio $B_n(t)$. If b_n is the upper limit of $b_n(t)$ for the whole t-range, then the numbers $b_1 b_2 \ldots b_N$ will form a suite of approximately the same kind as the suite $B_1 B_2 \ldots B_N$.

In any point t we evidently have

(9b) $$H_{nk}\sqrt{F_n} \lessapprox b_n \cdot H_{oj}\sqrt{F_o}$$

where j and k are defined by (5cd).

Suppose that the numeration of the trends has been chosen so as to make the suite of products $A_n B_n$ (or the previously considered suite $a_n b_n$) not increasing and such that $A_1 B_1$ (or $a_1 b_1$) has the least possible value. The product $A_n B_n$ (or $a_n b_n$) may then be taken as an expression for the order of the trend y_o measured in proportion to the trend y_n. The relative difference in order between y_o and y_n is great if $A_n B_n$ (or $a_n b_n$) is a small fraction. We shall especially have to consider the case in which the small value of $A_n B_n$ (or $a_n b_n$) is due to the fact that A_n (or a_n) is a small fraction and B_n (or b_n) is not great. In this case y_o and y_n may be said to have a great difference in order in the proper sense. On the contrary, y_0 and y_n must be considered as trends of the same trend order if A_nB_n (or a_nb_n) is close to unity or even greater than unity. In the special case where the gravitations are rigorously constant and the numeration of the trends is chosen in the way referred to, we evidently always have $A_nB_n (= a_nb_n) \geq 1$.

It was pointed out that the accordance between the actual and the fictive zero distance in y_n is all the better the smaller the proportional variation in F_n. Now to characterize the maximum proportional variation of F_n in i_{nk} we introduce the numbers p_{nk} and q_{nk} which measures respectively the ratio of the maximum value of F_n in i_{nk} to its average value $(\pi/D_{nk})^2$, and the ratio of the average to the minimum value of F_n in i_{nk}.

(10a) $F_{nk}^{(max)} = (\pi/D_{nk})^2 \cdot p_{nk}$ (10b) $F_{nk}^{(min)} = (\pi/D_{nk})^2 / q_{nk}$

Both p_{nk} and q_{nk} are obviously positive numbers not less than unity. For the sine function where F_n is constant, we have $p_{nk} = q_{nk} = 1$. For any other function we must have $p_{nk} > 1$ and $q_{nk} > 1$. This follows from (IIIb) of Section 4.

In i_{nk} we evidently have

(10c) $$(\pi/D_{nk})^2 / q_{nk} \leq F_n \leq (\pi/D_{nk})^2 \cdot p_{nk}$$

The product $P_{nk} = p_{nk} \cdot q_{nk}$ expresses the ratio of $F_{nk}^{(max)}$ to $F_{nk}^{(min)}$, i.e. $F_{nk}^{(max)} = F_{nk}^{(min)} \cdot P_{nk}$

Evidently $p_{nk} \leq P_{nk}$ and $q_{nk} \leq P_{nk}$. The greatest of the numbers p_{nk} and q_{nk} will be designated by p. The greatest of the numbers P_{nk} will be designated by P.

The significance of the numbers p and P might be rendered more intuitive by analyzing their relation to the ratio between two consecutive zero distances. We evidently have

(11a) $$D_{n,k+1}/D_{nk} = \sqrt{p_{n,k+1}q_{nk}} \cdot \sqrt{F_{nk}^{(min)}/F_{n,k+1}^{(max)}}$$

$F_{nk}^{(min)}$ cannot be greater than $F_{n,k+1}^{(max)}$ because F is continuous and hence $F_{nk}^{(min)} \leqq F_n(t_{n,k+1}) \leqq F_{n,k+1}^{(max)}$. Therefore the second factor on the right hand side of (11a) cannot be greater than unity, but it will be equal to unity if the value of F_n in any point in i_{nk} is not less than the value in any point in $i_{n,k+1}$, this is for instance the case if F_n is monotonically decreasing throughout both intervals.

Now if $F_{nk}^{(min)} = F_{n,k+1}^{(max)}$, we have

(11b) $$D_{n,k+1}/D_{nk} = \sqrt{p_{n,k+1}q_{nk}} \leqq p$$

This means that the ratio between two consecutive zero distances in y_n can never be greater than the upper limit p of the numbers p_{nk} and q_{nk}. But it might be equal to p.

On the other hand as

$$F_{nk}^{(min)} \leqq (\pi/D_{nk})^2 \leqq F_{nk}^{(max)}$$

(the equal signs holding good only in the case where F_n is constant), we have

$$1/P \leqq D_{n,k+1}/D_{nk} \leqq P$$

The difference between $D_{n,k+1}/D_{nk}$ and the limits can be rendered arbitrarily small, namely if F_n is nearly constant in both intervals i_{nk} and $i_{n,k+1}$, falling sharply or rising sharply from the last part of i_{nk} to the first part of $i_{n,k+1}$ (in which case p_{nk} and $q_{n,k+1}$ or $p_{n,k+1}$ and q_{nk} are close to unity). This means that the ratio between two consecutive zero distances in y_n can never be greater than P, but it might come near to P. The case in which $D_{n,k+1}/D_{nk}$ come near to P is of course a much more special case than that in which it is equal to p.

The numbers p_{nk}, q_{nk} and P_{nk} measure the total variation of F_n in i_{nk}, but gives no account of the differential variation of F_n (per unit of time). Even if an upper limit of p_{nk} and q_{nk} or of P_{nk} is fixed, the differential variation of F_n in any point in i_{nk} may take any value. A coefficient measuring the differential variation of F_n is now to be considered.

In economic theory as well as in statistical application a well known coefficient is the coefficient of elasticity or its reciprocal which might be termed the increase-proportion. The increase-proportion of a function f(t) is defined as the ratio between the relative variation in f(t) and the corresponding relative variation in t, i.e., as

$$\lim_{h \to \infty} (f(t+h)-f(t))/f(t) \div h/t = d \log f/d \log t = (f'(t)/f(t))t$$

This coefficient is independent of the unities with which f and t are measured but it is not independent of the origin of t. However for the purpose of measuring the differential variation in F_n what we need is a coefficient which is invariant for a linear transformation of t.

Such a coefficient would be the ratio between the relative variation in F_n and the corresponding relative variation in t, the variation in t being measured relative to some quantity which is independent of the origin of t and varies in the same proportion as t when the unit of t is changed. Such a quantity is the zero distance in y_n. The intuitive signification of an increase-proportion defined in this way would be that if r is the magnitude of the increase-proportion, and the increase in time amounts to say x% of the zero distance (x being a small quantity), then the corresponding proportional increase in F_n would be rx% (to a first approximation). I shall later offer another observation supporting the plausibility of measuring the variation in t proportionally to the zero distance.

Now the zero distance in proportion to which the variation in t is to be measured, might be the actual zero distance D_{nk} or the fictive zero distance $d_n(t) = \pi/\sqrt{F_n}$

Accordingly we shall have the two coefficients

$$R_n = (F_n'/F_n)D_{nk} \qquad r_n = \pi F_n'/F_n^{3/2}$$

The first is discontinuous in the zeros of y_n, but the definition is unambiguous as we have made the convention that k shall be determined by (5d). The second coefficient is continuous for all values of t. Both coefficients are invariant for a linear transformation of t.

Now let the positive numbers R and r be the upper limits of the functions R_n and r_n respectively (regardless of sign) for the whole t-range and for n = o,1...N. We consequently have in any point t

(12a) $|F_n'/F_n| \leq R/D_{nk}$ (12b) $|F_n'/F_n^{3/2}| \leq r/\pi$

The rough intuitive signification of the two coefficients R_n

and r_n are much the same. Multiplied by the relative increase in t (measured in proportion to the actual or fictive zero distance respectively) they give the corresponding proportional increase in F_n, to a first approximation. The significance of the upper limit R may further be illustrated in a very simple manner by its relation to p_{nk}, q_{nk}, P_{nk} and their upper limits p and P.

Let t' and t" be the points in i_{nk} where F_n reaches its maximum and minimum respectively, then

$$\log\left(F_{nk}^{(max)}/F_{nk}^{(min)}\right) = \left|\int_{t'}^{t''}(F'_n/F_n)dt\right| \leqq \int_{t_{nk}}^{t_{n,k+1}} |F'_n/F_n| dt \leqq R$$

so that

(13) $$P_{nk} = p_{nk} \cdot q_{nk} \leqq e^R$$

This holds good for all n and k, even if P_{nk} or p_{nk} or q_{nk} should be equal to their upper limits P and p respectively.

Hence

(14a) $P \leqq e^R$ and a fortiori (14b) $p \leqq e^R$

This shows the relation between the maximum total variation of the various gravitations and their maximum differential variation. If R is given, this entails an upper (but not a lower) limit for P and p. If one of the numbers P or p are given, this entails a lower (but not an upper) limit of R.

The limit (14a) is a precise limit in the sense that if P is given, the course of F_n may be chosen so as to have e^R not exceeding P and still make one of the P_{nk} equal to P. And if R is given the course of F_n may be chosen so as to have one of the P_{nk} equal to e^R. In fact, if F_n is monotonically increasing or decreasing throughout the whole interval i_{nk} at its maximum rate, then we have $P_{nk} = F_{nk}^{(max)} / F_{nk}^{(min)} = e^R$, and no other P_{nk} can exceed this value. The limit (14b) is not a precise limit (except in the trivial case where all the gravitations are constant).

We get perhaps a still better idea of the significance of R by analyzing its relation to the ratio between two consecutive zero distances in y_n.

From (11a) and (14b) we deduce

(14c) $D_{n,k+1}/D_{nk} \leqq e^R$ and $D_{nk}/D_{n,k+1} \leqq e^R$

This is not a precise limit. It might not be possible to choose F_n so that $D_{n,k+1}/D_{nk} = e^R$. But it is certainly possible to have

(14d) $D_{n,k+1}/D_{nk} = e^{R/2}$ or $D_{nk}/D_{n,k+1} = e^{R/2}$

This means that if the upper limit R of the differential variation of the various gravitations is fixed, this does not prevent the ratio between two consecutive zero distances in any trend y_n from becoming equal to $e^{R/2}$.

I shall show that the first equation of (14d) holds good if F_n is monotonically decreasing throughout both intervals i_{nk} and $i_{n,k+1}$ at its maximum rate R.

In i_{nk} and $i_{n,k+1}$ y_n would satisfy respectively the equations

$$0 = y'' + yM.e^{-R(t-t_{nk})/D_{nk}}$$

and
$$0 = z'' + z.e^{-R}M.e^{-R(t-t_{n,k+1})/D_{n,k+1}}$$

with the conditions

$$y(t_{nk}) = y(t_{n,k+1}) = z(t_{n,k+1}) = z(t_{n,k+2}) = 0, \qquad M = F_{nk}^{(max)}$$

Now put

$$F(x) = D_{nk}^2 M \cdot e^{-Rx}$$

$$y(t) = Y((t-t_{nk})/D_{nk})$$

$$z(t) = Z((t-t_{n,k+1}/D_{n,k+1})$$

Then Y(x) and Z(x) would satisfy the equations

$$Y'' + F(x)Y = 0$$

$$Z'' + hF(x)Z = 0 \qquad h = (D_{n,k+1}/D_{nk})^2 e^{-R}$$

with the conditions

$$Y(0) = Y(1) = Z(0) = Z(1) = 0$$

Since $x = 0$ and $x = 1$ are consecutive zeros of both Y and Z, h must be equal to unity. For if it was not, one of the functions F and hF would be identically greater than the other. Consequently by virtue of (Ia) of Section 4, at least one of the points $x = 0$ and $x = 1$ would be a point where only one of the functions Y and Z vanished. Hence $D_{n,k+1}/D_{nk} = e^{R/2}$.

If F_n is constantly increasing through both intervals i_{nk} and $i_{n,k+1}$, we should evidently have the second equation of (14d).

The precise limit (14a) shows quite clearly that the only plausible procedure of measuring the differential variation in the gravitation of a trend is by introducing the relative variation in t measured in terms of the zero distance, as we have done.

The kind of differential increase-proportion which is of real significance in time series analysis is certainly a coefficient that indicates what would be the total proportional variation in any of the gravitations F_n in the course of an interval between two consecutive zeros of the trend considered, if the rate of increase kept constant. We are not interested in a coefficient indicating what would be the total proportional variation in F_n during some definite length of time which is defined as a constant for all trends. We may put it another way too: R is now defined so as to give a limit for the maximum proportional variation of the zero distance (formulae (14c) and (14d)). Therefore it has a significance for <u>any</u> trend. If it had not been defined by measuring the relative variation in t in terms of the zero distance, it would have given a limit for the maximum <u>absolute</u> variation in the zero distance. Therefore it would not have been a coefficient with the same significance for trends of low and high order.

We finally have to consider increase-proportions of higher order. The definition of these will be readily suggested by analogy with the higher differential coefficients.

If we have a regular function f(t) and attribute to t some finite but small increase h, the corresponding absolute increase in the function itself may be developed to successive orders of approximation by taking account of the successive powers of h. And the coefficients of this expansion are just the differential coefficients $f^{(i)} = f^{(i)}(t)$, for we have

$$f(t+h) - f(t) = \sum_{i=1}^{\infty} (f^{(i)}/i!)h^i$$

Now if we want to express the <u>relative</u> increase $(f(t + h) - f(t))/f(t)$ of f(t) in terms of the corresponding <u>relative</u> increase h/a of t, where a is some quantity in proportion to which h is measured, we should get an expansion of the form

$$(f(t+h) - f(t))/f(t) = \sum_{i=1}^{\infty} (A^{(i)}/i!).(h/a)^i$$

The coefficients $A^{(i)}$ might be considered as the increase-proportions of higher order of the function f(t), when the increase in t is measured relative to a. We evidently have

$$A^{(i)} = (f^{(i)}a^{i})/f$$

If we apply this to the gravitation F_n, we get two sets of higher order increase-proportions accordingly as we define the relative variation of t in proportion to the actual zero distance D_{nk} or to the fictive zero distance $\pi/\sqrt{F_n}$.

Hence

(15a) $$R_n^{(i)} = F_n^{(i)}D_{nk}^{i}/F_n$$ and (15b) $$r_n^{(i)} = \pi^{i}F_n^{(i)}/F_n^{1+i/2}$$

may be taken as the definition of the higher order increase-proportions of F_n. It is readily seen that they are invariant for a linear transformation of t.

The intuitive significance of these coefficients is evidently that if the increase in t amounts to x% of the actual and the fictive zero distance respectively, then the expansion of the corresponding percentage increase in F_n will be

$$100 \sum_{i=1}^{\infty} (R_n^{(i)}/i!).(x/100)^{i} \text{ and } 100 \sum_{i=1}^{\infty} (r_n^{(i)}/i!).(x/100)^{i}$$

The first order increase-proportions $R_n^{(1)}$ and $r_n^{(1)}$ are of course identical to the increase-proportions R_n and r_n already considered.

The higher increase-proportions $r_n^{(i)}$ will be used in developing the criterion for the applicability of the method of moving differences.

3. PROPOSITIONS CONCERNING THE ZEROS OF A SUM OF TWO CONTINUOUS FUNCTIONS

This section and the next following are mainly concerned with certain propositions on which the argument of the other sections are founded. Some of the propositions are proved in a somewhat more general form than is strictly necessary for our purpose. It has been found that this could be done without introducing materially more complicated considerations.

Let

$$\Phi(t) = f(t) + R(t)$$

f and R continuous.

Let

$$t_k (k = 1,2...) \text{ and } T_K (K = 1,2...)$$

be the real zeros of f and Φ respectively, the zeros being arranged in an ascending order of magnitude, i.e., $t_k < t_{k+1}, T_K < T_{K+1}$. Shorter T will be written for T_K.

Let $D_k = t_{k+1} - t_k$, δ a lower limit of the D_k, δ may be a not precise limit, i.e. δ may even be smaller than the smallest of the D_k. Finally let l_k' $(<D_k)$ and l_k'' $(<D_{k-1})$ be positive numbers, l an upper (not necessarily precise) limit of the numbers l_k' and l_k''.

The intervals $(t_k - l_k'', t_k + l_k')$, $(t_k \pm l)$ and $(T_K \pm l)$ will be designated by i_k, I_k and J_K respectively, or shorter by i, I and J.

<u>I(a)</u> If

$$|f(t_k - l_k'')| > |R(t_k - l_k'')| \qquad |f(t_k + l_k')| > |R(t_k + l_k')|$$

and $$f(t_k - l_k'')f(t_k + l_k') < 0$$
$$k = 1,2\ldots$$

then each i and a fortiori each I contains at least one zero of Φ.

This means that if the i are separated, each zero of f is announced at least once by the vanishing of Φ.

I(b) The first condition of (Ia) is evidently satisfied if $f'(t)$ exists in each i and $|f'| > m_k''$ in $(t_k - l_k'', t_k)$, (f' continuous) $|f'| > m_k'$ in $(t_k, t_k + l_k')$ while $|R(t_k - l_k'')| < l_k'' m_k''$ and $|R(t_k + l_k')| < l_k' m_k'$, m_k'' and m_k' being positive numbers. In this case the i must be separated.

II. If $|f| > |R|$ outside the i, Φ cannot vanish outside the i and a fortiori not outside the I. Consequently if zeros of Φ exist, each J contains at least one zero of f.

This means that the vanishing of Φ never announces a false (non-existing) zero of f.

III(a) If $l < \delta/2$, the I and a fortiori the i are separated and each J contains at most one zero of f.

III(b) If further (II) is satisfied, each J contains exactly one zero of f. But an i may contain any number of T (eventually none). Furthermore two different J may be overlapping both in the case where the two corresponding T belong to the same i and in the case where they belong to two different i, so that still the knowledge of the T is insufficient to locate the t_k in separated intervals.

III(c) If further $l < \delta/4$, two different J cannot overlap unless the two corresponding T belong to the same i.

This means that if zeros of Φ exist, the intervals J occur in clusters in such a manner that the total length of a cluster does not exceed 4l. The clusters are separated and each cluster contains exactly one zero of f. This is situated in the interval j which is common to all J of the cluster.

<u>III(d)</u> If further (Ia) is satisfied, the intervals j of (IIIc) exhaust the zeros of f, f having no zero outside the j, so that the knowledge of the zeros of Φ is sufficient to locate all the zeros of f in separated intervals whose length does not exceed 2l.

<u>IV(a)</u> If Φ is monotonic in which i, each i contains at most one zero of Φ. This condition is evidently satisfied if in each i f is monotonic while f' and R' exist, $|f'|$ being $> |R'|$.

<u>IV(b)</u> If further (IIIa-d) are satisfied, not only the I but also the J are separated, each I containing exactly one zero of Φ and each J exactly one zero of f. Consequently f has no zero outside the J and Φ no zero outside the I. The clusters considered in (IIIc-d) now contain exactly one T each.

