COWLES COMMISSION FOR RESEARCH IN ECONOMICS

MONOGRAPH NO. 5

PUBLISHED JOINTLY BY THE
COWLES COMMISSION AND THE
DEPARTMENT OF ECONOMICS AND
SOCIOLOGY, IOWA STATE COLLEGE

THE VARIATE DIFFERENCE METHOD

By

GERHARD TINTNER

PRINCIPIA PRESS, INC.
BLOOMINGTON, INDIANA
1940

SET UP AND PRINTED
IN THE UNITED STATES OF AMERICA BY
THE DENTAN PRINTING COMPANY
COLORADO SPRINGS, COLORADO

PREFACE

This volume is intended as a modest contribution to the much discussed problem of the analysis of economic time series. It does not, of course, try to offer anything but a very tentative solution of one particular problem out of this great complex. It deals with the role and importance of the random element. But this seems to be one of the most fundamental and important aspects of the whole problem, whose importance for economics cannot be overemphasized. It arises from the fact that the analysis of the random element can give us somewhat secure foundations for statistical tests of our empirical results and enables us to establish their validity from the point of view of modern statistics. The author has also endeavoured to give an economic foundation, also rather tentative, to the statistical procedures involved.

The book may be of some interest to a variety of scholars. The economist, especially the nonmathematical economist, is referred to Chapter I and Appendixes VI and VII. The economic and rather nonmathematical statistician will perhaps be interested in Chapters V to X which present the Variate Difference Method in a form comparatively free from mathematics. Appendix I gives a summary of computations. It is also hoped that some of the tables may be useful in empirical applications and reduce the amount of calculations, which is still considerable. The mathematical statistician is referred to Appendixes II and V. The historian of statistics will possibly be interested in Chapter II.

Most of the material contained in the book goes back to the original work of Professor O. Anderson and Dr. R. Zaycoff, of Sofia, Bulgaria. The author is much obliged to them for the permission to use some of the material published first in the *Publications of the Statistical Institute for Economic Research,* State University of Sofia. A new method is presented in Chapter VIII and Chapter X, Section B.

The work on this book was carried out under the auspices of the Iowa State College Agricultural Experiment Station, Project Number 557. The author has to thank Mr. Norman Strand (Ames, Iowa) for helping him with a W.P.A. project, O.P. No. 665-72-3-46, which did most of the calculation of the tables. Mrs. Gevernia Richardson (Ames, Iowa) supervised and checked the calculations and proofread the book. She acted as the author's secretary and it is due to her, if

the number of errors in the book and especially the tables is only a minimum. Dr. I. Lubin (U. S. Bureau of Labor Statistics) was very helpful in connection with the collection of the agricultural price material used as examples in this book.

Mr. Alfred Cowles and Professor H. T. Davis of the Cowles Commission and Professor T. W. Schultz and Dean E. W. Lindstrom of Iowa State College made possible the publication of this book which is a common venture of the Cowles Commission for Research in Economics and the Department of Economics and Sociology at Iowa State College. Mr. D. H. Leavens of the staff of the Cowles Commission, is to be credited with the editing of the book. He also very kindly arranged the drawing of the figures.

The author wants to express his sincere thanks to the following who were very helpful in connection with the present book: Professor Edward S. Allen (Iowa State College), Professor Eugen Altschul (University of Minnesota), Mr. R. L. Anderson (Iowa State College), Professor E. C. Bratt (Lehigh University), Mr. W. G. Cochran (Iowa State College, formerly Rothamsted), Mr. Alvin E. Coons (Iowa State College), Professor H. T. Davis (Northwestern University), Professor Harold Hotelling (Columbia University), Mr. Sven Laursen (Copenhagen), Professor Edward S. Lynch (Iowa State College), Mr. Herbert E. Jones (Cowles Commission), Mr. W. G. Madow (Columbia University), Dr. Horst Mendershausen (Colorado College), Professor Erich Rothe (Penn College), the late Professor Henry Schultz (University of Chicago), Professor Theodore W. Schultz (Iowa State College), Professor George W. Snedecor (Iowa State College), Mr. Herman M. Southworth (Iowa State College), Professor George Stigler (University of Minnesota), Dr. A. Wald (Columbia University), Professor S. S. Wilks (Princeton University), Dr. R. Zaycoff (Sofia). Needless to say, none of these who very kindly helped me in various ways are to blame for any mistakes, errors, omissions, and general blunders in this book, which the author claims as his very own.