This means that not only is the knowledge of the zeros of Φ sufficient to locate separately all the zeros of f (as in the case of (IIId)), but also will each zero of f be announced once and only once by the vanishing of Φ.

The proposition (IIId) might be called the one-sided and (IVb) the general or reciprocal zero proposition.

The proofs of these propositions are evident.

4. UPPER AND LOWER LIMITS FOR REAL SOLUTIONS OF THE LINEAR DIFFERENTIAL EQUATION OF THE SECOND ORDER

Let y and z be real solutions of the two differential equations

$$y'' + Fy = 0$$

$$z'' + Gz = 0$$

F(t) and G(t) being regular and (F - G) not negative in an interval L extending over the whole t-range to be considered.

Let t_o be a value of t in L

$$y(t_o) = a \qquad z(t_o) = b$$

$$y'(t_o) = A \qquad z'(t_o) = B$$

We assume that either are a and b both $\neq 0$, or $a = b = 0$. This means that the curves y and z start from points lying either none or both on the t-axis. We further assume that A and B both $\neq 0$ and finite.

Since a solution of a linear equation cannot have singularities in a point which is not a singular point for at least one of the coefficients, y and z are regular in L. Consequently there must exist an interval of positive length, $i' = (t_o, t_1)$ extending upwards of t_o such that both y and z keep a constant sign in i', limits excepted, at least one of the functions y and z vanishing at the upper limit t_1. Similarly there exists an interval $i'' = (t_2, t_o)$ extending downwards of t_o, such that both y and z keep a constant sign in i''.limits excepted, yz being = 0 at t_2.

We interpret $y'(t_1)/y(t_1)$ as $\lim_{t \to t_1 - o} y'/y$, and $y'(t_2)/y(t_2)$ as $\lim_{t \to t_2 + o} y'/y$.

The sign ε of yz is evidently the same in i' as in i", both in the case where $a = b = 0$ and in the case $a \neq 0$, $b \neq 0$.

A point t in i' or i" such that (F - G) is identically zero between t_0 and t, will be termed a point of identity. Otherwise t will be called a point of non-identity. There must exist a point t' in i' (eventually $t' = t_0$) such that all points between t_0 and t' are points of identity and all points between t' and t_1 are points of non-identity. An analagous point t" exists in i". If t is a point of non-identity, there must exist a finite interval between t_0 and t where all points are points of non-identity. For F and G are supposed continuous. We assume that the first zero of y to the right and left of t_0 is a point of non-identity (if it belongs to (i"+i')).

Let $\Delta = \Delta(t_0) = \varepsilon \begin{vmatrix} a & A \\ b & B \end{vmatrix}$

I. If $\Delta \geqslant 0$

then in i', upper limit included $y'/y < z'/z$ except when t is a point of identity and $\Delta = 0$, in this case $y'/y = z'/z$. The ratio $|z/y|$ is monotonically increasing as t increases, except when $\Delta = 0$ and t increases through points of identity, in this case z/y is constant.

The function y is $= 0$, and $|z| > 0$ at the upper limit t_1.

If $\Delta \leqslant 0$

then in i" lower limit included $y'/y > z'/z$, except when t is a point of identity and $\Delta = 0$, in this case $y'/y = z'/z$. The ratio $|z/y|$ is monotonically increasing as t decreases except when $\Delta = 0$ and t decreases through points of identity, in this case z/y is constant.

The function y is $= 0$, and $|z| > 0$ at the lower limit t_2.

Corollary (a)

Let t_0 be an arbitrary point, $\Delta = \Delta(t_0)$. Then to the right of t_0 y must vanish before z if $\Delta \geqq 0$. To the left of t_0 y must vanish before z if $\Delta \leqq 0$. Therefore if $\Delta = 0$, y must vanish before z both to the right and left of t_0.

The condition $\Delta = 0$ is evidently fulfilled if for instance z is a first order approximation curve to y in the point t_0 (i.e., $a = b$, $A = B \neq 0$) or if t_0 is a common zero of y and z (i.e., $a = b = 0$), no other assumptions being made as to A and B than $A \neq 0$, $B \neq 0$. In this case, therefore, y must vanish before z both to the right and left of t_0.

Corollary (b)

Suppose that $\lim_{t \to t_0} z/y = 1$, i.e. either $a = b \neq 0$ or $a = b = 0$ and $A = B \neq 0$.

If $\Delta \geqq 0$, then

$y'/y < z'/z$ and $|y| < |z|$ in points of non-identity in i' upper limit included.

If $\Delta \leqq 0$, then

$y'/y > z'/z$ and $|y| < |z|$ in points of non-identity in i" lower limit included.

Consequently if $a = b = 0$ and $A = B \neq 0$ (hence $\Delta = 0$), then $|y| < |z|$ in points of non-identity in the total interval (i' + i") limits included, with exception of the point t_0 where $y = z$. And $y'/y < z'/z$ in points of non-identity in i', while $y'/y > z'/z$ in points of non-identity in i".

Further let $a = b = 0$ and $A = B \neq 0$ and let j' and j'' designate the intervals between t_0 and the first points s' s'' of extremum of y to the right and left of t_0 respectively.

Then y'/y is positive in j', negative in j'' (limits excepted) and equal to zero at the upper limit of j' and the lower limit of j''. Consequently if s' and s'' are points of non-identity, z'/z must be positive (not zero) in s' and negative (not zero) in s'', so that the first extremum of z to the right and left of t_0 must lie outside j' and j'' respectively.

In points of non-identity in the total interval $(j' + j'')$ we evidently have

$$(16) \qquad |y'/y| < |z'/z|$$

To prove the preceding propositions let us consider the function

$$(17) \qquad Q(t) = yz' - zy'$$

We have

$$Q'(t) = yz'' - zy'' = yz(F - G)$$

hence

$$(18) \qquad Q(t) = \int_{t_0}^{t} yz(F - G)dt \quad + \varepsilon\Delta$$

The function under the sign of integration in (18) has the sign ε (not zero) in points of non-identity in $(i' + i'')$ except in the points t_2, t_0 and t_1 where yz vanish. Consequently $Q(t)$ has the sign ε (not zero) in points of non-identity in i' if $\Delta \geqslant 0$, and the sign $-\varepsilon$ in points of non-identity in i'' if $\Delta \leqslant 0$. In the first case we have consequently $0 < \varepsilon(yz' - zy')$, and in the second case $\varepsilon(yz' - zy') < 0$. Dividing by $\varepsilon \cdot yz$ (which is positive, not zero, in $(i' + i'')$, limits excepted), we find that

$y'/y < z'/z$ in i' (limits excepted) if $\Delta \gtrless 0$, and $y'/y > z'/z$ in i" (limits excepted) if $\Delta \lessgtr 0$. The only exception is when t is a point of identity and $\Delta = 0$, in which case we evidently have $y'/y = z'/z$.

Hence, $T_1 < T_2$ and $T_3 < T_4$ being arbitrary values of t in i' and i" respectively, (limits excepted), we have

$$|z(T_1)/y(T_1)| < |z(T_2)/y(T_2)| \text{ if } \Delta \gtrless 0$$

$$|z(T_4)/y(T_4)| < |z(T_3)/y(T_3)| \text{ if } \Delta \lessgtr 0$$

The only exception is when T_2 and T_3 (hence T_1 and T_4) are points of identity and $\Delta = 0$, in which case the inequalities are reduced to equations.

Furthermore by virtue of the definition of i', at least one of the two functions y and z must vanish at t_1. If $\Delta \gtrless 0$, they cannot vanish both at t_1. For if they did, Q would vanish too by virtue of (17). But this is impossible, for if both functions vanish at t_1, this point will be a zero of y, hence according to our assumptions a point of non-identity. Consequently Q would have the sign ε (not zero) as shown by (18).

The function that vanishes at t_1 must be y, for if it were z, $\lim_{t \to t_1 - 0} |z/y|$ would be $= 0$ and $|z/y|$ would consequently be decreasing in a finite part of i'. Similar argument regarding i".

That $y'/y < z'/z$ in t_1 if $\Delta \gtrless 0$ and $y'/y > z'/z$ in t_2 if $\Delta \lessgtr 0$ follows from

$\lim_{t \to t_1 - 0} y'/y = +\infty$ and $\lim_{t \to t_2 + 0} y'/y = +\infty$, z'/z being finite at both limits.

The proofs of the corollaries are evident.

<u>II(a)</u> Let y be a solution of $y'' + Fy = 0$, t_0 an arbitrary value

of t, m and M two positive numbers, $0<m<M$.

$$t_m = t_o + (\pi - \arctan a\, m/A)/\, m$$

$$t_M = t_o + (\pi - \arctan a\, M/A)/\, M$$

$a = y(t_o)$, $A = y'(t_o)$, $a/A>0$, and arcus being interpreted as lying between 0 and $+\pi/2$.

Then $t_o < t_M < t_m$.

If $m \leq \sqrt{F} \leq M$ in $(t_o t_m)$ limits included, y keeps a constant sign in $(t_o t_M)$ and vanishes at least once in $(t_M t_m)$. The same proposition holds good in analagous intervals to the left of t_o.

As a special instance we have the proposition that if t_o is a zero of y and $m\leq\sqrt{F}\leq M$ in $(t_o, t_o + \pi/m)$, the only assumptions as to $A = y'(t_o)$ being $A \neq 0$, then y keeps a constant sign in $(t_o, t_o + \pi/M)$ and vanishes at least once in $(t_o + \pi/M,\ t_o + \pi/m)$. If $m\leq\sqrt{F}\leq M$ in $(t_o - \pi/m, t_o)$, then y keeps a constant sign in $(t_o - \pi/M, t_o)$ and vanishes at least once in $(t_o - \pi/m, t_o - \pi/M)$.

These propositions are easily proved by the corollary (Ia) if y is compared with the solution

$$y_M = \sqrt{a^2 + A^2/M^2}\,\sin(\cdot M(t - t_o) + \arctan a\, M/A)$$

of $y'' + M^2 y = 0$ and the solution y_m of $y'' + m^2 y = 0$, where

$$a = y(t_o) = y_m(t_o) = y_M(t_o)$$

$$A = y'(t_o) = y'_m(t_o) = y'_M(t_o)$$

<u>II(b)</u> A consequence of the above propositions is that if $\lim_{t\to\infty} F>0$, zeros of y must always exist. On the other hand, a non-identically vanishing solution of $y'' + Fy = 0$, F being regular, can only have simple zeros. For we have

$$y^{(n)} = -\sum_{i=0}^{n-2} \binom{n-2}{i} F^{(n-2-i)}\, y^{(i)} \qquad (n\geq 2)$$

so that if $y(t_o) = y'(t_o) = 0$ we should also have $y^{(n)}(t_o) = 0$ $(n = 2,3\ldots)$.

The same proposition may be proved without using the expansion of y. In fact z may be chosen so that in (I) $\Delta(t_0) \gtrless 0$, then t_1 is zero of y. If y' vanishes at t_1, Q would vanish too by (17), but would be $\neq 0$ by (18), because G may be chosen so that t_1 is a point of non-identity.

I now proceed to certain propositions concerning limits for a solution y of $y'' + Fy = 0$ which may be formulated by introducing the consecutive zeros of y.

III(a) Let r be a zero of y and s the first point of extremum of y to the right of r ($r<s$). If $0<m\leq\sqrt{F}\leq M$ in $(r, r + \pi/2m)$, then s is situated in the interval $(r + \pi/2M,\ r + \pi/2m)$ and the value of the extremum y(s) lies between A/M and A/m, where $A = y'(r)$. The same proposition holds good for analagous intervals to the left of r.

These propositions follow immediately from (16) of (I b), if y is compared with the functions y_M and y_m defined under(IIa).

The condition that $0<m<F<M$ in $(r,r+\pi/2m)$ may evidently be replaced by the condition that $0<m\leq\sqrt{F}\leq M$ in (rs).

III(b) Let r and r'($r<r'$) be two consecutive zeros of a solution y of $y''+Fy=0$. Then F takes the value $c=(\pi/(r-r'))^2$ at least once in the interval (rr').

Let $z(t)=C\sin\sqrt{c}(t-r)$, C being an arbitrary constant $\neq 0$. Then z satisfies $z''+cz=0$.

Suppose that $F<c$ everywhere in (rr'). As r and r' are consecutive zeros of z, y could vanish in only one of the points r and r' by virtue of (Ia). On the other side if $F>c$ everywhere in (rr'), it would follow that y had a zero between r and r', so that r and r' could not be consecutive zeros of y. Hence F must at least once assume a value > c, and at least once a value < c, and being continuous it must at least once assume the value c.

It is also readily seen that if F is not $\equiv$ c in (rr'), then F must be > c in some finite part of the interval and < c in some other finite part.

The above proposition can also be proved by (IIa).

III(c) Let r'>r>r" be three consecutive zeros of a solution y of y"+Fy=0 (F>0), s' and s" the points of extremum in the intervals i'=(rr') and i"=(r"r) respectively. One and only one point of extremum exists in each of the intervals i' and i" because F>0. Let H' and H" be the values of y(regardless of sign) in s' and s".

Finally let m'<M' and m"<M" be the (not necessarily precise) limits of $\sqrt{F}$ in i' and i" respectively. The limits are supposed to be positive (not zero).

Then we have to the right of r

$$\text{(19a)}\quad (r, r+\pi/M)\qquad Hm/M.\sin M(t-r) < |y(t)| < HM/m.\sin m(t-r)\quad (rr')$$

$$\text{(20a)}\quad (r, r+\pi/2M)\qquad Hm.\cos M(t-r) < |y'(t)| < \begin{cases} HM.\cos m(t-r) & (rs') \\ HM.\cos m(r'-t) & (s'r') \end{cases}$$

The intervals in which the lower and upper limits hold good are indicated in the parentheses to the extreme left and right respectively; m, M and H stand for m', M' and H'.

To the left of r we have

$$\text{(19b)}\quad (r-\pi/M, r)\qquad Hm/M.\sin M(r-t) < |y(t)| < HM/m.\sin m(r-t)\quad (r"r)$$

$$\text{(20b)}\quad (r-\pi/2M, r)\qquad Hm.\cos M(r-t) < |y'(t)| < \begin{cases} HM.\cos m(r-t) & (s"r) \\ HM.\cos m(t-r") & (r"s") \end{cases}$$

where now m, M and H stand for m", M" and H".

We only prove the limits to the right of r. The argument is analagous to the left of r.

Let

$$y_M = A/M \,.\, \sin M(t-r)$$

$$y_m = A/m \,.\, \sin m(t-r)$$

$A = y'(r)$ ($\neq 0$ because the zeros of y are simple)

$$m = m' \qquad\qquad M = M'$$

We have

$$y(r) = y_M(r) = y_m(r) = 0$$

$$y'(r) = y_M'(r) = y_m'(r) = A \neq 0$$

y_M and y_m satisfy respectively

$$y_M'' + M^2 y_M = 0$$

and $$y_m'' + m^2 y_m = 0$$

Noticing that $r+\pi/M < r' < r+\pi/m$ by virtue of (IIIb), we have by corollary (Ib)

$(r, r+\pi/M)$ $\quad |y_M| < |y| < |y_m|$ $\quad (r, r')$

As by (IIIa)

$$mH' < |A| < MH'$$

we immediately deduce (19a).

As

$$\frac{d}{dt}\frac{y'}{y} = -\frac{Fy^2 + y'^2}{y^2} \qquad (0<F),$$

y'/y is decreasing for every value of t. It decreases monotonically from $+\infty$ to $-\infty$ as t runs from one zero of y to the next following, passing zero at the point where y reaches its extremum. To establish the upper limits of (20a) it will therefore be necessary to consider separately the intervals (rs') and (s'r').

Let y_m be defined as above, and let

$$\dot{y}_m = y'(r')/m.\sin m(t-r')$$

Then

$$y(r') = \dot{y}_m(r') = 0$$

$$y'(r') = \dot{y}_m'(r') \neq 0$$

$$\dot{y}_m'' + m^2\dot{y}_m = 0$$

We therefore have by (16) of (Ib)

$$|y'/y| < |y_m'/y_m| \qquad (rs')$$

$$|y'/y| < |\dot{y}_m'/\dot{y}_m| \qquad (s'r')$$

i.e.,

$$|y'/y| < \begin{cases} m \text{ cotan } m(t-r) & (rs') \\ m \text{ cotan } m(r'-t) & (s'r') \end{cases}$$

Now multiply by

$$|y| < H'M/m.\sin m(t-r) \qquad (rr')$$

$$|y| < H'M/m.\sin m(r'-t) \qquad (rr')$$

respectively, the last inequality being obtained from (19b), which holds good also to the left of r', in which case m, M and H would stand for m', M' and H'.

This gives the upper limits of (20a).

To prove the lower limit of (20a) we only have to notice that from (16) of (Ib)

$$|y_M'/y_M| < |y'/y| \qquad (r, r+\pi/2M)$$

<u>III(d)</u> Let y and z be solutions of the equations

$$y'' + F(t)y = 0$$

$$z'' + h^2F(ht)z = 0$$

where $h > 0$ is an arbitrary finite constant.

If y and z are determined so as to be equal to 0 in the point $t = 0$, then

$$z(t) = \frac{z'(0)}{hy'(0)}\, y(ht)$$

From this formula we see that if y and z are determined so as to be equal to zero in the point $t = r$ ($r = 0$ or $\neq 0$), the distance to the next zero of z is $1/h$ of the distance to the next zero of y. This holds good both to the left and right of r and regardless of the values ($\neq 0$) which have been attributed to $z'(r)$ and $y'(r)$.

To prove the above formula we put $Z(t) = z(t/h)$.

It is readily seen that both Z and y are solutions of the same equation

$$Z'' + F(t)Z = 0$$

with $Z(0) = y(0) = 0$. Hence Z and y can only differ by the determination of $Z'(0)$ and $y'(0)$.

Now let Z here stand for z of (I). If r is a common zero of Z and y we have $\Delta(r) = 0$. Hence if r' is the next following zero of y, the ratio Z/y is constant as t increases in (rr') through points of identity. Consequently for points in (rr')

$$Z/y = \lim_{t \to r} Z/y = Z'(r)/y'(r)$$

From the demonstration of (I) it is seen that this holds good even if the next zero of y is a point of identity. Hence

$$\lim_{t \to r'} Z/y = Z'(r')/y'(r') = Z'(r)/y'(r)$$

Consequently for all values of t

$$Z(t)/y(t) = Z'(r)/y'(r)$$

which proves the proposition.

5. THE METHOD OF NORMAL POINTS

This section will be concerned with the following problem. Given a composite curve $w = \sum y_n$ where the y_n are trends of successive orders, each y_n being a solution of $y_n'' + F_n y_n = 0$. The various F_n are supposed positive and of descending order of magnitude in the sense explained in Section 2 (formula (9a)). How far will the zeros of the second differential coefficient w'' of the composite curve indicate approximately the location of the normal points (i.e., the zeros) of the trend of lowest order y_0? If the approximation is accepted as sufficiently accurate, how can the normal points (i.e., the zeros) and the ordinates of the successive trends y_n be computed?

I shall first make an intuitive approach to the problem and develop the computing scheme which may be derived therefrom. I then proceed to the rigorous proof of the theorem which establishes the criterion of the closeness of the approximation used.

An idea which will naturally present itself is that according to the conception of y_n as trends of different orders, the curvature of y_0 will generally be far greater than the curvature of y_1, and still greater than the curvature of y_2, etc. Therefore, y_0'' will generally be the predominant term in w''. However, this does not hold good in the vicinity of the very points with which we are here concerned, viz. the zeros of y_0. For in these points y_0'' will vanish on account of its containing y_0 as a factor besides the finite factor $(-F_0)$. But only a slight displacement from the zero of y_0 will generally suffice to reestablish the predominance of y_0'' over the other terms in w". And the reason is that in the zeros of y_0 the first differential coefficient y_0' is at its maximum (regardless of sign).

What we have to do, is therefore, to consider intervals encircling the zeros of y_0, as small as possible and just large enough to reestablish the predominance of y_0'' outside these intervals. It then follows that between a zero of w" (which is known) and the corresponding zero of y_0 (which is not known but lies somewhere in the neighborhood of the zero of w") the distance cannot exceed the intervals considered.

A priori it seems obvious that these intervals may be determined all the smaller the greater is the difference in order between the successive trends. (Difference in order being defined by (6ab) or (9ab) of Section 2). The closeness of approximation, therefore, seems to depend on the relative difference in order, and this prima facie impression will be verified by the exact carrying out of the analysis.

Now if we suppose that the difference in order is sufficiently great to insure a good approximation, then the normal points of y_0 may be considered as determined by the location of the zeros of w".

Further, since $y_0 = 0$ in the normal points, the data relating to these points may be considered as a new series where the trend of lowest order y_0 is eliminated.

And this new series may be treated in the same way, thus eliminating the trend of next higher order y_1 , and so forth.

Let W_0 be the original series where all the trends $y_0 y_1 \cdots y_N$

are present. And let $W_1 W_2 \dots W_N$ be the series successively derived from W_0 by the method of normal points. Then W_n contains only the trends $y_n y_{n+1} \dots y_N$. This is how the successive trends can be eliminated. Now to actually construct the ordinates of the trends we may proceed in the following way.

When the difference in order between the successive trends is great, the ordinate of the trend of lowest order y_0 will be represented approximately by the deviation of the actual data W_0 from a line interpolated in some way or another through the series W_1 which represents the normal points for y_0. We may, for instance, draw a straight line between every two consecutive normal points (i.e., between every two consecutive data in W_1), or draw a m-th order parabola through (m+1) consecutive normal points, or use any other method of curve fitting. The essential point is that whatever the method used may be, the data W_1 determining the interpolation line, are data <u>where the lowest order trend y_0 is already eliminated.</u>

Similarly the ordinates of a higher order trend y_n is determined by the deviation of W_n from a line interpolated through W_{n+1} by one of the methods referred to.

This is the essense of the method of normal points. In Section 6 we shall consider another method of determining the trends y_n, which may be used even in the case where the difference in order between any of the trends is not great. This method, however, is subject to more restrictive assumptions as to the differential nature of the gravitations of the successive trends.

I now proceed to some points connected with the actual computation of the normal points.

In practically all cases the data in the given series $w(=W_0)$ will be equidistant. I shall, therefore, make this assumption. Let w_t be the items in the given series for the moments of time $t = 1,2...$

The second differential coefficient of w has to be approximated by the second difference of w_t. It will hardly ever be necessary to carry the approximation further than simply putting the second differential coefficient of w equal to the second difference.

$$w''(t) = \Delta^2 w_{t-1} = w_{t+1} - 2w_t + w_{t-1} \tag{21}$$

To study the location of the zeros of $w''(t)$, we therefore have to plot the variation of $\Delta^2 w_{t-1}$ and determine the points where this curve passes zero.

It should be noticed that the operation (21) might be carried out graphically in the plot of the original data simply by measuring the deviation of w_t from the straight line through w_{t+1} and w_{t-1}.

It is quite clear that a successful determination of the zeros of w" by the variation of $\Delta^2 w$ depends on the following conditions

(a) The distance between the data in w (which will also be the distance between the known values of $\Delta^2 w$) must be small in relation to the actual distance between the zeros in w".

(b) The accidental errors in the data must not be so dominating that a number of false zeros is indicated in $\Delta^2 w$.

By a false zero is meant a zero in $\Delta^2 w$ which does not correspond to actually existing zeros in w".

Let us take an illustration. Suppose the data are monthly and we intend to eliminate the seasonal fluctuations (considered as the trend of lowest order y_0).

Even if the accidental errors are quite insignificant, the variation in $\Delta^2 w$ will not give a good location of the zeros of w" if the actual seasonal fluctuations y_0 are very complicated say have peaks and vallies at the short distance of one or two months, because in this case y_0'' and hence w" will change sign at so short notice that it might not be indicated in $\Delta^2 w$. This follows from the simple fact that if we only know the value of a function in certain discrete points, we are not able to trace the variation of this function within intervals whose length does scarcely exceed the distance between the discrete points where the value of the function is known.

To this must be added the effect of accidental errors. If accidental errors are already present in w, they will be all the more dominating in $\Delta^2 w$. Now it can be proved that the magnifying of the error-effect, which is introduced by the operation Δ^2, will be approximately balanced if the original data are smoothed by a twice iterated, moving quarterly average. (Smoothing the original data or the series $\Delta^2 w$ evidently amounts to the same.)

I shall not prove this proposition, which would necessitate the opening of a new section. I hope, however, to be able to revert to the general question of the smoothing of time series in another article.

Now if a twice iterated, moving quarterly average is applied, the approximation formula for w" will be

(22) $$w''(t) = \Delta^2\dot{w}_{t-1} = (w_{t+3} - 2w_t + w_{t-3})/9$$

$\dot{w}_t$ designating the smoothed series.

As is readily seen $\Delta^2\dot{w}$ may be determined graphically from the plot of the original data simply by measuring the deviation of w_t from the straight line through w_{t+3} and w_{t-3}.

If a smoothing of this kind is necessary, the zeros of w" cannot be determined with any degree of accuracy unless the zero distance in the seasonal fluctuations amounts to say three or four months at least, the degree of accuracy depending, of course, also to some extent on the shape of w".

If weekly data are available, or if the data are monthly while the trend of lowest order y_0 to be eliminated is the ordinary business cycle (the seasonal fluctuations being insignificant or already eliminated by some other method, for instance by a moving yearly average, arithmetic or geometric), then the case is quite different. The number of data in each period is then greater, and the accuracy will be much closer.

It should be pointed out that if the plot of $\Delta^2 w$ or of $\Delta^2\dot{w}$ shows quite distinctly certain intervals of positive and others of negative values, while the curve passes zero in a cluster of points lying between an interval of distinct positive and an interval of distinct negative values, this cluster should only count for one normal point (i.e. one zero in w") If, for instance, the trend of lowest order y_0 to be eliminated is the business cycle, and the plot of $\Delta^2 w$ or of $\Delta^2\dot{w}$ shows a distinct positive period say from the beginning of 1901 to May 1903 and a distinct negative period from August 1903 to the end of 1905, while the item for June 1903 is negative and for July positive, this should count only for one normal point lying between June and July, 1903.

When the trend of lowest order is eliminated and we proceed to the determination of the normal points in the derived series $W_1, W_2 \ldots$, then the data will no longer be equidistant and we have to approximate the second differential coefficient by divided differences (instead of ordinary differences). But otherwise the procedure will be the same as for the series W_0.

It is obvious that the number of data in the successive series $W_0, W_1 \ldots$ is rapidly diminishing. We finally arrive at a series W_N containing only some few data. This series W_N may be taken as representing the trend of highest order y_N.

I now proceed to establish the theorem regarding the closeness of approximation with which the zeros of y_0 may be determined from the knowledge of the zeros of w".

Let p, a_n, b_n, A_n, B_n, A and R be the numbers defined in Section 2.

The criterion for the error committed by taking the zeros of w" as the zeros of y_0, may be stated by introducing either the numbers a and b or the numbers A and B.

The limits established by a and b are the sharpest. In return the limits established by A and B have the advantage that the signification of the numbers of A and B is more intuitive. It would certainly be possible to prove limits much sharper than those of the following theorem, by introducing more complicated expressions. Our aim has been rather to establish formulae that might be easily interpreted than to obtain the sharpest possible limits.

It is comparatively easy to establish an upper limit for the ratio between the error committed and the actual zero distance in y_0, in which the zero of w" in question is situated. This limit is furnished by (A) of the following theorem, which already gives a valuable information regarding the accuracy.

However, we want to know more than this. We want to express the error in proportion to the actual zero distance in w", and to distinguish the case in which the absolute error intervals encircling the zeros of w" are not overlapping or at any rate only overlapping in such a manner as not to prevent the unambiguous location of the zeros of y_0 (case of the one-sided zero proposition indicated in Section 3.), further to distinguish the case in which the correspondence between the zeros of w" and y_0 is a one to one correspondence.

This is done by (B) and (C) of the following theorem. The proposition (B) is uniform in the sense that it holds good for the whole t-range without regard to the variation in the length of the zero distance in y_0 or w". In (C) account is taken of the relative length of the consecutive zero distances in w" (which are known). In return (C) assumes less than (B) with regard to the difference in order between the successive trends.

Before formulating the theorem I shall define some notations, and state certain conditions, which are involved in various combinations in the following propositions:

The first conditions are

(23a) $$p \sum a_n b_n < 1$$

(23b) $$p^3 \sum A_n B_n < 1$$

where $\sum$ designates $\sum_1^N$

These conditions are fundamental. One of them must be fulfilled if any of the propositions below shall hold good.

Further we shall have to consider the conditions

(24a) $$p\sum a_n b_n < \sin \pi/4$$

(24b) $$p^3 \sum A_n B_n < \sin \pi/4$$

(25a) $$p \sum a_n b_n < \sin(\lambda \pi/4)$$

(25b) $$p^3 \sum A_n B_n < \sin(\lambda\pi/4)$$

where $\lambda = D^{(min)}/D^{(max)}$ designates the ratio of the smallest zero distance in y_0 to the greatest.

Further,

(26a) $$\begin{cases}(D_{K+1}+D_{K-1})/D_K < (\pi/u) - 2 \\ (D_K + D_{K-2})/D_{K-1} < (\pi/u) - 2\end{cases}$$

(26b) $$\begin{cases}(D_{K+1} + D_{K-1})/D_K < (\pi/U) - 2 \\ (D_K + D_{K-2})/D_{K-1} < (\pi/U) - 2\end{cases}$$

where

$$D_K = T_{K+1} - T_K$$

T_K $(K = 1,2\ldots)$ being the zeros of w'' arranged in an ascending order of magnitude, and u and U designating

(27a) $u = \arcsin p \sum a_n b_n$ u being interpreted as $0<u<\pi/2$

(27b) $U = \arcsin p^3 \sum A_n B_n$ U being interpreted as $0<U<\pi/2$

For reasons which will be evident by the following, the conditions (24ab), (25ab) and (26ab) will be called the conditions for non-overlapping, or shorter, the non-overlapping conditions.

Further,

(28a) $$R\sqrt{p} + A(R+\pi) < \pi \cos u/u$$

(28b) $$R\sqrt{p} + A(R+\pi) < \pi \cos U/U$$

If (24a) or (24b) is satisfied, then (28a) or (28b) may be replaced by

(28c) $$R\sqrt{p} + A(R+\pi) < 2\sqrt{2}$$

For $\pi \cos x/x$ is monotonically decreasing from $+\infty$ to $2\sqrt{2}$, when x runs from zero to $\pi/4$. Evidently (28ab) assumes less than (28c).

The conditions (28abc) will be called the conditions for one to one correspondence or shorter, the one to one conditions.

The combinations of these conditions to be considered will always be either a combination where only (a) formulae enter or a combination where only (b) formulae enter. The first will be called an (a) case, the second a (b) case.

Theorem of normal points.

A. If (23a) is satisfied, the presence of a normal point in the lowest order trend y_0 must be announced at least once by the vanishing of w". A vanishing of w" which announces a zero r of y_0 must take place in the interval (r-h", r+h') where

$$h' = (u/\pi)D'$$

$$h'' = (u/\pi)D''$$

D' and D" designating the distance from r to the next following zero of y_0 to the right and left of r respectively, and u being defined by (27a). The interval (r-h", r+h') does not overlap with any of the analagous intervals encircling the other zeros of y_0.

Further, w" cannot vanish outside the intervals considered, i.e., the vanishing of w" never announces a false (non-existing) zero of y_0. Hence the normal points of y_0 can never be displaced from the zeros of w" by more than a time interval amounting to a fraction u/π of the zero distance in y_0 in which the zero of w" in question is situated.

If (23b) is satisfied, the same proposition holds good when u is replaced by U (formula(27b)).

B. If (25a, 28a) or (25b,28b) is satisfied, then each normal point (i.e. zero) of the lowest order trend y_o will be announced once and only once by the vanishing of w".

If T_K (K=1,2...) designates the zeros of w", and we put

(29a) $l = l_u = D^{(max)}u/\pi$ in the (a) case

(29b) $l = l_U = D^{(max)}U/\pi$ in the (b) case

where $D^{(max)}$ is the greatest of the zero distances in y_o, then the intervals $(T_K \pm l)$ do not overlap and y_o has one and only one normal point in each of these intervals and no normal point outside them.

If (25a) or (25b) but neither (28a) nor (28b) is satisfied, the knowledge of the zeros of w" is still sufficient to locate all the normal points of y_o in separated intervals in the manner indicated in (IIIcd) of Section 3. But a normal point of y_o may now be announced more than once by the vanishing of w".

C. If (23a, 28a) or (23b, 28b) is satisfied, then each normal point in y_o will be announced once and only once by the vanishing of w".

If further for a particular zero T_K of w" (26a) or (26b) is satisfied, and we encircle T_K by the interval

$$J_K = (T_K - L''_K,\ T_K + L'_K)$$

where

(30a) $L'_K = D_{K-1}/(\pi/u-2)$ $L''_K = D_K/(\pi/u-2)$ in the (a) case

(30b) $L'_K = D_{K-1}/(\pi/U-2)$ $L''_K = D_K/(\pi/U-2)$ in the (b) case

$$D_K = T_{K+1} - T_K$$

then J_K will not overlap either with the analagous interval J_{K-1} to the left or the interval J_{K+1} to the right; y_o has one and only one normal point in J_K and no normal point neither between J_{K-1} and J_K nor between J_K and J_{K+1}.

To prove these propositions we first notice that the zeros of y_o are the very points in which y_o'' vanishes. It is therefore sufficient to consider the zeros of y_o''.

Let
$$z = \sum_{n=1}^{N} y_n$$

then
$$w'' = y_o'' + z''$$

Further, let t_{oj} $(j=1,2,\ldots)$ be the zeros of y_o''.

The propositions (ABC) will be proved if it is possible to specify non-overlapping intervals

$$(t_{oj} - l_{oj}'', t_{oj} + l_{oj}')$$

such that the conditions (Ia), (II), (IIIc) and (IVa) of Section 3 are fulfilled. The functions w'', y_o'' and z'' here stand for Φ, f and R of Section 3 respectively.

In particular (A) will be established when it can be shown that (Ia) and (II) of Section 3 hold good, and further that the intervals $(t_{oj}-l_{oj}'', t_{oj}+l_{oj}')$ are included in the intervals of proposition (A), these intervals being separated. (B) will be established if it can be shown that not only (Ia) and (II) but also that (IIIc) and (IVa) of Section 3 hold good. If (Ia), (II) and (IVa) (but not (IIIc)) hold good, and the intervals $(t_{oj}-l_{oj}'', t_{oj}+l_{oj}')$ are separated, the first part of (C) is established. The second part of (C) is then to be proved by (26a) or (26b).

What we have to do is therefore to show that (Ia) and (II) of Section 3 follows from (23a) or (23b), (IIIc) from (25a) or (25b) and finally (IVa) from (28a) or (28b).

I shall first show that the conditions of Section 3 referred to hold good if the (a) conditions of the present section are fulfilled, and we choose

$$l'_{oj} = (u/\pi)D_{oj}\sqrt{p_{oj}}$$
$$l''_{oj} = (u/\pi)D_{o,j-1}/\sqrt{p_{o,j-1}}$$

where D_{oj} is defined by (5a) and p_{oj} by (10a).

Let s_{nk} be the point of extremum of y_n in the intervals i_{nk} defined in (5b). And let $H_{nk} = |y_n(s_{nk})|$.

The apostrophes ' and " will be used throughout the following analysis to designate quantities to the right and left of t_{oj} respectively.

For the sake of brevity we put

	p'	p''	q'	q''
for	p_{oj}	$p_{o,j-1}$	q_{oj}	$q_{o,j-1}$

and r''	s''	t''	r	t'	s'	r'
for $t_{o,j-1}$	$s_{o,j-1}$	$t_{oj}-l''_{oj}$	t_{oj}	$t_{oj}+l'_{oj}$	s_{oj}	$t_{o,j+1}$

The quantities in the last two lines are written in an ascending order of magnitude, for it will be shown that $s'' < t''$ and $t' < s'$; q_{oj} is defined by (10b).

Further we write

	l'	l''	D'	D''	H'	H''
for	l'_{oj}	l''_{oj}	D_{oj}	$D_{o,j-1}$	H_{oj}	$H_{o,j-1}$

By (10c) the upper limit of F_o in (rr') is $(\pi/D')^2 p'$. The lower limit of (19a) Section 4 therefore holds good in $(r, r+D'/\sqrt{p'})$ and a fortiori in (rt') because $u < \pi$ and hence $l' < D'/\sqrt{p'}$. We consequently have by putting $m = \pi/D'\sqrt{q'}$,

$M = \pi\sqrt{p'}/D'$ and $t = t'$ in (19a)

$$|y_o''(t')| > (H'F_o(t')/\sqrt{p'q'}) \sin(\pi\sqrt{p'}l'/D')$$

Since $\qquad \sin(\pi\sqrt{p'}l'/D') = \sin u = p\sum a_n b_n$

and $\qquad \sqrt{p'q'} \leq p$

we have

(31) $$|y_o''(t')| > H'F_o(t')\sum a_n b_n$$

On the other side we have for any value of t in (rr')

$$|y_n''(t)| \leq H_{nk}F_n(t)$$

where k is defined by (5d).