GERHARD TINTNER

Departments of Economics and Mathematics
Iowa State College
Ames, Iowa
December, 1939

TABLE OF CONTENTS

CHAPTER X
CORRELATION

APPENDIX I
SUMMARY OF COMPUTATONS

APPENDIX II
MATHEMATICAL NOTES

APPENDIX III

APPENDIX IV

APPENDIX V
ALTERNATIVE METHODS

APPENDIX VI

APPENDIX VII

TABLE OF CONTENTS

LIST OF FIGURES

THE VARIATE DIFFERENCE METHOD

CHAPTER I

INTRODUCTION

A. The Problem of Time Series

The variate difference method is a statistical tool for the analysis of the random element in time series. The statistical methods with which we deal in the treatment of time series are of special interest to economists since most economic data come in the form of items that are ordered in time. We use, for instance, monthly series of wholesale prices, yearly series of production of certain commodities, and so on. It is, therefore, not surprising that a great deal of attention has in the past been given to this problem.

It is, however, regrettable that no closer collaboration has prevailed between the economists and the economic statisticians. The economist feels that he has to take into account whatever he knows from economic theory about the nature of the behavior of economic quantities through time.[1] The mathematical statisticians in general do nothing of this kind. Their contributions, which are sometimes very valuable and often display great ingenuity, have been mainly mathematical and mechanical. Most statistical methods fail to make sense if interpreted from the point of view of the economist.

We shall try in this study to take the strictly economic point of view. That is to say, every statistical or mathematical operation, e.g., taking an average or calculating a trend, has in our opinion a very definite economic meaning. We think that it should be based on certain economic assumptions which should be clearly stated and which, if clearly stated, in many cases will make us much more cautious in the appraisal of particular statistical methods and results. The application of any specific statistical method to an economic problem should be made only if it can be handled and understood in economic terms.

But we should realize at the outset that we are not always going to get a unique and definite answer from economics. The main reason for this is that our knowledge of the forces that act on the development of economic phenomena through time is extremely limited. The part of economic theory that refers to it is known (not too properly)

[1] O. Morgenstern, *Wirtschaftsprognose*, Vienna, 1928.

as the theory of economic dynamics.[2] Most of its problems are still very far from even a theoretical solution in spite of the substantial progress that has been made in this field in the last ten years. We know almost nothing about the numerical values of the specific quantitative parameters and coefficients which occur in dynamic economics.[3]

Hence, the economic assumptions on which we have to base our statistical methods are unfortunately not very refined to start with, and in any case very indefinite. This is certainly a handicap. If we propose, nevertheless, certain methods for dealing with a specific aspect of the problem of time series we do so because we feel that economic theory, too, can make progress if the right or at any rate better statistical methods are used. Only then can we hope to get at least an idea of the order of magnitude of the quantities that enter into dynamic economics and possibly, also, some indications for the further theoretical treatment of such problems.

B. The Components of Economic Time Series

In general, four components have been distinguished in an economic time series: secular trend, cyclical fluctuation, seasonal component, and a remainder.[4] The variate difference method is particularly concerned with this remainder.

There seems to be no objection from the economic point of view to the distinction between these four components. In fact, we can really imagine that they refer to longer- and shorter-run considerations in the economic system.[5] If we take this approach we can think of the longest run as referring to the secular trend (which, by the way, may also contain so-called long waves) and the shortest as referring to the remainder term which, for reasons to be stated later, we shall call the random variation.