Hence by (9ab)

$$|y_n''(t)| \leq H'F_o(t)a_n b_n$$

Consequently

(32) $$|z''(t)| \leq \sum |y_n''(t)| \leq H'F_o(t) \sum a_n b_n$$

for any value of t in (rr').

Therefore

$$|y_o''(t')| > |z''(t')|$$

An analagous argument shows that

$$|y_o''(t'')| > |z''(t'')|$$

Furthermore we have by (IIIa) of Section 4, $s'-r > D'/2\sqrt{p'}$ which is $>l'$ since $u < \pi/2$. Similarly $r-s'' > l''$.

Therefore y_o is monotonic in $(t''t')$. Consequently $y_o(t')$ and $y_o(t'')$ are of opposite sign. As F_o is essentially positive, $y_o''(t')$ and $y_o''(t'')$ must also be of opposite sign.

The conditions (Ia) of Section 3 are therefore fulfilled.

In $(t's')$ we have

$$|y_o(t)| > |y_o(t')| \text{ which by (31) is} > H'\sum a_n b_n.$$

In $(s',r'-l''_{o,j+1})$ we have

$$|y_o(t)| > |y_o(r'-l''_{o,j+1})|$$

As $l''_{o,j+1} = l'_{oj} = l'$, we get by applying the lower limit of (19b) for the point r' instead of r

$$|y_o(r'-l''_{o,j+1})| > H'\sum a_n b_n$$

so that the limit

(33a) $$|y_o(t)| > H'\sum a_n b_n$$

holds good in the whole interval $(t', r'-l''_{o,j+1})$.

Consequently in the same interval

(33b) $$|y''_o(t)| > H'F_o(t)\sum a_n b_n \text{ which by (32) is} \geqslant |z''(t).|$$

An analagous argument shows that

$$|y''_o(t)| > H''F_o(t)\sum a_n b_n \geqslant |z''(t)|$$

in $(r''+l'_{o,j-1}, t'')$.

The condition (II) of Section 3 is therefore fulfilled. Hence w'' has at least one zero in each $(t_{oj}-l''_{oj},\ t_{oj}+l'_{oj})$ and no zero outside these intervals. The same will be true of any set of intervals which are defined so that each of them contains a $(t_{oj}-l''_{oj},\ t_{oj}+l'_{oj})$. Such a set of intervals is $(t_{oj}-h''_{oj}, t_{oj}+h'_{oj})$ where

$$h'_{oj} = (u/\pi)D_{oj}$$
$$h''_{oj} = (u/\pi)D_{o,j-1}$$

It is readily seen that each of the new intervals contains one of the original for we have $p_{oj} \geqslant 1$, hence $h'_{oj} \geqslant l'_{oj}$ and $h''_{oj} \geqslant l''_{oj}$.

And the new intervals do not overlap. For $h'_{oj} + h''_{o,j+1} = 2h'_{oj}$ is less than D_{oj} because $u < \pi/2$.

This proves the proposition (A) in the (a) case.

The intervals $(t_{oj}-l''_{oj}, t_{oj}+l'_{oj})$ must evidently a fortiori be non-overlapping.

To demonstrate that (Ia) and (II) of Section 3 are fulfilled, we have only used (23a).

That (IIIc) of Section 3 holds good when (25a) is satisfied, simply follows from the fact that $l_u = D^{(max)} u/\pi$ is an upper limit of the numbers l'_{oj} and l''_{oj} and this limit is $< D^{(min)}/4$ if $u < \lambda\pi/4$.

The foregoing limits could be developed without taking account of the differential variations of the various F_n (measured by R). To establish (IVa) of Section 3 we shall have to introduce R and use the one to one condition (28a).

We have to establish an upper limit of $|z'''(t)|$ holding good for any point in $(t''t')$ and show that $|y_o'''(t)|$ is greater than this limit in any point in the same interval. As the upper limit of $|z'''(t)|$ must be essentially positive (not zero), it will follow that y_o'' is monotonic in the interval considered.

Evidently for any value of t

$$|y_o'''(t)| \geqslant F_o(|y_o'| - |y_o| \cdot |F_o' / F_o|)$$

Let us first consider the interval (rt'). This interval is included in $(r, r+D'/2\sqrt{p'})$ since $u < \pi/2$, so that we may use the upper limit of (19a) and the lower limit of (20a) by putting $m = \pi/D'\sqrt{q'}$ and $M = \pi\sqrt{p'}/D'$. Further by (12a) $|F_o'/F_o| \leqslant R/D'$

This gives in (rt')

$$|y_o'''(t)| >$$
$$F_o(t)\, H'/D' \left[(\pi/\sqrt{q'}) \cos(\pi\sqrt{p'}(t-r)/D') - R.\sqrt{p'q'}\, \sin(\pi(t-r)/D'\sqrt{q'}) \right].$$

Since in (rt')

$$\pi\sqrt{p'}(t-r)/D' \leqslant \pi\sqrt{p'}\,l'/D' = u < \pi/2$$

and $$\pi(t-r)/D'\sqrt{q'} = \pi\sqrt{p'}(t-r)/D'\sqrt{p'q'} \leqslant u/\sqrt{p'q'} < \pi/2$$

we have a fortiori

$$|y_o'''(t)| > F_o(t) H'/D' \left[(\pi/\sqrt{q'}) \cos u - R\sqrt{p'q'} \sin(u/\sqrt{p'q'}) \right]$$

Since $o<u/\sqrt{p'q'}<\pi/2$ we have $\sin(u/\sqrt{p'q'})<u/\sqrt{p'q'}$.

Further $$1/\sqrt{q'} \geqslant 1/\sqrt{p}$$

Hence a fortiori in any point t in (rt')

(34) $$|y_o'''(t)|> F_o(t)H'/D'\left[(\pi/\sqrt{p})\cos u - Ru\right]$$

On the other side we have for any point t

$$|y_n'''(t)| \leqslant F_n(t)(|y_n'| + |F_n'/F_n|\cdot|y_n|)$$

Now let t be any point in (rt'), and let k be defined by (5d).

As $\cos \leqslant 1$, we have by the upper limits of (20a)

$$|y_n'(t)|<\pi H_{nk}\sqrt{p_{nk}}/D_{nk} \leqslant \pi H_{nk}\sqrt{p}/D_{nk}$$

Further

$$|F_n'/F_n|\cdot|y_n|< H_{nk}R/D_{nk}$$

That is

(35) $$|y_n'''|<(F_nH_{nk}/D_{nk})(\pi\sqrt{p} + R)< (F_nH_{nk}/D_{nk})\sqrt{p}(\pi+R)$$

Using (6a) and (9ab) we get

$$|y_n'''|<(F_oH'/D')\sqrt{p}(\pi+R)Aa_nb_n$$

so that in any point t in (rt')

(36)

$$|z'''(t)|<(F_o(t)H'/D')A(\pi+R)\sqrt{p}\sum a_nb_n= (F_o(t)H'A(\pi+R)/D'\sqrt{p})\sin u$$

Comparing with (34) we see that in order to ensure $|y_o'''(t)|>|z'''(t)|$ in any point t in (rt') it is sufficient that

(37) $$A(\pi+R)\sin u<\pi\cos u - R\sqrt{p}\,u$$

As $\sin u<u$ when $o<u$, it is a fortiori sufficient that

$$A(\pi+R)u<\pi\cos u - R\sqrt{p}\,u$$

which is the condition (28a) of the theorem.

A quite analogous argument shows that $|y_o'''(t)|>|z'''(t)|$ in any point in (t"r). The condition (IVa) of Section 3 is therefore fulfilled.

That the second part of (C) follows from (26a), may be proved thus.

Consider the intervals $(t_{oj}-l''_{oj}, t_{oj}+l'_{oj})$. These intervals are separated and w" has one and only one zero in each of them, and no zero outside them. For it has been shown that (Ia), (II) and (IVa) of Section 3 follows from (23a) and (28a) and further that the intervals considered are separated.

Now let T_K be the zero of w" in $(t_{oj}-l''_{oj}, t_{oj}+l'_{oj})$. If we move from any point in this interval, a distance l'_{oj} to the left and a distance l''_{oj} to the right, we shall surely cover t_{oj}. Therefore if we put $L'_K \geqslant l''_{oj}$ and $L''_K \geqslant l'_{oj}$, the intervals J_K of proposition (C) will surely contain at least one zero of y_o.

Now $D_K = T_{K+1} - T_K$ must be greater than

$$(t_{o,j+1}-l''_{o,j+1}) - (t_{oj}+l'_{oj}) = D_{oj}-2l'_{oj} \geqslant D_{oj}(1-(2u/\pi))$$

Hence $l'_{oj} \leqslant (u/\pi)D_{oj}$ is $\leqslant (u/\pi)D_K/(1-(2u/\pi))$

It is therefore sufficient to put

$$L''_K = D_K/(\pi/u-2)$$

Similarly it is sufficient to put

$$L'_K = D_{K-1}/(\pi/u-2)$$

If L'_K and L''_K are chosen in this way, it is readily seen that not only must y_o have at least one zero in each of the intervals J_K, but it can have no zero outside them.

Furthermore if any particular one of these intervals does not overlap either with the analagous interval J_{K-1} to the left or the interval J_{K+1} to the right, then J_K must contain exactly one zero of y_o.

Now the condition that J_K shall not overlap with J_{K+1} is

(38a) $$L'_K + L''_{K+1} < D_K$$

i.e. $$(D_{K-1} + D_{K+1})/D_K < \pi/u-2$$

And the condition that J_K shall not overlap with J_{K-1} is

(38b) $$L'_{K-1} + L''_K < D_{K-1}$$

i.e. $$(D_{K-2} + D_K)/D_{K-1} < \pi/u-2$$

which are the conditions of the proposition.

It should be noticed that while the numbers 1' and 1" are defined as a fraction of the zero distance in y_o which lies to the same side of t_{oj} as 1' and 1" respectively, the numbers L' and L" are defined as a fraction of the zero distance in w" which lies to the opposite side of T_K as L' and L" respectively.

The demonstration of the theorem of normal points by using the (b) conditions is not very different from the foregoing. The main difference is that we now approximate F_n by (10c) and therefore have to introduce the numbers A_n and B_n.

We shall use the same notation as before. The numbers $1'_{oj}$ and $1''_{oj}$ have now to be taken equal to

$$1'_{oj} = (U/\pi)D_{oj}/\sqrt{p_{oj}}$$
$$1''_{oj} = (U/\pi)D_{o,j-1}/\sqrt{p_{o,j-1}}$$

In this case too we have $1' < D_{oj}/\sqrt{p_{oj}}$. Therefore the lower limit of (19a) may be used, which gives

(39) $$|y''_o(t')| > H'(\pi/D')^2 p \sum A_n B_n$$

On the other side we have for any point t in (rr')

$$|y''_n(t)| \leqslant H_{nk}(\pi/D_{nk})^2 p_{nk}$$

where k is defined by (5d).

Hence by (6ab)

$$|y''_n(t)| \leqslant H'(\pi/D')^2 p A_n B_n$$

Consequently for any point t in (rr')

(40) $$|z''(t)| \leqslant H'(\pi/D')^2 p \sum A_n B_n$$

This formula corresponds to the formula (32) developed by the (a) conditions; (39) and(40) shows that $|y_0''(t')| > |z''(t')|$. Same argument to the left of r. As $y_0''(t')y_0''(t'') < 0$, (Ia) of Section 3 is fulfilled.

It is further easily proved that the limit

(41) $$|y_0''(t)| > H'(\pi/D')^2 p \sum A_n B_n$$

which is analagous to (33b), holds good in the whole interval $(t;r'-l''_{o,j+1})$. As (40) holds good in the same interval, and a similar argument may be applied to the left of r, (II) of Section 3 is satisfied.

To establish that (IVa) issatisfied, we notice that to the right of r the formula (34) evidently holds good if we change u in U. Hence by (10c)

(42) $$|y_0'''(t)| > (H'/D')(\pi/D')^2 p^{-1} \left[(\pi/\sqrt{p})\cos U - RU\right]$$

Further from (35) by using (6ab) and (10c)

$$|y_n'''| < (H'/D')(\pi/D')^2 p^{3/2}(\pi + R) A A_n B_n$$

i.e.

(43) $$|z'''(t)| < (H'/D')(\pi/D')^2 p^{-3/2} A(\pi + R) \sin U$$

in any point t in (rt').

Comparing with (42) we see that we fell back on the condition (28b) of the theorem.

Analagous argument to the left of r. Hence (IVa) of Section 3 is fulfilled.

I now proceed to an interpretation of the significance of some of the formulae obtained. I shall try to form an idea about the closeness of approximation with which the method of normal points may be actually carried out.

The main feature of the accuracy of approximation is the magnitude of the angles u and U respectively, (formulae (27ab)). It is quite clear that we cannot make any exact computation of these angles because we do not know the exact value of p, a_n, b_n and A_n, B_n respectively. The value of these numbers evidently varies from one series to another. But it is possible to make a rough estimation of the value which these quantities will have in a time series of the usual type when the components are really trends of different orders.

Let us first see what the result would be if the numbers a_n and b_n had the values suggested in Section 2 for the analagous numbers A_n and B_n, viz., $a_n = (1/7)^n$ and $b_n = (\frac{1}{2})^n$. This means that each cycle contains at least seven cycles of the next lower order, for instance each major cycle contains at least seven ordinary cycles, and that the intensity (distinctness) of the oscillations of each trend is at most half that of the next lower. Let us further assume p = 1,5. This means that the total fluctuation in the various gravitations is such that the ratio between the greatest and smallest of two consecutive zero distances in any trend can go up to 1,5 but not more. The closeness of approximation will be all the better the smaller the number of trends, but let us make no assumption in this respect. Let us calculate u as if the number of trends were infinite.

In this case we should have

$$p \sum a_n b_n = 1,5 \sum_{n-1}^{\infty} (1/14)^n = 1,5/13 = 0,115$$

Hence the fundamental condition (23a) is largely satisfied. We further get (formula (27a)) $\sin u = 0{,}115$, hence $u = 6^{0}35'$. Consequently $(u/\pi) = 0{,}036 = 3{,}6\%$.

This means (proposition A) that the normal points in the trend of lowest order y_0 cannot be displaced from the known zeros of the second differential coefficient w'' of the composite curve by more than a time interval amounting to at most 3,6% of the actual zero distance in y_0 in which the zero of w'' in question is situated.

Now let us drop the assumption that the intensity of the oscillations of the higher trends is smaller than that of the lower trends, i.e., we now assume that the ratio between the maximum amplitude and the zero distance in any trend may be as great as the same ratio in y_0. Hence $b_n = 1$ and

$$p \sum a_n b_n = 1{,}5 \sum_{n=1}^{\infty} (1/7)^n = 1{,}5/6 = 0{,}25$$

The fundamental condition (23a) is still largely satisfied. We get $u = 14^{0}30'$, hence $(u/\pi) = 8\%$.

Finally let us make some still more unfavorable assumptions. Suppose that $a_n = (1/5)^n$, $b_n = 1$. This means that each cycle may contain only five cycles of the next lower order. And the intensity of the oscillations of the higher order trends may be as great as that of y_0. Let us further suppose $p = 2$. This means that two consecutive zero distances in y_0 may differ by 100%. In this case too the fundamental condition (23a) is largely satisfied and we get $u = 30^{0}$, hence $(u/\pi) = 16{,}7\%$.

The accuracy of the approximation is now decidedly weaker, but still it is not quite useless. Furthermore it must be remembered that the limits used in the demonstration of the theorem have been very large. Therefore if the error in the actual determination of any particular normal point of y_0 should amount to the full value of u/π, this would certainly be an extraordinary case. It would mean that all the various possible deviations we have considered in the course of the demonstrations should happen to be present at the same time and accumulate their effect. In most practical cases the error would probably amount to only a small fraction of u/π.

The foregoing considerations were only concerned with the magnitude of u/π, which determines the degree of accuracy. Now let us see what would be the condition for the fulfillment of the one to one condition (28a).

When a_n and b_n are given, then (28a) will hold good if

(44) $$R < \pi (\cos u/u-A)/\sqrt{p}+A$$

If the numbers A_n are decreasing, we evidently have $A = A_1$. According to our assumption concerning the numbers A and a we might therefore put $A = a_1 = 1/7$ and $1/5$ respectively in the two first and the last of the three cases above. This gives

$$R < 20$$

$$R < 8,7$$

$$R < 2,8$$

respectively.

Now the assumption that R shall not exceed but might be equal to 20 has the significance that the differential rate of variation of the gravitation of any trend may be so great that if the variation kept on at this rate through two consecutive zero distances of the trend considered, the ratio between the greatest and the least of the two zero distances would be a number written with five digits (formula (14d)). It would evidently be quite meaningless to assume that the differential rate of variation in any gravitation should be greater than this.

In the second and third case the ratio between two consecutive zero distances would be equal to 76 and 4 respectively. This, too, is largely sufficient to sustain the assumption that (28a) is fulfilled.

This analysis suggests the conclusion that the one to one conditions (28ab) are in ordinary cases far less restrictive than the condition that u/π shall be a small fraction. This means that if u/π is really a small fraction, so that the location of the normal points of y_0 may be determined fairly accurately, then in nearly all cases the correspondence between the zeros of y_0 and w'' will theoretically be a one to one correspondence, each zero of y_0 being announced once and only once by the vanishing of w''. Consequently if in an actual case w'' should apparently pass zero in a cluster of points lying within an interval that is short in relation to the zero distance in the lowest order trend (as in the illustration given in connection with the computation scheme for the normal points), then this should be interpreted as due to accidental errors.

Finally let us consider one of the conditions for non-overlapping, for instance the condition (26a) and let us see how the criterion of accuracy may be formulated by introducing the zero distances in w'' (which are known) instead of the zero distances in y_0. Suppose that $u/\pi = 0{,}036$ as in the first illustration above, then we have $\pi/u-2 = 26$ hence $1/(\pi/u-2) = 0{,}039$. This means (proposition (C)) that if we encircle each of the zeros T_K of the second differential coefficient w" of the known composite curve by an interval J_K, extending to the right of T_K a distance equal to 3,9% of the distance from T_K to the next zero of w" to the left, and to the left of T_K a distance equal to 3,9% of the distance from T_K to the next zero of w" to the right, then each of these intervals will contain at least one normal point for y_o, and y_o has no normal point outside these intervals. Further if one of these intervals, for instance J_K, does not overlap with the analagous intervals J_{K-1} and J_{K+1} encircling the next following zero of w" to the left and right respectively, then J_K contains exactly one normal point for y_o. And y_o has no normal point neither between J_K and J_{K-1} nor between J_K and J_{K+1}.

Now from the condition (26a) is seen that overlapping between the intervals considered can only take place if, of three consecutive zero distances in w", the center one is less than 3.9% of the sum of the two others. Such a possibility cannot be seriously considered. It would mean that of three consecutive zero distances in the lowest order trend y_o, the center one amounted to only some few per cents of the sum of the two others.

In the two other cases considered above, viz., $u/\pi = 0,08$ and $u/\pi = 0,167$, the intervals encircling the zeros of w" have to be made larger and the accuracy would be accordingly diminished, however, not to such an extent as to make the case of overlapping intervals very probable.

The case is quite different if two or more of the trends present are approximately of the same trend order. Suppose for instance that besides the ordinary business cycle with a duration of 6 to 10 years, there is present a minor cycle with an average duration of 40 months. Let y_0 be the minor cycle and y_1 the ordinary cycle. Then the magnitude of a_1 would probably lie somewhere between 1/2 and 1/3. And if the comparatively great value of a_1 is not balanced by a very small value of b_1, then the product $a_1 b_1$ and hence $\sum a_n b_n$ would not be a small fraction. This means that if the small difference between the average magnitudes of the periods in y_0 and y_1 is not balanced by the fact that the curve y_1 is very flat as compared with the curve y_0, i.e. has oscillations much less distinct than the oscillations of y_0, then y_0 and y_1 will be of approximately the same trend order. The method of normal points will therefore at best only permit a very rough estimation of the location of the zeros of y_0. And if the product $a_1 b_1$ is close to unity or even greater than unity, the method of normal points will not work at all. To trace the components y_0 and y_1 in this case it will be necessary to have recourse to the method of moving differences developed in Section 6.

We can make an interesting application of the general theorem by assuming that w is rigorously a sum of sine functions.

$$w = \sum y_n = \sum_{n=0}^{N} C_n \sin \sqrt{c_n}(t-t_n)$$

In this special case we should have $R = 0$, $p = 1$ and

$$d_n(t) = D_{nk} = \pi / \sqrt{c_n}$$

$$H_{nk} = C_n$$

Hence

$$a_n(t) = \sqrt{c_n/c_o}$$

$$b_n(t) = (C_n/C_o)\sqrt{c_n/c_o}$$

and consequently $a_n b_n = (C_n c_n)/(C_o c_o)$

If we introduce the half-periods (i.e. the zero distances) $D_n = \pi / \sqrt{c_n}$, we get

$$a_n b_n = (C_n/D_n^2)/(C_o/D_o^2)$$

Therefore

$$\sin u = (\sum_{n=1}^{N} C_n/D_n^2)/(C_o/D_o^2)$$

This means that the accuracy with which the zeros of y_o may now be determined by the location of the zeros of w", will be close if the ratio between the maximum amplitude and the second power of the period in y_o is predominant as compared with the same ratio in the various y_n.

When w is rigorously a sum of sine functions, we have $u = U$.

Further in this case the fulfillment of the one to one condition of proposition (B) is an immediate consequence of the non-overlapping condition. In fact, the non-overlapping condition involved in (B), namely (25a), may now be written $u < \pi/4$. And the one to one condition (28a) may always be replaced by the slightly less assuming (37) which in the case here considered reduces to $A < \cot u$. The one to one condition is therefore a fortiori fulfilled if only $A \leqslant 1$. And this does not imply any assumption as to the relative difference in order. Whatever be the periods of the components y_n, the numeration of the trends may always be chosen so as to have the period in y_o smaller than the other periods.

It would be interesting as an illustration to choose w as a sum of sine functions, compute a series of numerical values of w and then by the method of normal points work backwards and plot the various components. The comparison between these plots and the values of the various components calculated directly would give an illustration of how the method works.

6. THE METHOD OF MOVING DIFFERENCES

The problem of this section is to investigate under which conditions the ordinates and the gravitations (or which amounts to the same, the fictive zero distances) of the various components of the composite curve $w = \Sigma y_n$ may be determined approximately in a given point from the knowledge of the magnitude of the successive differential coefficients of w in this point, without introducing interpolation formulae of the type used in the method of normal points for determining the ordinates of the trends. Further to point out how the computation may be carried out if the conditions referred to are satisfied, and finally to establish a criterion which permits the classification of the trends present in a given series.

The conditions for the applicability of the methods of the present section will be more restrictive than those of the preceding sections with regard to the differential variations of the gravitation F_n of a particular trend y_n. In return the conditions regarding the difference in order between successive trends will be less restrictive. In fact in its most general version, the method of moving differences developed in the present section makes it theoretically possible to decompose the given series even in the case where two or more of the components present have periods whose magnitudes do not differ widely.

This is an important feature which distinguishes the method of moving differences from the method of normal points. The method of normal points is at a disadvantage as compared with the Fourier analysis when periods of the same order of magnitude are present. But this is not the case with the general method of the present section. The method of moving differences may therefore perhaps prove to be useful also in the analysis of certain kinds of physical phenomena which are now usually studied by harmonic analysis.

The practical applicability of the method will of course depend on the condition that the data must not be afflicted with accidental errors to such an extent as to completely obscure the finer traits of the curve, which are represented by the magnitude of the successive differential coefficients.

I shall first make an intuitive approach to some special and very simple cases.

Let us revert to the case where the composite curve w is made up of a straight line and a sine curve (formula (1)). In this case the ordinate of the lowest order trend is simply proportional to the second differential coefficient w and the <u>factor of proportionality is the reciprocal of c which is the gravitation of the lowest order trend</u>, the gravitation here being constant. The magnitude of the factor of proportionality is therefore given by the distance between two consecutive zeros of the lowest order trend (here y). And this distance is in turn equal to the distance between two consecutive zeros of w" which are known. Let D be this distance. Then we have

(45) $$y = -(D/\pi)^2 w''$$

Therefore, to plot the ordinates of y in this case, we only have to plot w" with the sign reversed and multiply by the known constant $(D/\pi)^2$.

We may also express y in proportion to the second difference of the composite curve

$$\Delta^2 w_{t-1} = w_{t+1} - 2w_t + w_{t-1}$$

Since

$$\sin(x+h) - 2\sin x + \sin(x-h) = -2(1-\cos h)\sin x$$
$$= -(2\sin(h/2))^2 \sin x$$

we have

(46) $$y_t = -(2\sin(\pi/2D))^{-2}\Delta^2 w_{t-1}$$

And this formula is exact (not approximate) when w is rigorously of the form (1). In the case (1) the ordinates of y may therefore be determined rigorously either from the plot of w" or from the plot of $\Delta^2 w_{t-1}$ by multiplication with a constant which can be computed when w is known.

This suggests the idea that when we have a composite curve $w = \Sigma y_n$ where the gravitations F_n are of descending order of magnitude in the sense explained in Section 2, we should have an approximation to the ordinate of the lowest order trend y_0 in a given point if we multiply -w" by the reciprocal of the lowest order gravitation F_0 in this point.

The conditions under which this really will give an approximation to y_0 shall be investigated later. At present I shall follow up the idea suggested.

The magnitude of the gravitation F_0 in a given point is not known, but may itself be approximated in one of the two following ways. In the first we may replace F_0 by its average value, the average being taken over the zero distance $i_{0j} = (t_{0j}, t_{0,j+1})$ in y_0 in which the point considered is situated. Now by formula (7) the average of F_0 taken over the zero distance i_{0j} is equal to $(\pi/D_{0j})^2$. For any point in i_{0j}

we should consequently have an approximation to y_o by putting

(47) $$y_o = -(D_{oj}/\pi)^2 w''$$

If the normal points of y_o are determined by the method of the preceding section, the ordinates of y_o would consequently be determined from the plot of w" by multiplication with a quantity which is constant between two consecutive normal points but might change from one zero distance to the next following. This would not entail a discontinuous variation in y_o because the points in which the factor of proportionality is changed, are just the points where y_o is equal to zero.

The next step would be to consider a factor of proportionality which not only moved by steps from one zero distance to the next, but moved continuously, which in practice would mean moved from one point of observation to the next.

Let us again revert to the case (1) where w is made up of a straight line and a sine curve. If we differentiate w four times we get

$$w'' = -cy$$
$$w^{(4)} = c^2 y$$

Hence

$$c = -w^{(4)}/w''$$
$$y = w''^2/w^{(4)}$$

This suggests the idea that if in the general case $w = \sum y_n$ the difference in order between the successive trends is great, we should have an approximation to the lowest order gravitation F_o in the vicinity of a point by putting it equal to

(48) $$F_o = -w^{(4)}/w''$$

Consequently we should have an approximation to y_o by putting it equal to

(49) $$y_o = w''^2/w^{(4)}$$

It should be pointed out that the kind of approximation here considered is quite different from the kind of approximation used in curve fitting with constant parameters. We do not consider the lowest order trend as capable of being represented by a sine function throughout an interval of definite length. This would involve the assumption that the gravitation is constant throughout the interval considered. But this is just the assumption which we now have dropped. In fact by (48) the gravitation is determined as an essentially moving feature of the lowest order trend.

The kind of approximation here used may rather be characterized as a method of instantaneous approximation, valid only in the vicinity of a point. Or it may be characterized as a method of curve fitting with moving parameters, the value of the parameters being continuously changed and determined in each point by the differential properties of the given series in this point.

The difference between the two points of view may perhaps be rendered more precise by the following general considerations.

Suppose we have an analytical expression

$$V = f(t; c_1 c_2 \dots c_n)$$

which is a function of the variable t and contains the parameters $c_1 c_2 \dots c_n$. If the function V is to be applied to a given series w (or a given analytical function) for the purpose of smoothing or interpolation, the values of the parameters (supposed constant) have to be determined by one of the known methods, for instance the method of moments. The procedure would be to express the parameters in terms of the theoretical moments of V and then introduce in these expressions the empirical moments of the given series w for the theoretical ones, the empirical moments being computed with or without corrections as the case may be, (or introduce the moments of the given analytical function for the moments of V).

On the contrary, the procedure of a general method of instantaneous approximation would be the following.

First the parameters of V have to be expressed in terms of the theoretical differential coefficients of V (of order 0,1...), these expressions being such as to hold good identically in t. Generally (n-1) differentiations would give the necessary number of equations. The differentiations have to be performed as if the parameters were independent t. If in the expressions obtained, the values of the empirical differential coefficients of the given series w for a certain point t (or the values of the differential coefficients of some given analytical function) are introduced instead of the theoretical differential coefficients of V, then a set of values of the parameters would be determined. And the function V thus defined (with constant parameters) would be such as to have contact of high order with the given series (or function) w in the point t considered. The same procedure may be applied in any point t. Thus a function $\bar{V}$ is determined, which is of the form V but with parameters that are functions of t.

Now any series (or function) w may evidently be expressed in the form $\bar{V}$ with variable parameters (generally in an infinity of ways if $n > 2$). In any point the value taken by w will evidently be rigorously equal to the value taken by $\bar{V}$. Further under a certain condition, the values of the parameters of w in a given point will be approximately equal to the values of the parameters of $\bar{V}$ in this point. The condition is evidently that when w is expressed in the same form as $\bar{V}$, then the equations obtained by differentiating w a certain number of times as if the parameters were constant, shall hold good approximately in the given point t. The accuracy of the approximation in a given point t depends therefor only on the differential properties of w and its parameters in this point, and is independent of how the approximation works out in other parts of the t-range. The procedure of determining the parameters of w in any point by putting them equal to the parameters of $\bar{V}$ in the same point, may be called the method of instantaneous approximation.

The difference between the assumptions underlying the ordinary method of curve fitting and the method of instantaneous approximation, is now obvious. In the first case it is assumed that the given series (or function) w can be represented over the whole t-range by the function V with rigorously constant parameters. In the second case the assumption is only that when the given series (or function) w is expressed in the form V (which is always possible when the parameters are considered as functions of t), then the differential variations in the parameters shall be of a certain kind.

This leads up to the idea of a method by which the curve $w = \Sigma\, y_n$ may be decomposed even in the case where two or more of the components present have periods whose magnitudes do not differ widely, i.e. the case where the difference in order between two consecutive trends is not great.

In fact if we suppose that the given composite curve $w = \Sigma y_n$ in the vicinity of a point can be represented differentially (i.e. not only with respect to its ordinate, but also with respect to a certain number of its differential coefficients) by a sum of sine functions and a linear function

(50) $$V = \sum_{n=0}^{N-1} S_n + at + b$$

where

$$S_n = C_n \sin\sqrt{c_n}(t-t_n)$$

C_n, c_n, t_n, a and b being parameters, then the method indicated above would permit the moving determination of these parameters. In particular we should have the possibility of the moving determination of the gravitations c_n, i.e. of F_n, and the moving determination of the ordinates S_n, i.e. of y_n. And this determination would not be subject to the condition that the difference in order between successive trends shall be great.

The plan of the following analysis is first to investigate the general properties of an expression of the form (50) and to show how the gravitations c_n (now supposed rigorously constant) and the ordinates S_n themselves can be expressed in terms of the successive differential coefficients of even order $V^{(2h)}$. It will be seen that this problem has a certain resemblance to the Stieltjes problem of moments for the case of discrete distributions.

When these questions are settled, I proceed to show under which conditions the curve $w = \sum y_n$ (where the gravitations are not rigorously constant) may be represented differentially by an expression of the form (50).

We introduce the following notations

(51) $$v_h = V^{(2h)} = \frac{d^{2h}V}{dt^{2h}} = \sum_{n=0}^{N-1} (-c_n)^h S_n \qquad h = 1,2\ldots\infty$$

(52) $$\Delta_n = \begin{vmatrix} v_1 & v_2 & \cdots & v_n \\ v_2 & v_3 & \cdots & v_{n+1} \\ \cdots & \cdots & \cdots & \cdots \\ v_n & v_{n+1} & \cdots & v_{2n-1} \end{vmatrix} \qquad n = 1,2\ldots\infty$$

Instead of Δ_n we write Δ_{nN} when it is necessary to emphasize the number N of sine terms in V.

Further we designate by d_{hn} ($h=0,1\ldots n$) the determinantes obtained by letting out the (h+1)th column in the matrix

(53a) $$\begin{pmatrix} v_1 & v_2 & \cdots & v_{n+1} \\ v_2 & v_3 & \cdots & v_{n+2} \\ \cdots & \cdots & \cdots & \cdots \\ v_n & v_{n+1} & \cdots & v_{2n} \end{pmatrix} \qquad n = 1,2\ldots\infty$$

Evidently

(53b) $$d_{nn} = \Delta_n$$

If $n = N$ we write in brief

(53c) d_h instead of d_{hN}

The h-th order elementary symmetric function of the numbers $c_0 c_1 \ldots c_{n-1}$, i.e., the sum of all the products of h factors which can be formed by picking out in all possible ways h of the n numbers $c_0 c_1 \ldots c_{n-1}$, will be designated

(54a) $$a_{hn} = (c_o c_1 \dots c_{n-1})_h \qquad h = o,1\dots n$$

$$a_{oo} = a_{on} = 1 \qquad n = o,1\dots N$$

If $n = N$ we write in brief

(54b) a_h instead of a_{hN}

Further let

(55a) $$R_{hn} = \sum_{i=0}^{N-1} (c_o - c_i)(c_1 - c_i)\dots(c_{n-1} - c_i)(-c_i)^h S_i \qquad h = 1,2\dots\infty$$

$$n = 0,1,2\dots N$$

$$R_{ho} = v_h = \sum_{i=0}^{N-1} (-c_i)^h S_i$$

The effective number of terms in R_{hn} is evidently $(N-n)$, R_{hn} containing only $S_n S_{n-1} \dots S_{N-1}$. In particular we have

(55b) $$R_{hN} = 0 \qquad h = 1,2\dots\infty$$

(55c) $$R_{h,N-1} = (c_o - c_{N-1})(c_1 - c_{N-1})\dots(c_{N-2} - c_{N-1})(-c_{N-1})^h S_{N-1} \qquad h=1,2\dots\infty$$

Finally let $p,q\dots s$ be n different numbers from the suite $0,1\dots(N-1)$, $(1 \leqslant n \lessgtr N)$; $p,q\dots s$ are supposed to be arranged in an ascending order of magnitude, i.e. $p < q \dots < s$. Then $(p,q\dots s)$ will be used to designate

(56a) $$(p,q\dots s) = c_p c_q \dots c_s \begin{vmatrix} 1 & c_p & c_p^2 \dots c_p^{n-1} \\ 1 & c_q & c_q \dots c_q^{n-1} \\ \dots & \dots & \dots \\ 1 & c_s & c_s^2 \dots c_s^{n-1} \end{vmatrix}^2$$

Developing the Vandermonde determinant to the right in (56a) we get

(56b) $$(p,q\dots s) = c_p c_q \dots c_s \left(\prod_{\kappa < \beta} (c_\kappa - c_\beta) \right)^2$$

where κ runs through the numbers $p,q\dots s$ with exception of the last one, and β through those of the numbers $p,q,\dots s$ which are greater than κ. Hence $(p,q\dots s)$ can only vanish if one of the c is zero or if two of the c are equal. If all the c are positive (not zero) and different, then (p,q...s) is essentially positive (not zero). Evidently it does not restrict generality if this assumption regarding the numbers c is made.

I shall first prove that if N is the number of sine terms in (50), all the quantities c being different and different from zero, then the determinants $\Delta_1 \Delta_2 \ldots \Delta_N$ can only vanish in discrete points, i.e., they cannot vanish identically in any interval of finite length however small. On the other hand, $\Delta_{N+1} \Delta_{N+2} \ldots$ and the higher determinants will all vanish identically in t. This evidently furnishes a necessary and sufficient criterion for the number of sine terms in V.

To establish the criterion we shall use the equation

(57) $$a_{nn}v_h + a_{n-1,n}v_{h+1} + \ldots + a_{1n}v_{h+n-1} + a_{on}v_{h+n} = R_{hn}$$
$$h = 1,2\ldots\infty$$
$$n = 0,1\ldots N$$

This equation evidently holds good for n = 0 and $h = 1,2\ldots\infty$, because in the case n = 0, it reduces to $v_h = R_{ho}$ (h 1).

Now suppose that it holds good for n and h. Successive derivations with respect to t show that it holds good for any higher value of h. On the other hand, adding the equation for h multiplied by c_n to the equation for h+1, and noticing that

$$a_{h+1,n+1} = a_{h+1,n} + c_n a_{hn}$$
$$R_{h,n+1} = R_{h+1,n} + c_n R_{hn}$$

we get the equation (57) for n+1 and h. Hence (57) holds good in general.