But the reader should be warned at the outset against overdoing

[2] F. H. Knight, "Statics and Dynamics" in *The Ethics of Competition*, New York, 1935, pp. 161 ff. See also G. C. Evans, *Mathematical Introduction to Economics*, New York, 1930; C. F. Roos, *Dynamic Economics*, Bloomington, Indiana, 1934; and J. R. Hicks, *Value and Capital*, Oxford, 1939.

[3] See, however, H. Schultz, *The Theory and Measurement of Demand*, Chicago, 1938, and some of the work of the econometrists as published in *Econometrica*.

[4] See, for instance, F. C. Mills, *Statistical Methods Applied to Economics and Business*, revised edition, New York, 1938, pp. 225 ff.; W. C. Mitchell, *Business Cycles, The Problem and Its Setting*, New York, 1927; C. F. Roos, "Correlation and Analysis of Time Series," *Econometrica*, Vol. 4, 1936, pp. 368 ff.; E. C. Bratt, "The Divisibility of Time Series," *Review of Economic Studies*, Vol. 5, 1938, pp. 79 ff.; W. Winkler, *Theoretische Statistik*, Berlin, 1931, pp. 106 ff.

[5] A. Marshall, *Principles of Economics*, 8th ed., London, 1920, pp. 378 ff.

this conventional distinction between the four components which suggests itself mainly from the mere visual impression of most economic time series when plotted graphically. We certainly are in no position to assume, for instance, that the components are independent. In fact, it has been established (and it is highly probable from a priori economic considerations) that there are definite relationships—for instance, between the secular trend and the cyclical movement. The point of view on these matters will depend essentially on the specific theory of the business cycle that is adopted. If, for instance, we accept Professor Schumpeter's theory of the cyclical movement,[6] there will be no reason at all for distinguishing between the trend and the business cycle since they are the same phenomenon. A similar connection exists between seasonal and cycle.[6a]

Another conclusion which we may draw from the study of economic theory and especially business-cycle theories is the following: On the basis of a priori economic considerations, it is highly improbable that economic time series should follow a strict and rigid pattern.[7] It is, on the contrary, probable that the conditions at different periods of time are not the same. We have no reason to believe that the economic forces that are effective during the boom are essentially the same as those that are at work during the slump. This refers to the cyclical movement. But there is also good reason to believe, for instance, that a long-run period with rising prices produces an entirely different situation from that arising from a period with falling prices. This has to do with long waves or the secular trend.[8]

Hence, we shall have to look for the statistical method that does least violence to our data and is not too strict and rigid. This is one of the reasons why the writer is, in general, opposed to the use of rigid seasonal coefficients and, similarly, to the calculation of trends according to some rigid formulae that cannot be justified on economic grounds. A curve like the logistic, on the other hand, which has definite, economically significant properties, could possibly be justified. In spite of many imperfections the writer prefers moving averages on grounds of flexibility, as he has pointed out in a previous publication.[9]

[6] J. A. Schumpeter, *The Theory of Economic Development*, Cambridge, Mass., 1934. See also W. C. Mitchell, *op. cit.*, pp. 412 ff.

[6a] S. Kuznets, *Seasonal Variations in Industry and Trade*, New York, 1933, pp. 303 ff.; J. Wiśniewski, "Interdependence of Cyclical and Seasonal Variation," *Econometrica*, Vol. 2, 1934, pp. 176 ff.

[7] See on this point especially G. Haberler, *Prosperity and Depression*, New Edition, Geneva, 1939, pp. 5 ff.

[8] G. Haberler, *op. cit.*, pp. 272 ff.; N. D. Kondratieff, "The Long Waves in Economic Life," *Review of Economic Statistics*, Vol. 17, 1935, pp. 105 ff.

[9] G. Tintner, *Prices in the Trade Cycle*, Vienna, 1935, pp. 22 ff.

C. *Economic "Errors" and Time Series*

The author has tried to give elsewhere[10] an approach to a theoretical study of the relationship between errors in economic life and the shape of economic time series. This is indeed an important problem as its solution affords a necessary foundation to the statistical methods that we propose to use here. We can, however, give only a short indication of the problems involved.