As $R_{hN} = 0$ by (55b), we get in particular

(58) $$a_N v_h + a_{N-1} v_{h+1} + \ldots + a_1 v_{h+N-1} + v_{h+N} = 0$$
$$h = 1,2\ldots\infty$$

Now if $n \leq N+1$, add to the last line of the determinant (52) the first line multiplied by $a_{n-1,n-1}$, the second line multiplied by $a_{n-2,n-1}$... finally the (n-1)th line multiplied by $a_{1,n-1}$. The quantities in the last line will then by (57) be

$R_{1,n-1} R_{2,n-1} \ldots R_{n,n-1}$.

Hence we have the formula

$$\Delta_n = \begin{vmatrix} v_1 & v_2 & \ldots v_n \\ v_2 & v_3 & \ldots v_{n+1} \\ \ldots & \ldots & \ldots \\ v_{n-1} & v_n & \ldots v_{2n-2} \\ v_n & v_{n+1} & \ldots v_{2n-1} \end{vmatrix} = \begin{vmatrix} v_1 & v_2 & \ldots & v_n \\ v_2 & v_3 & \ldots & v_{n+1} \\ \ldots & \ldots & \ldots & \ldots \\ v_{n-1} & v_n & \ldots & v_{2n-2} \\ R_{1,n-1} & R_{2,n-1} & \ldots & R_{n,n-1} \end{vmatrix} \qquad (59)$$

$$n = 1,2 \ldots N+1$$

From this and (55b) we immediately deduce that if N is the number of sine terms in (50), then $\Delta_{N+1} = 0$ identically in t.

If $n > N+1$, add to the last line of the determinant (52) the (n-N)th line multiplied by a_N, the (n-(N-1))th line multiplied by a_{N-1}... finally the second last line multiplied by a_1. By virtue of (58) each element in the last line will be zero, hence

$\Delta_{N+1} = \Delta_{N+2} = \ldots = 0$ identically in t.

It is seen that if R_{hn} $(h=1,2\ldots\infty)$ is defined as being equal to zero for $n > N$, the formula (59) will hold good for any value of $n=1,2\ldots\infty$.

Further we get by developing Δ_{nN}

$$\Delta_{22} = (01)S_0S_1$$
$$\Delta_{23} = (01)S_0S_1 + (02)S_0S_2 + (12)S_1S_2$$

$$-\Delta_{33} = (012)S_0S_1S_2$$

and generally

$$\Delta_n = \Delta_{nN} = (-)^n \sum (p,q\ldots s)S_pS_q\ldots S_s \qquad (60a)$$

the sommation being extended to combinations without repetition of the n numbers p,q...s chosen amongst the N numbers 0,1...(N-1), i.e.

$$p = 0,1\ldots(N-n)$$
$$q = p+1,p+2\ldots(N-n+1)$$
$$\ldots\ldots$$
$$s = \ldots(N-1)$$

From (60a) we get in particular

(60b) $\Delta_N = \Delta_{NN} = (-)^N(0,1\ldots(N-1))S_0S_1\ldots S_{N-1}$

As (0,1...N-1) is different from zero, we immediately deduce that the highest of the non-identically vanishing determinants, namely Δ_N can only vanish in points in which at least one of the sine terms vanishes. To show that the lower determinants cannot vanish identically, we proceed in the following way.

We first notice that a linear combination of non-identically vanishing sine terms $\sum_{n=0}^{N-1} S_n$ where the periods are different, i.e. all the c_n are different, can never vanish identically in any finite interval however small.

For suppose that $\sum_{n=0}^{N-1} S_n = 0$ identically in t in a certain interval. By 2N-2 differentiations we should have in any point in the interval considered

$$\sum_{n=0}^{N-1}(-c_n)^h S_n = 0 \qquad h = 0,1\ldots(N-1)$$

This is a homogeneous system of N equations linear in the N quantities S_n. The determinant of the system is a Vandermonde determinant which cannot vanish when all the c are different. Hence in any point in the interval considered each of the S_n would be equal to zero.

The same evidently holds good of a linear combination of cosine terms.

Further by successive application of the formulae

$$2 \sin x.\sin y = \cos (x-y) - \cos (x+y)$$

$$2 \sin x.\cos y = \sin (x+y) + \sin (x-y)$$

we see that $\sin x_1.\sin x_2 \ldots \sin x_n$ can be developed as a linear combination of sine terms (if n is odd) or a cosine term (if n is even). The expansion contains amongst others the term $\sin(x_1+x_2+\ldots +x_n)$ or $\cos(x_1+x_2+\ldots +x_n)$ with a non-vanishing coefficient. Further the expansion will contain a certain number of terms of the form sine or cosine to the angle $(\pm x_1 \pm x_2 \ldots \pm x_n)$ where the signs are combined in different ways, however not so that all the signs are minus.

Hence the product of n factors $S_p S_q \ldots S_s$ in (60a) may be developed as a linear combination of sine terms or cosine terms. One of the terms being sine or cosine to the angle

(61a) $$(\sqrt{c_p} + \sqrt{c_q} + \ldots + \sqrt{c_s})t - (\sqrt{c_p}.t_p + \sqrt{c_q}t_q + \ldots + \sqrt{c_s}t_s)$$

and the coefficient of this term is non-vanishing.

Now without restricting the generality we may assume the periods in the various S_n to be positive, i.e. all the quantities $\sqrt{c_n}$ may be taken positive.

Hence the pre-factor of t in (61a) is not equal to, but greater than the pre-factor of t in any other term which occurs in the expansion of $S_p S_q \ldots S_s$. Further the numbering of the quantities $c_o c_1 \ldots c_{N-1}$ may be chosen so that the quantities c_n are arranged in a descending order of magnitude, $c_o > c_1 \ldots > c_{N-1}$. In the expansion of (60a) there must therefore be a sine or cosine term (with non-vanishing coefficient) where the pre-factor of t

(61b) $$\sqrt{c_o} + \sqrt{c_1} + \ldots + \sqrt{c_{n-1}}$$

is not equal to, but greater than the pre-factor of t in any other term, and further different from zero, because all the $\sqrt{c_n}$ are positive (not zero).

If in the rest of the expansion there should occur terms where the pre-factor of t is zero, these terms would be constants (eventually zero). If there should occur terms with equal pre-factors of t, i.e. with equal periods, these terms might evidently be expressed as one single sine or cosine term with the same period. The coefficient of such a term may eventually vanish.

When the rest of the expansion is ordered in this way, the rest is either constant (eventually zero) or it is the sum of a constant and a linear combination of non-identically vanishing sine or cosine terms, where the periods are all different, different from zero and different from the period of the first term, where the pre-factor of t is (61b) and where the coefficient is certainly non-vanishing. In the first case Δ_n is the sum of a constant and a simple non-identically vanishing sine or cosine function. In the second case it is the sum of a constant and a linear combination of such functions with different periods. In neither case can Δ_n vanish identically.

The criterion for the number of sine terms in V is therefore established.

Corollary regarding the moments of a frequency distribution.

From the formula (60a) may be derived a direct and simple proof of the Watanabe criterion (1) for the order of a discrete frequency distribution.

(1) Watanabe, Tôhoku Mathem. Journal 15(1919). See also the present writers essay "Sur les semi-invariants et moments employés dans l'étude des distributions statistiques." Skrifter utg. av det Norske Videnskapsakademi. Oslo II 1926. No. 3, p. 7.

Let

$$x_0 x_1 \dots x_{N-1}$$
$$P_0 P_1 \dots P_{N-1}$$

be a frequency distribution (empirical or a priori), $x_0 x_1 \dots x_{N-1}$ designating the possible values of the attribute and $P_0 P_1 \dots P_{N-1}$ the frequencies (absolute or relative). All the P are supposed positive (not zero), all the x are supposed different. N is called the order of the distribution.

Let

$$m_h = \sum_{n=0}^{N-1} x_n^h P_n$$

be the moments of the distribution.

Further let

$$\Delta_n = \begin{vmatrix} m_0 & m_1 \dots m_{n-1} \\ m_1 & m_2 \dots m_n \\ \dots & \dots \\ m_{n-1} & m_n \dots m_{2n-2} \end{vmatrix}$$

Then Watanabe's necessary and sufficient criterion for the order of the distribution is that if N is the order of the distribution, Δ_{N+1} and all the higher order determinants are equal to zero, while all the lower order determinants are different from zero.

If we put

$$-c_n S_n = P_n \qquad \text{and } -c_n = x_n$$

we get by (51)

$$m_h = v_{h+1}$$

Introducing in (55a), (59) and (60a) we have

$$R_{hn} = \sum_{i=o}^{N-1} (x_i - x_o)(x_i - x_1)\dots(x_i - x_{n-1}) x_i^{h-1} P_i$$

(62a) $$\Delta_n = \begin{vmatrix} m_o & m_1 & \dots & m_{n-1} \\ m_1 & m_2 & \dots & m_n \\ \dots & \dots & \dots & \dots \\ m_{n-2} & m_{n-1} & \dots & m_{2n-3} \\ m_{n-1} & m_n & \dots & m_{2n-2} \end{vmatrix} = \begin{vmatrix} m_o & m_1 & \dots & m_{n-1} \\ m_1 & m_2 & \dots & m_n \\ \dots & \dots & \dots & \dots \\ m_{n-2} & m_{n-1} & \dots & m_{2n-3} \\ R_{1,n-1} & R_{2,n-1} & \dots & R_{n,n-1} \end{vmatrix}$$

and

(62b) $$\Delta_n = \sum (p,q\dots s)\ P_p P_q \dots P_s$$

where the summation is extended to combinations without repetition of the n numbers p,q...s chosen amongst the N numbers 0,1,...(N-1), and where now (p,q...s) designates

$$(p,q\dots s) = \begin{vmatrix} 1 & x_p & x_p^2 & \dots & x_p^{n-1} \\ 1 & x_q & x_q^2 & \dots & x_q^{n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_s & x_s^2 & \dots & x_s^{n-1} \end{vmatrix}^2 = \left(\prod_{\alpha<\beta} (x_\alpha - x_\beta) \right)^2$$

From (62a) is seen that $\Delta_{N+1} = \Delta_{N+2} = \dots = 0$, because $R_{hn} = 0$ for $n \geq N+1$.

And (62b) shows that the lower determinants are different from zero. For (p,q,...s) is essentially positive when the various x_n are all different (positive, negative or zero). And the frequencies P_n are all positive.

*

I now proceed to the determination of the periods in the sine terms of (50). This amounts to determining the numbers c_n.

By virtue of the definition of the quantities $a_h = a_{hN}$

(formulae (54ab)), the numbers c_n are the N roots of the equation

$$c^N - a_1 c^{N-1} + a_2 c^{N-2} \dots + (-)^N a_N = 0$$

Now, from (58) we get

(63) $$a_N v_h + a_{N-1} v_{h+1} + \dots + a_1 v_{h+N-1} = -v_{h+N}$$

$$h = 1,2 \dots N$$

This is a system of equations which are linear in the quantities a. The system cannot be homogeneous except in a point where all the sine terms vanish separately. For if $v_{h+N} = 0$ for $h = 1,2\dots N$, we should have

$$\sum_{n=0}^{N-1} (-c_n)^{h+N} S_n = 0 \qquad h=1,2\dots N$$

which is a homogeneous linear system in the S_n with non-vanishing determinant.

The determinant of the linear system (63) is $\Delta_N = d_{NN} = d_N$, which by (60b) can only vanish in points in which at least one of the sine terms vanishes. With exception of these points we therefore have

(64) $$a_h = (-)^h d_{N-h}/d_N$$

where the d_h are defined by (53ac).

Consequently with exception of the points where one or more of the sine terms vanishes, the numbers $c_o c_1 \dots c_{N-1}$ are determined as the N roots of the equation

(65) $$d_o + d_1 c + d_2 c^2 + \dots + d_N c^N = 0$$

This equation will be called the period equation.

In the points where one or more of the sine terms vanishes, all the coefficients of the period equations, i.e., all the determinants $d_o d_1 \dots d_N$ vanish simultaneously. In any other point they are all different from zero. This means that if one of the coefficients of the period equation vanishes in a given point, then all the other coefficients must vanish too in this point. Further in points where the coefficients do not vanish simultaneously they have alternating signs, i.e., d_0 and d_1 have opposite signs, d_1 and d_2 have opposite signs and so forth.

In fact we deduce from (63) that in any point

$$d_N a_h = (-)^h d_{N-h}$$

hence

$$d_h = (-)^{N-h} a_{N-h} d_N \tag{66}$$

Here a_{N-h} is essentially positive (not zero) because the periods of the sine terms in (50) are supposed real and consequently all the c_n positive. And $d_N = \Delta_N$ vanishes only in points where one or more of the sine terms vanishes.

That the determinants d_h must have alternating signs in points where they do not vanish, can also be seen by applying the Descartes rule of signs to the equation (65).

I shall now show how the ordinates S_n in (50) can be determined.

In analogy with (54a) we define $a_h^{(n)}$ as the h-th order elementary symmetric function of the numbers $c_0 c_1 \dots$ ((except c_n)) $\dots c_{N-1}$

$$a_h^{(n)} = (c_0 c_1 \dots \text{((except } c_n\text{))} \dots c_{N-1})_h \tag{67}$$

From (57) we get for $n = N-1$

$$R_{h,N-1} = \sum_{i=0}^{N-1} a_i^{(N-1)} v_{N+h-i-1}$$

But $R_{h,N-1}$ contains only one single sine term, namely, S_{N-1}. We therefore get by (55c)

$$(c_0 - c_{N-1})(c_1 - c_{N-1}) \dots (c_{N-2} - c_{N-1})(-c_{N-1})^h S_{N-1} = \sum_{i=0}^{N-1} a_i^{(N-1)} v_{N+h-i-1}$$

Now the successive elimination of the sine terms which is performed by the equation (57) for $n = 0,1\ldots N-1$, may evidently be so arranged that any of the S_n is the last term to remain in $R_{h,N-1}$. Further h may be chosen as any of the numbers 1,2... If we put $h = 1$, we must therefore have

$$(68a)\quad (c_0-c_n)(c_1-c_n)\ldots((c_n-c_n))\ldots(c_{N-1}-c_n)(-c_n)S_n = \sum_{h=0}^{N-1} a_h^{(n)} v_{N-h}$$

$$n = 0,1\ldots N-1$$

here $((c_n-c_n))$ indicates that this factor is excluded in the product on the left hand side of the equation.

The pre-factor of S_n in (68a) is different from zero because all the c (the roots of the period equation) are different and different from zero. The formula (68a) therefore determines all the S_n unambiguously.

We may also obtain another expression for S_n. In fact if we put $n = N$ and use the last determinant of (59), we see that we get an expression for S_{N-1}. And analagous expressions hold good if the successive elimination of the sine terms is arranged so that any of the terms is the last one to remain in $R_{h,N-1}$. We therefore have quite generally

$$(68b)\quad (c_0-c_n)(c_1-c_n)\ldots((c_n-c_n))\ldots(c_{N-1}-c_n)(-c_n)S_n = \Delta_N/\Delta_N^{(n)}$$

where

$$\Delta_N^{(n)} = \begin{vmatrix} v_1 & v_2 & \ldots & v_N \\ v_2 & v_3 & \ldots & v_{N+1} \\ \ldots & \ldots & \ldots & \ldots \\ v_{N-1} & v_N & \ldots & v_{2N-2} \\ 1 & (-c_n) & \ldots & (-c_n)^{N-1} \end{vmatrix}$$

In the foregoing analysis we have only used the quantities v_h for $h \geqq 1$. That is, we have only considered the second order and higher differential coefficients of the composite curve V. The formulae obtained are therefore independent of the "secular trend" at+b contained in (50). The ordinates of this "secular trend" may be determined as the residuum left in V when the ordinates of the various S_n are subtracted.

Further we have made no assumptions as to the relative difference in order between the various S_n. The various c_n may therefore be of the same order of magnitude. If this is the case, the various S_n may be considered as a trend group, the various components in the group being of the same trend order, and the group as such being of a low trend order as compared with the "secular trend", i.e. the straight line at+b, which can be looked upon as a solution of $y''+cy = 0$ with $c = 0$.

I shall now consider the case in which the smallest of the c_n, i.e., c_{N-1}, is very small as compared with the other c_n, and the case in which the greatest of the c_n, i.e., c_o, is very great as compared with the other c_n. This analysis will throw a light on the notion of trend group.

If c_{N-1} tends towards zero, one of the roots of (65) must tend toward zero. As all the determinants d are finite, d_o must also tend towards zero for any value of t.

Now if in an equation of the N-th degree in c the coefficients are considered as independent, and the term independent of c is put equal to zero, then one root of the equation is $c = 0$ and the remaining (N-1) roots may be determined as the roots of the equation obtained by dividing the original equation by c. In our case, however, this procedure is not possible. For if d_o vanishes identically on account of one of the c_n becoming zero, then $d_N = \Delta_N$ must vanish identically by virtue of (60b). Consequently all the other coefficients d_h must vanish identically, as is seen by (66). Hence the period equation will be of no use whatsoever in determining the periods.

If the various S_n are considered as trends in a composite curve, the only possible interpretation of the fact referred to is that S_{N-1} no longer belongs to the same trend group as the other sine terms, but belongs to the group of "secular trends" represented by the straight line. This interpretation is also in conformity with the criterion that if $\Delta_N \equiv 0$, the number of sine terms cannot exceed (N-1).

If the periods of the remaining (N-1) sine terms $S_0 S_1 \ldots S_{N-2}$ are different and different from zero, the periods will evidently be determined by the equation

$$d_{o,N-1} + d_{1,N-1}c + \ldots + d_{N-1,N-1}c^{N-1} = 0$$

Now suppose that c_0 is tending towards ∞. As seen by (60b) the fluctuations in Δ_N will be more and more violent the greater c_0. And the moments of time in which Δ_N passes from positive to negative values will be more and more frequent. For Δ_N passes zero in all the points where S_0 passes zero. If not only c_0 but also some of the next following c_n are great, then the fluctuations in Δ_N will not only be violent and with frequent changes of sign, but also more or less irregular and confusing. The practical interpretation of this fact would be that the first components $S_0 S_1 \ldots$ belong to a group of lower order than the rest of the S_n. If the ordinates of the composite curve are only known in discrete points, the lowest order group must necessarily be considered as a group of accidental components, and it will not be possible to analyze the periods and ordinates of the rest of the S_n before the lowest order components are eliminated (as far as it can be done) by some sort of smoothing, for instance by applying a moving average to the given composite curve.

If this is done, the significance of the vanishing of the various determinants is changed. The point of real interest now is not if the determinants themselves vanish rigorously in all points t, but if the average value of any of the determinants considered, taken over any small interval is practically zero.

The vanishing of the various determinants, taken in this practical sense, will be a criterion for the number of terms which is left after the moving average elimination of the components that are considered as accidental.

In the case where the function V is smoothed before submitted to analysis, a correction must be introduced in the formulae (68ab). This correction shall now be considered.

If we take the sum of (2k+1) equidistant terms in the function sin t, the interval between consecutive terms being h, and the terms being multiplied by arbitrary weights λ_j $(j = -k, -(k-1) \ldots 0, 1 \ldots k)$, we get

$$\sum_{j=-k}^{k} \lambda_j \sin(t + jh) = \sum_{j=-k}^{k} \lambda_j (\sin t . \cos jh + \cos t . \sin jh)$$

$$= \sin t \left(\lambda_0 + \sum_{j=1}^{k} (\lambda_j + \lambda_{-j}) \cos jh \right)$$

$$+ \cos t \sum_{j=1}^{k} (\lambda_j - \lambda_{-j}) \sin jh$$

If the weights to the right and left of t are symmetrical, i.e. $\lambda_j=\lambda_{-j}$, the last term in the right hand side of the formula will vanish.

Therefore if we apply a central moving average with (2k+1) equidistant terms and symmetrical (but otherwise arbitrary) weights to the function considered, the item sin t will be replaced by ρ sin t , where ρ is a constant independent of t

$$\rho = (\lambda_0 + 2 \sum_{j=1}^{k} \lambda_j \cos jh)/(\lambda_0 + 2 \sum_{j=1}^{k}\lambda_j)$$

It is readily seen that ρ is always less than unity. Further ρ is certainly positive if $2kh < \pi$. This means that if the length of the interval covered by the moving average does not exceed the zero distance in the sine function, then the only effect of the moving average operation is that the oscillations of the function are damped, i.e. depressed, in a constant proportion, the ordinate being multiplied by a positive factor which is independent of t and less than unity. The factor ρ will be called the coefficient of damping.

Now suppose that a central moving average of the kind considered and with h=1 is applied to the function (50). This means that the item V(t) is replaced by the item

$$\dot{V}(t) = \frac{\lambda_0 V(t) + \lambda_1(V(t+1)+V(t-1)) + \dots + \lambda_k(V(t+k)+V(t-k))}{\lambda_0 + 2(\lambda_1 + \lambda_2 + \dots + \lambda_k)}$$

where the weights are positive but otherwise arbitrary.

If the length of the average viz. 2k is less than $\pi/\sqrt{c_0}$, i.e. less than the zero distance in the term S_0(which is the term with shortest zero distance), then the only effect of the moving average operation is that each of the sine terms S_n is multiplied by a coefficient of damping ρ_n (< 1) which is positive and independent of t

$$\rho_n = (\lambda_0 + 2\sum \lambda_j \cos j\sqrt{c_n})/(\lambda_0 + 2\sum \lambda_j)$$

where $\sum$ designates $\sum_{j=1}^{k}$.

The linear term at+b in V will of course not be affected by the moving average operation.

Hence the function $\dot{V}(t)$ will be a function of the same kind as the original function V(t). The periods in $\dot{V}$ will be the same as the periods in V. The only difference is that the constants $\dot{C}_n$ in $\dot{V}$ will have other values than the constants C_n in V. Therefore $\dot{V}$ may be decomposed and the ordinates $\dot{S}_n$ of its components determined by the method developed above. When this is done, we may revert to the ordinates S_n of the original function simply by using the formula

$$\rho_n S_n = \dot{S}_n$$

Consequently if a moving average of the kind considered is applied to V and now $v_1 v_2 \ldots$ designate the differential coefficients of the smoothed function $\dot{V}$, then the periods of the original function V are still determined by the period equation (65) and the ordinates S_n are determined by the equations obtained from (68a) or (68b) by introducing the factor ρ_n on the left hand side.

It is readily seen that the coefficient of damping ρ_n for the term S_n lies all the closer to unity the shorter the interval covered by the moving average is as compared with the zero distance in S_n. Therefore the damping effect is heaviest for the first term S_0. If the interval covered by the moving average is very short as compared with the zero distance in the first term S_0, then all the corrections ρ_n will be insignificant and may be left out, i.e. all the ρ_n may be put equal to unity.

The foregoing analysis was concerned with the case where the gravitations were rigorously independent of t. In this case the composite curve has the form (50) and the gravitations c_n and the ordinates S_n may be determined rigorously from the knowledge of the differential coefficients of the composite curve.

I now proceed to show under what conditions the curve $w = \sum y_n$ with not rigorously constant gravitations may be represented differentially by (50) in the vicinity of a point and hence may be analyzed by the method of instantaneous approximation, the approximation curve with moving parameters being (50).

The function w is capable of being represented differentially by (50) in the vicinity of a point, if w is of such a kind that k differentiations of w carried out as if the various F_n were constants, will give approximately the right value of $w^{(k)}$, k being not greater than the order of the highest differential coefficient which it is necessary to introduce in the formulae (65) and (68ab), i.e., $k = 4N + 2$.

As the method only involves even order differential coefficients of w, it is sufficient to consider $k = 2h$ $(h = 1,2...2N+1)$.

It is easily seen that when the differentiations of w are carried out as if the various gravitations were constants, then we get

(69) $$w^{(2h)} = \sum (-F_n)^h y_n$$
$$(h = 1,2\ldots 2N+1)$$

The problem is therefore to show under which conditions the actual value of $w^{(2h)}$ is approximately equal to the value given by (69).

Let us consider an isolated trend y_n. For the sake of brevity we drop the subscript n.

It is easily seen that if y is a solution of $y'' + Fy = 0$, then the n-th differential coefficient of y $(n = 0,1\ldots\infty)$ is of the form

(70) $$y^{(n)} = A_n y + B_n y'$$

where A_n and B_n are polynomials in F and its differential coefficients. In particular

(71a) $A_0 = 1$ $\quad A_1 = 0$ $\quad A_2 = -F$

(71b) $B_0 = 0$ $\quad B_1 = 1$ $\quad B_2 = 0$

From

$$y^{(n+2)} = -\sum_{i=0}^{n} \binom{n}{i} F^{(n-i)} y^{(i)}$$

we derive

(72a) $$A_{n+2} = -\sum_{i=0}^{n} \binom{n}{i} F^{(n-i)} A_i$$

(72b) $$B_{n+2} = -\sum_{i=0}^{n} \binom{n}{i} F^{(n-i)} B_i$$

This recurrence formulae and the initial conditions (71ab) give an easy means of calculating the A_n and B_n successively.

Differentiating (70) we further get the recurrence formulae

(73a) $$A_{n+1} = A'_n - FB_n$$

(73b) $$B_{n+1} = A_n + B'_n$$

from which we deduce

(74a) $$A_{2(h+1)} = -FA_{2h} + A''_{2h} - 2FB'_{2h} - F'B_{2h}$$

(74b) $$B_{2(h+1)} = -FB_{2h} + B''_{2h} + 2A'_{2h}$$

It is easily seen that A_{2h} and B_{2h} respectively is a sum of terms of the form

(75) $$F^{n_o} F'^{n_1} \dots F^{(k)n_k}$$

where the exponents $n_o n_1 \dots n_k$ are not negative integers.

The sum

(76) $$2n_o + 3n_1 + 4n_2 + \dots + (k+2)n_k$$

will be called the weight of the term (75).

I shall show that all the terms in A_{2h} have the same weight, namely 2h, and all the terms in B_{2h} have the same weight, namely (2h-1); 2h and (2h-1) will be called the weight of A_{2h} and B_{2h} respectively.

We first notice that if the term (75) is multiplied by F, then the weight is augmented by 2, and if the term is multiplied by F', the weight is augmented by 3.

Further if the term (75) is differentiated with respect to t, the result will be a sum of terms whose weight is one unit greater than the weight of (75). In fact the differentiation of (75) with respect to t may be performed by first differentiating with respect to F and multiplying the result by F', then differentiating with respect to F' and multiplying the result by F", etc. Now if (75) is differentiated with respect to $F^{(j)}$ and the

result is multiplied by $F^{(j+1)}$, we get a term where the expression for the weight is the same as (76) with the only exception that the term $(j+2)n_j$ in (76) is replaced by $(j+2)(n_j-1)$ and the term $(j+3)n_{j+1}$ is replaced by $(j+3)(n_{j+1}+1)$.
As

$$(j+2)(n_j-1) + (j+3)(n_{j+1}+1) - ((j+2)n_j + (j+3)n_{j+1}) = 1$$

we see that the weight of the term obtained will be one unit greater than the weight of (75).

Two differentiations of (75) will consequently transform the term (75) in a sum of terms whose weight is two units greater than the weight of (75).

The proposition regarding the weights of A_{2h} and B_{2h} may now be proved by complete induction using (74ab).

In fact the proposition holds good for h = 1 because $A_2 = -F$ and $B_2 = 0$. Supposing it to hold good up to h we see that in the right hand side of (74a)

FA_{2h} has the weight $(2h) + 2 = 2h+2$

A''_{2h} has the weight $(2h) + 2 = 2h+2$

FB'_{2h} has the weight $(2h-1) + 1 + 2 = 2h+2$

$F'B_{2h}$ has the weight $(2h-1) + 3 = 2h+2$

Consequently $A_{2(h+1)}$ has the weight $2h+2$.

In the right hand side of (74b)

FB_{2h} has the weight $(2h-1) + 2 = 2h+1$

B''_{2h} has the weight $(2h-1) + 2 = 2h+1$

A'_{2h} has the weight $(2h) + 1 = 2h+1$

Consequently $B_{2(h+1)}$ has the weight $2h+1$.

For any term in A_{2h} and B_{2h} we therefore have respectively

$$(77) \qquad 2n_o + 3n_1 + 4n_2 + \ldots + (k+2)n_k = \begin{cases} 2h & (\text{in } A_{2h}) \\ 2h-1 & (\text{in } B_{2h}) \end{cases}$$

As $n_o n_1 \ldots n_k$ are not negative integers, we see from (77) that the highest possible value of n_o in A_{2h} is $n_o = h$ (hence $n_1 = n_2 = \ldots = n_k = 0$) and in B_{2h} $n_o = h-2$ (hence $n_1=1$, $n_2=\ldots=n_k=0$).

It is further easily proved that the term with $n_o = h$, i.e. the term F^h is always present in A_{2h} and has the coefficient $(-)^h$.

We first notice that in the sum resulting from one differentiation of (75) none of the terms can have a n_o which is greater than n_o in (75). In particular, if in (75) n_o is the only non-vanishing exponent, then one differentiation of (75) gives a single term where n_o is one unit less than n_o in (75).

Utilizing this remark the proposition concerning the presence of the term $(-F)^h$ in A_{2h} may be proved by complete induction. The proposition evidently holds good for $h = 1$. Now suppose that it holds good up to h. Then the first term $-FA_{2h}$ in the right hand side of (74a) contains the term $(-F)^{h+1}$. The second term A''_{2h} cannot contain F^{h+1}. Further B'_{2h} does not contain a higher power of F than h-2 (because in B_{2h} we must have $n_o \leqslant h-2$), hence FB'_{2h} does not contain F^{h+1}. Finally $F'B_{2h}$ does not contain a higher power of F than h-2. Consequently in $A_{2(h+1)}$ only one term with F^{h+1} is present, namely the term $(-F)^{h+1}$.

From (77) is further seen that in A_{2h} we must have $k \leqslant 2h-2$, and in B_{2h} we must have $k \leqslant 2h-3$. The highest order differential coefficient of F that occurs is therefore $F^{(2h-2)}$ in A_{2h} and $F^{(2h-3)}$ in B_{2h}. That these differential coefficients really occur is easily seen from (72ab).

Consequently $y^{(2h)}$ is of the form

$$y^{(2h)} = (-F)^h y(1 + P_{2h} + (y'/y\sqrt{F})Q_{2h}) \tag{78}$$

where P_{2h} is a sum of terms whose denominator is F^h and whose numerator is of the form (75) with weight 2h and with not all the $n_1 n_2 .. n_k$ equal to zero, and Q_{2h} is a sum of terms whose denominator is $F^{h-\frac{1}{2}}$ and whose numerator is of the form (75) with weight 2h-1 and with not all the $n_1 n_2 ... n_k$ equal to zero.

Any term in P_{2h} or Q_{2h} may therefore be written

$$\frac{F^{n_o} \, F'^{n_1} \quad \dots \quad F^{(k)n_k}}{F^{n_o} \, F^{(1+\frac{1}{2})n_1} \quad \dots \quad F^{(1+k/2)n_k}}$$

Now the higher order increase-proportions of F defined by (15b) of Section 2 are

$$r^{(k)} = \pi^k F^{(k)}/F^{1+k/2}$$

For the sake of brevity we write

$$r_k = r^{(k)}$$

If the increase-proportions r_k are introduced, the general term of P_{2h} and Q_{2h} will be

$$(r_1^{n_1} \, r_2^{n_2} \, \dots \, r_k^{n_k}) \pi^{-(n_1+2n_2+\dots+kn_k)}$$

where not all the $n_1 n_2 ... n_k$ are zero, and where the highest k in any term in P_{2h} is $k = 2h-2$ and in Q_{2h} $k = 2h-3$.

For $h=3$, we have for instance

$y^{(6)} =$

$$(-F)^3 y \left[1-((4/\pi^2)r_1^2+(7/\pi^2)r_2-(1/\pi^4)r_4)+(y'/y\sqrt{F})((6/\pi)r_1-(4/\pi^3)r_3)\right]$$

Now $y'/y\sqrt{F}$ is the first order increase-proportion of y itself. This is finite for any point which is not a zero of y. We therefore see that if the increase-proportions of F up to the order 2h-2 are small as compared with unity, and also so small that the product obtained by multiplying them by the first order increase-proportion of y, is small as compared with unity (in points not lying in the vicinity of a zero of y), then we shall approximately have

$$y^{(2h)} = (-F)^h y$$

in points not lying in the vicinity of a zero of y.

If this holds good for any of the trends y_n and for $h = 1,2...(2N+1)$, the equation (69) will hold good approximately in points not lying in the vicinity of a zero for any of the trends. Therefore in this case w may be analyzed by the method of instantaneous approximation, the approximation curve with moving parameters being (50). And the moving determination of the gravitations and ordinates of the trends in a point not lying in the vicinity of a zero of one of the trends, is to be performed by (65) and (68a or b).

The method developed in this section may be called the method of moving differences because if a time series w_t of discrete values is given, the differential coefficients have to be determined by the successive differences of the given series.

7. SUMMARY

In this section I shall give a summary of the results obtained in the preceding sections and state some practical computation rules which may be derived therefrom. The order to be followed will not be the logical order of the preceding sections but rather the order in which the actual computations have to be performed. All demonstrations will be left out in the summary. I further believe that it should not be necessary to indicate for each step in the analysis the conditions under which the various approximations hold good. These conditions can easily be formulated by reverting to the text of the preceding sections.

When a time series of discrete values w_t is given, the differential coefficients have to be determined from the successive differences (eventually divided differences) by one of the known methods. In most practical cases it will probably be found sufficient simply to put the k-th order differential coefficient in the point t equal to the k-th order difference (eventually divided difference) which extends over an interval whose center is t. Let v_h be the empirically determined differential coefficient of order 2h.

In the simplest case (method of normal points, indicated below) it is sufficient to take account of v_1 only, i.e. the second differential coefficient of the composite curve. If a more refined analysis has to be performed, then a certain number of the following $v_2 v_3 \ldots$ must be introduced.

The first thing to do, is to plot the variation of v_1 or the variation of a certain number of the first $v_1 v_2 \dots$ as the case may be, If the fluctuations in these quantities are very violent and irregular, all the v_h considered changing sign frequently within intervals so short as to contain only some few original data, this will indicate the presence of components of such a low order that they cannot be investigated with the material available.

These components are not necessarily all of a true accidental character. In fact, if data were available at shorter intervals it might be possible to trace real periodic fluctuations in some of these components. But in the actual material they have to be considered as accidental. Consequently the group of these components will be called the accidental group.

Before proceeding to the analysis of the higher components, the accidental group has to be eliminated (so far as it can be done) by a moving average smoothing of the original series.

NB

I shall consider separately the two cases referred to.

The first case is the case where a priori considerations make it plausible to assume that the trend of lowest order y_0 which will be present after elimination of the accidental group, is of considerably lower trend order than the next following trend y_1. This means that the period in y_0 is small and the distinctness of the oscillations in y_0 (in the sense indicated in Section 2) is not very small as compared with the period and the distinctness of the oscillations of y_1. In this case y_0 may be eliminated and its normal as well as its ordinate determined by the method of normal points which only involves v_1.

If only v_1 has to be introduced and if the data are monthly, (hence equidistant) it will probably in most cases be found sufficient to smooth the original series by a twice iterated moving quarterly average. This means that before determining v_1 the original item w_t has to be replaced by the item

$$\dot{w}_t = (w_{t+2} + 2w_{t+1} + 3w_t + 2w_{t-1} + w_{t-2})/9$$

If in this case the second differential coefficient $w''(t)$ is approximated by simply putting it equal to the second difference, we have

$$w''(t) = v_1(t) = \Delta^2 \dot{w}_{t-1} = \dot{w}_{t+1} - 2\dot{w}_t + \dot{w}_{t-1} \quad = (w_{t+3} - 2w_t + w_{t-3})/9$$

In this case therefore the composite operation of first smoothing and then forming the second difference may be performed graphically in the plot of the original data w_t simply by measuring the deviation of w_t from the straight line through w_{t+3} and w_{t-3}, and multiplying by 2/9.

If a heavier smoothing should be found necessary or if the data are not equidistant, then the smoothing and the difference operation can not be performed by such a simple rule.

Now if v_1 is plotted, the procedure of first eliminating and then determining the normal and the ordinate of y_o will be this.

Determine the points where v_1 passes zero. These points are taken as the normal points of y_o. Since $y_o = 0$ in the normal points, the data relating to these points may be considered as a new series where the trend of lowest order is eliminated. Let W_o be the series from which v_1 was determined, and let W_1 be the new series derived from W_0 by the method of normal points. Since the difference in order between y_0 and the following trend is great, the ordinate of y_0 will be represented approximately by the deviation of W_0 from a line interpolated in some way or another through the series W_1 which represents the normal points for y_0. We may, for instance, draw a straight line between every two consecutive normal points (i.e. between every two consecutive data in W_1), or draw a m-th order parabola through (m+1) consecutive normal points, or use any other method of curve fitting. The essential point is that whatever the method used may be, the data W_1 determining the interpolation line, are data where the lowest order trend y_0 is already eliminated.

The series W_1 or the line interpolated through W_1 may evidently be considered as the normal of y_0.

If the further assumption can be made that the lowest order trend present in W_1 namely y_1 is of considerably lower trend order than the following trend y_2, then W_1 may be treated in exactly the same manner as W_0, thus eliminating y_1 and so forth.