In the publication mentioned the author distinguished essentially two types of errors that occur in economic behavior. Errors of the first kind result from the fact that an individual who otherwise acts rationally and makes correct forecasts of all relevant future data fails for some reason to make all adaptations of the economic factors which he controls in such a manner as to give him the maximum utility or profit. The reasons for this are certain institutional obstacles, the influence of tradition, the sometimes imperfect work of organization, negligence, and similar "frictional" causes. We can make the assumption that those "errors of the first kind" will be more or less independent for the economic activities of the same individual and still more between different individuals. Since we know furthermore that those errors are probably of a rather small order of magnitude and as there are many of them, we may expect that this type of errors is of more or less random character. They are then much like the errors that occur in the natural sciences and which are treated in the so-called theory of errors.[11] We should not be surprised if their effects followed approximately the normal law of errors or at least a symmetrical law. But they are probably of greater magnitude in time of great economic change. Their variance can be expected to be somewhat correlated with the general fluctuations. They are possibly not homoskedastic but heteroskedastic in time.[12] They represent, so to speak, the influence of the nonpermanent causes that are working on the economic time series. It is clear, of course, that the term nonpermanent has a different meaning with different units of time, e.g., with yearly or monthly data. The fundamental units are very important for the definition of randomness. It is really the fourth component, the random element, or the "remainder" term in our time series, that corresponds roughly to these economic properties. (See also Appendixes VI, VII.)

[10] G. Tintner, "A Note on Economic Aspects of the Theory of Errors in Time Series," *Quarterly Journal of Economics*, Vol. 53, November, 1938, pp. 141 ff.

[11] See, e.g., D. Brunt, *The Combination of Observations*, Cambridge, 1931, pp. 11 ff.

[12] A. A. Tschuprow, *Grundbegriffe und Grundprobleme der Korrelationstheorie*, Leipzig, 1925, p. 31.

Besides this we have to distinguish errors of the second kind which are the effect of erroneous forecasts. The influence of those errors goes much deeper and it has been assumed by many theorists of the cycle that they are intimately connected with at least the duration and amplitude of cyclical variations if they are not their cause. Hence they have to do essentially with the second component of the economic time series, the business cycle, and fall outside the scope of our present study which is concerned only with the random part.[13]

Systematic deviations which enter into consecutive items with opposite signs would fall under the category of systematic errors or results of errors of the second kind, which are the outcome of faulty expectations and forecasts. As Mr. von Szeliski points out,[14] alternating errors, for instance in price series, may result from a tendency of violent changes in prices of speculative commodities and stocks on one day being followed by opposite changes on the next day. A similar point of view has also been presented by Mr. S. Kuznets.[14a] They are evidently results of mistakes in expectations and forecasts of noncontrollable factors and hence of errors of the second kind. If such errors are important, they will not produce a smooth course of the time series but some kind of a zigzag effect, where subsequent items are strongly negatively correlated with each other. Phenomena of this nature cannot be treated by the variate difference method, which can only separate purely random fluctuations and the "smooth" part of the time series. But they could be treated by the correlogram methods proposed by Mr. Wold[15] in his recent book. (See also Appendix V, Section B.)

The random element as defined above does not necessarily have to do with the results of "noneconomic" causes and should not be confused with them. These outside forces would only have effects which come under the category of random variations if they were not correctly foreseen and if their influence was not lasting and small.

[13] R. Frisch, "Propagation Problems and Impulse Problems in Dynamic Economics" in *Economic Essays in Honour of Gustav Cassel*, London, 1933, pp. 171 ff.; M. Kalecki, "A Macrodynamic Theory of Business Cycles," *Econometrica*, Vol. 3, July, 1935, pp. 327 ff.; J. Tinbergen, "Annual Survey: Suggestions on Quantitative Business Cycle Theory," *Econometrica*, Vol. 3, July, 1935, pp. 241 ff.; E. Slutzky, "The Summation of Random Causes as the Source of Cyclic Processes," *Econometrica*, Vol. 5, April, 1937, pp. 105 ff.; H. Working, "A Random Difference Series for Use in the Analysis of Time Series," *Journal of the American Statistical Association*, Vol. 29, 1934, pp. 11 ff.