Let $W_0 W_1 W_2$ ... be the series successively obtained by the method of normal points under the assumption that each trend is of considerably higher order than the preceding . Then W_n contains only the trends $y_n y_{n+1}$ ... . And the ordinate of y_n will be represented by the deviation of W_n from a line interpolated through W_{n+1} by one of the methods referred to. W_{n+1} evidently represents the normal of y_n.

The number of data in the successive series $W_0 W_1 W_2$ ... is rapidly diminishing. We finally arrive at a series W_N containing only some few data. This series W_N may be taken as representing the trend of highest order y_N which can be traced in the given series.

If no plausible a priori assumption can be made concerning the relative difference in order between y_o and the following trends, we shall have to introduce the higher v_h in order to perform a classification of the trends present. The practical applicability of the method involving the higher v_h is of course contingent upon a more reliable and more regular material than the method which only involves v_1.

The quantities v_h are to be interpreted as the differential coefficients which are determined empirically after the smoothing of the original data. It will generally be found that the smoothing of the original series must be all the heavier the higher is the order of the highest v_h to be considered.

As in most practical cases the original data w_t are equidistant, I shall make this assumption. I shall further assume that the graduation of the original data is performed by a central moving average with (2k+1) terms and with symmetrical but otherwise arbitrary weights λ_j. This means that the original item w_t is replaced by

$$\bar{w}_t = (\lambda_o w_t + \lambda_1(w_{t+1}+w_{t-1}) + \ldots + \lambda_k(w_{t+k}+w_{t-k}))/(\lambda_o + 2(\lambda_1+\lambda_2+\ldots+\lambda_k))$$

The length of the interval covered by the average, i.e. 2k, is supposed to be less than the shortest zero distance in the lowest order trend y_o, which is present after moving average elimination of the accidental group.

If a graduation of this kind is performed, the effect will be not only the elimination of the accidental fluctuations but also to a certain extent the damping, i.e. the depression, of the oscillations in the various higher trends present. The proportion in which the ordinate of the trend y_n is shortened in the point t, is approximately measured by the coefficient of damping ρ_n

$$(79) \qquad \rho_n = (\lambda_0 + 2\sum_{j=1}^{k} \lambda_j \cos j\sqrt{F_n})/(\lambda_0 + 2\sum_{j=1}^{k} \lambda_j)$$

where F_n designates the gravitation of the trend y_n.

It is seen that ρ_n is always less than unity. It is certainly positive if $2k < \pi/\sqrt{F_n}$, i.e. if the length of the moving average interval is less than the fictive zero distance in y_n in the point t considered (the fictive zero distance having the significance indicated in Section 2). If the length of the moving average interval is very small as compared with the zero distance in y_n, then ρ_n is close to unity. This means that the moving average operation has nearly no effect on the ordinate of y_n. In this case, the correction ρ_n in the formulae (81a), (82), (87ab), (89) and (91) below may be left out by putting $\rho_n = 1$.

I now procede to state the method by which the various trends may be classified and their gravitations and ordinates determined.

When a certain number of the first v_h are plotted, the determinants $\Delta_1 \Delta_2 \dots$ (formula (52)) should be formed and the fluctuations of the first of these determinants should be plotted. The question is if any of the Δ_n vanishes identically in t.

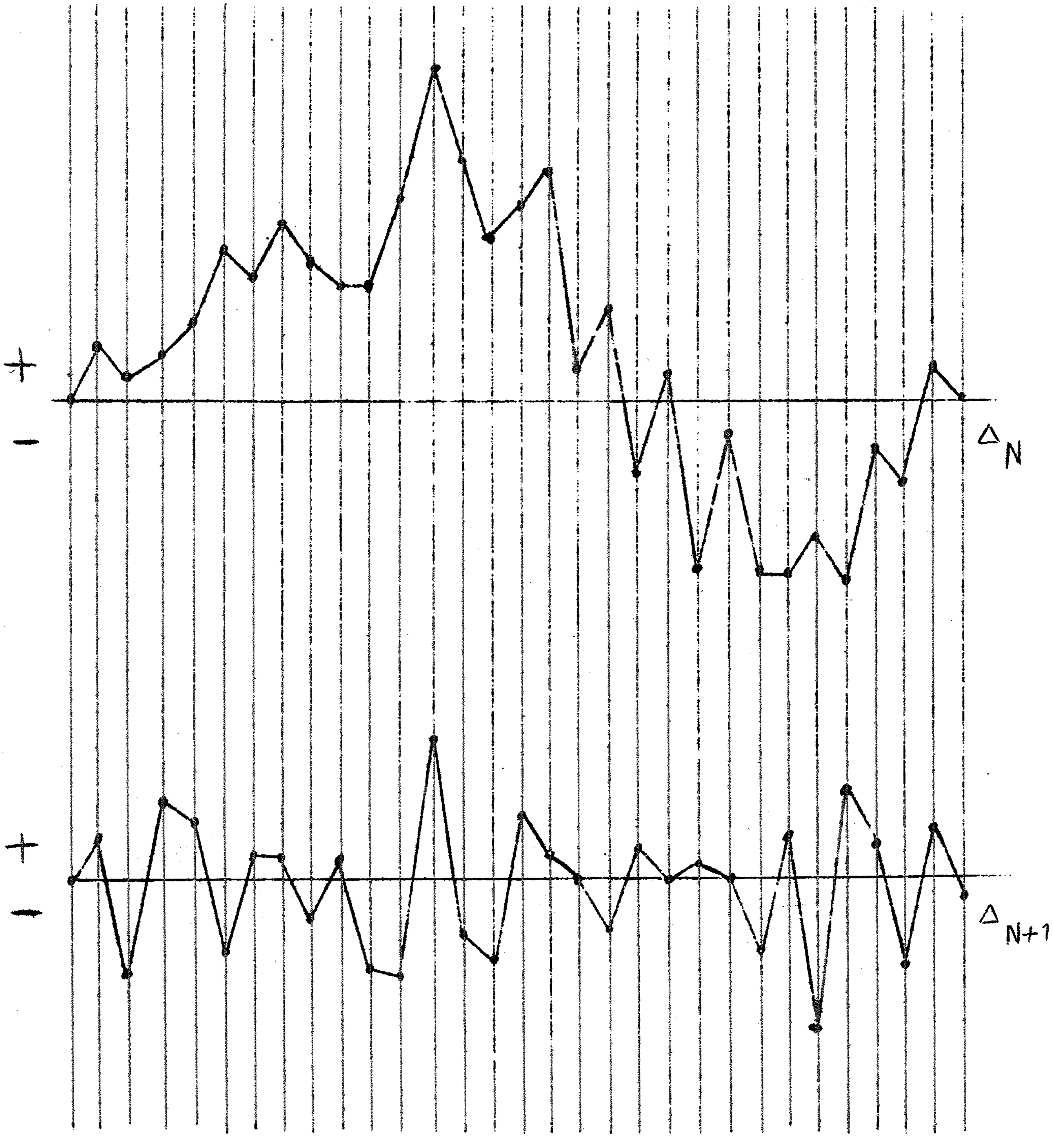
In practice, none of the Δ_n will be rigorously zero for all values of t. The practical criterion will be if the fluctuations in any of the Δ_n, say Δ_{N+1}, are of such a kind that Δ_{N+1} changes sign frequently, say in points between which there are only one or some few points of observation, and further such that the average of Δ_{N+1} taken over small intervals is practically zero over the whole t-range. When in the following reference is made to the identical vanishing of determinants, it should always be understood as identical vanishing in this practical sense. In the illustration below Δ_{N+1} would represent an identically vanishing determinant (in the practical sense), and Δ_N would represent a non-identically vanishing determinant.

Now if Δ_{N+1} vanishes identically and it can be traced quite distinctly certain intervals where the next lower determinant Δ_N, or the smoothed curve representing Δ_N, is essentially positive, or intervals where it is essentially negative, then the number of trends in the lowest trend group will be equal to N.

This means that there are present N trends which do not differ widely with respect to thelength of the period. They are approximately of the same trend order. But the difference in order between these trends and the next following is great.

The trends in the lowest trend group have to be eliminated and their ordinates determined before we proceed to analyze the higher order trends.

I shall first consider the two special cases N = 1 and N = 2.


+
-
Δ_N
+
-
Δ_{N+1}

which will probably be the most frequent cases when the time series to be analyzed represents an economic phenomenon.

If N = 1, i.e., if $\Delta_1 = v_1$ shows distinct positive intervals or distinct negative intervals while

(80)
$$\Delta_2 = \begin{vmatrix} v_1 & v_2 \\ v_2 & v_3 \end{vmatrix} = v_1 v_3 - v_2^2$$

vanishes identically, then the lowest trend group consists of only one single trend y_0. And the difference in order between y_0 and the next following trend is great. In this case the lowest order trend y_0 may be determined either by the method of normal points or by the method of moving differences.

The practical computations involved in the determination of y_0 by the method of normal points have already been indicated.

The method of moving differences may be applied in two different ways to approximate the values of y_0. In the first place we may consider the gravitation F_0 as constant in the interval between two consecutive normal points, and here put F_0 equal to its average value in this interval. This average value is $(\pi/D)^2$, where D designates the distance between the two normal points considered. In this case the ordinate of y_0 in the interval considered would be

(81a)
$$\rho_0 y_0 = -(D/\pi)^2 v_1$$

where ρ_0 is determined by (79). If the length of the moving average interval is small as compared with D, then the correction ρ_0 may be left out by putting $\rho_0 = 1$.

The method of formula (81a) only involves v_1. If v_2 is introduced, we may consider F_0 not only as moving by steps from one zero distance in y_0 to the next, but as moving continuously from one point of observation to the next. In this case F_0 would be determined by

(81b) $$F_0 = -v_2/v_1$$

If (81b) is used, the determination of F_0 will be uncertain in the vicinity of the normal points, because in all practical cases v_1 and v_2 will not vanish rigorously in the same points (as assumed by theory). Therefore F_0 has to be plotted and its value in a point lying in the vicinity of one of the normal points has to be determined rather by interpolation in taking account of the general shape of the curve F_0 than by calculating the actual value of $-v_2/v_1$ in the point considered. In determining the general shape of the curve F_0 it should be remembered that the average value of F_0 between two consecutive normal points is $(\pi/D)^2$ where D is the distance between the two normal points considered. F_0 should never vanish, but be essentially positive.

When the variation in F_0 is determined, y_0 is given by

(82) $$\rho_0 y_0 = -v_1/F_0$$

I now proceed to the second special case, viz. N = 2.

If Δ_2 shows distinct positive intervals or distinct negative intervals while

(83) $$\Delta_3 = \begin{vmatrix} v_1 & v_2 & v_3 \\ v_2 & v_3 & v_4 \\ v_3 & v_4 & v_5 \end{vmatrix}$$

vanishes identically, then the lowest trend group consists of two components y_o and y_1 between which the differences in trend order is not great. But the difference in order between this first group and the next following is great. In this case y_o can not be eliminated before y_1, y_o and y_1 have to be considered simultaneously.

We now have to form the determinants d_o, d_1 and d_2 which are obtained from the matrix

$$\begin{pmatrix} v_1 & v_2 & v_3 \\ v_2 & v_3 & v_4 \end{pmatrix}$$

by letting out the first, second and third column respectively. Hence

(84) $$d_o = \begin{vmatrix} v_2 & v_3 \\ v_3 & v_4 \end{vmatrix} = v_2 v_4 - v_3^2$$

$$d_1 = \begin{vmatrix} v_1 & v_3 \\ v_2 & v_4 \end{vmatrix} = v_1 v_4 - v_2 v_3$$

$$d_2 = \begin{vmatrix} v_1 & v_2 \\ v_2 & v_3 \end{vmatrix} = v_1 v_3 - v_2^2 = \Delta_2$$

The variations in the two determinants d_o and d_1 have to be plotted on the same chart as $d_2 = \Delta_2$ already plotted. The three determinants, d_o, d_1 and d_2 should generally show alternating signs, i.e. either d_o positive, d_1 negative, and d_2 positive or inversely. And they should pass zero in approximately the same points. The method will work all the better the more closely this condition is satisfied.

The next step is to plot the variation in the ratio

(85) $$x = d_o d_2/d_1^2$$

In the vicinity of the points where d_1 vanishes, the determination of x by (85) will be uncertain because the three determinants will not in practice vanish rigorously in the same points (as assumed by theory). Therefore the value of x in the vicinity of the zeros of d_1 has to be interpolated rather than determined by (85). If necessary the curve x should be smoothed. The ratio x should generally be a fraction between 0 and $\frac{1}{4}$. If x is greater than $\frac{1}{4}$ throughout a definite interval and no plausible smoothing will bring it down to $\frac{1}{4}$, then the method will not work.

The moving determination of the two gravitations F_o and F_1 has to be performed by the following formulae

(86a) $$F_o = -(d_1/2d_2)(1 + \sqrt{1-4x})$$

(86b) $$F_1 = -(d_1/2d_2)(1 - \sqrt{1-4x})$$

In the vicinity of the zeros of d_2 the determination of F_o and F_1 by (86ab) will again be uncertain. Interpolation should therefore be used in the vicinity of these points. F_o and F_1 should never vanish, but be essentially positive.

It is seen that the ratio between the two gravitations will be all the greater the smaller x. The ratio $1/\sqrt{x}$ or better $(1-x)/\sqrt{x}$ will indicate approximately the number of (fictive) y_o periods which are contained in each y_1 period.

The ordinates of the trends y_o and y_1 are given by

(87a) $$\rho_o y_o = (v_2 - F_1 v_1)/F_o(F_o - F_1)$$

(87b) $$\rho_1 y_1 = -(v_2 - F_o v_1)/F_1(F_o - F_1)$$

where ρ_o and ρ_1 are defined by (79).

If x is small, i.e., if the ratio F_0/F_1 is great, then the first term in the expression for y_0 will be predominant over the second, but in y_1 both terms will be of the same order.

I now proceed to the general case where the lowest trend group consists of any number of trends N.

As has already been pointed out the number N is determined by the criterion that Δ_N shall show distinct positive periods or distinct negative periods while Δ_{N+1} shall vanish identically.

We now have to form the determinants $d_0 d_1 \ldots d_N$ which are obtained by letting out the first, second... and finally the last (i.e. the (N+1)th) column respectively in the matrix

(88)
$$\begin{pmatrix} v_1 & v_2 & \cdots v_{N+1} \\ v_2 & v_3 & \cdots v_{N+1} \\ \cdots & \cdots & \cdots \\ v_N & v_{N+1} & \cdots v_{2N} \end{pmatrix}$$

The various d_h have to be plotted. These plots should be such that the quantities $d_0 d_1 \ldots d_N$ generally show alternating signs and pass zero in approximately the same points. The method will work all the better the more closely this condition is satisfied.

Now in a point where the d_h have alternating signs, the gravitations $F_0 F_1 \ldots F_{N-1}$ of the N trends in the lowest group are determined as the N roots of the period equation (65). The variations in these roots have to be plotted. In the vicinity of a point where one of the d_h vanishes, the determination of the F_n will be uncertain. The magnitudes of the various F_n in the vanishing intervals of the d_h have therefore to be interpolated by taking account of the general shape of the F_n curves outside these intervals.

When the N roots of (65) are plotted and if necessary the curves smoothed, the ordinates $y_0 y_1 \dots y_{N-1}$ of the trends in the lowest trend group are given in any point by

(89) $$\rho_n (F_o - F_n)(F_1 - F_n)\dots((F_n - F_n))\dots(F_{N-1} - F_n)(-F_n) y_n = \sum_{h=0}^{N-1} a_h^{(n)} v_{N-h}$$

where $((F_n - F_n))$ indicates that this product shall be excluded in the left hand side of the equation. The corrections ρ_n are defined by (79). And

(90) $$a_h^{(n)} = (F_o F_1 \dots ((\text{except } F_n))\dots F_{N-1})_h$$

designates the h-th order elementary symmetric function of the quantities $F_o F_1 \dots ((\text{except } F_n))\dots F_{N-1}$. That is the sum of all the products of h factors which can be formed by picking out in all possible ways h of the N-1 quantities $F_o F_1 \dots ((\text{except } F_n))\dots F_{N-1}$.

The value of the y_n in any point can also be calculated from the formula

(91) $$\rho_n (F_o - F_n)(F_1 - F_n)\dots((F_n - F_n))\dots(F_{N-1} - F_n)(-F_n) y_n = \Delta_N / \Delta_N^{(n)}$$

where

$$\Delta_N^{(n)} = \begin{vmatrix} v_1 & v_2 & \dots v_N \\ v_2 & v_3 & \dots v_{N+1} \\ \dots & \dots & \dots \\ v_{N-1} & v_N & \dots v_{2N-2} \\ 1 & (-F_n)\dots & (-F_n)^{N-1} \end{vmatrix}$$

The values of y_n given by the two formulae (89) and (91) should approximately coincide.

If the ordinates of the trends in the lowest trend group are subtracted from the original data, we shall have a series where only the trends of higher order are present. This series may be considered as the normal of the composite curve which represents the first trend group. None of the components in the first trend group can be considered as a component in the curve which represents the normal for the other components in the first group. The first N components must be considered as a unity. This distinguishes the present case from the case $\Delta_2 = 0$.

As the difference in order between the first group and the following is great, the number of original data contained in each period is far greater for the next trend group than it was for the lowest trend group. This is an important point. It makes it possible to perform a new and effective smoothing of the residuum left by subtraction of the lowest group from the original data. In fact, in this case the correction (79) for the ordinates in the next group would be close to unity and could be left out. It would even be possible to analyze the higher order trend by only taking account of isolated points at intervals amounting to several times the interval between consecutive data in the original series.

Now let W_1 be the series obtained after the elimination of the lowest trend group. The series W_1 may be treated in exactly the same manner as the series of original data, thus eliminating the next trend group and so forth. If Δ_2 for W_1 should vanish identically, then the lowest trend group in W_1 would consist of one single trend which might be eliminated by the method of normal points or the method of moving differences. If Δ_3 for W_1 should vanish identically but not Δ_2, then the lowest trend group in W_1 would consist of two trends, and so forth.

Let W_0 be the series from which the original v_1 v_2 ... were determined, and let $W_1 W_2 \ldots$ be the series successively obtained from W_0 after elimination of the lowest trend group, the second trend group, and so forth. If Δ_2 should vanish identically not only for W_0 but also for $W_1 W_2 \ldots$ and the following series, then we should have the special case where each group consists of one single trend only. This is just the case considered under the development of the method of normal points. The series $W_0 W_1 \ldots$ now considered are therefore analagous to but more general than the series $W_0 W_1 \ldots$ of the method of normal points. There we have eliminated one trend at a time, here we have eliminated one trend group at a time.

pro tem. Ragnar Frisch

New York City, April, 1927