[14] V. von Szeliski, "Analysis of Random Errors in Time Series," Appendix II in C. F. Roos, *Dynamic Economics*, Bloomington, Indiana, 1934, pp. 251 ff.

[14a] S. Kuznets, "Time Series," in *Encyclopaedia of the Social Sciences*, Vol. 14, New York, 1935, especially p. 635.

[15] H. Wold, *A Study in the Analysis of Stationary Time Series*, Uppsala, 1938, pp. 147 ff.

D. The Variate Difference Method[16]

A thorough analysis of the random element in economic time series is very important for the following reason: Any statistical comparison of time series or their components or characteristics must be based on the theory of probability on which all statistical methods necessarily rest. But it is impossible to get any comparisons between characteristics of economic time series that are valid, from the point of view of probability, without having at least an idea about the order of magnitude of the random element involved. Suppose we want to compare, for instance, two different seasonal patterns or the slope of two different trends or to test the degree of agreement between the business cycles in two different prices. It is then most important to know something about the standard errors of the statistical parameters involved, which result from random fluctuations. Only these standard errors or some equivalent information can give us an idea as to how valid ultimately our conclusions are from the point of view of probability. Hence, whereas the random element itself is probably not very important for economics, its importance arises from the probability nature of every statistical comparison and hence also of the comparison of certain characteristics of economic time series. It must be the basis of tests of significance and of statistical tests of hypotheses.

Another use to which the random variances of two time series may be put is the following: If we correlate two variables each of which is affected by errors, we sometimes desire not the two regression lines but one line of common relationship. The knowledge of the ratio of the random variances or of the matrix of the random variances and covariances may help us to determine the direction in which to minimize the errors.[17]

[16] See especially O. Anderson, "On the Logic of Decomposition of Statistical Series into Separate Components," *Journal of the Royal Statistical Society*, Vol. 90, 1927, pp. 548 ff.; O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, Bonn, 1929; G. Tintner: *Prices in the Trade Cycle*, pp. 9 ff.; G. Tintner, "On Tests of Significance in Time Series," *Annals of Mathematical Statistics*, Vol. 10, 1939, pp. 139 ff.

[17] See for this general problem H. Schultz, *The Theory and Measurement of Demand*, Chicago, 1938, p. 148, note; C. F. Roos: "A General Invariant Criterion of Fit for Lines and Planes where All Variates Are Subject to Error," *Metron*, Vol. 13, 1937, pp. 3 ff.; H. E. Jones, "Some Geometrical Considerations in the General Theory of Fitting Lines and Planes," *ibid.* pp. 21 ff. For related problems see also R. Frisch, *Confluence Analysis*, Oslo, 1934; H. Hotelling, "Relations between Two Sets of Variates," *Biometrika*, Vol. 28, 1936, pp. 321 ff.; T. Koopmans, *Linear Regression Analysis of Economic Time Series*, Haarlem, 1937; R. G. D. Allen, "The Assumption of Linear Regression," *Economica*, Vol. 6, 1939, pp. 191 ff. See also the forthcoming article of A. Wald, "The Fitting of Straight Lines if Both Variables are Subject to Error."

The variate difference method is designed to deal with the random element or the effect of errors of the first kind as defined above. We shall at this point give a brief indication of its nature and its methods.

The variate difference method starts with the fundamental assumption that every economic time series consists essentially of two parts. The first part is the mathematical expectation and is the effect of the somewhat permanent causes in economic life. We can in general expect that it will show a more or less smooth shape. It will give a nearly continuous curve if it is plotted against time. Consecutive items are highly positively, but not necessarily linearly, correlated with each other.

The other component is called the random element and is essentially the effect of the nonpermanent causes which work in economic life. It represents among other things the result of errors of the first kind. There is no reason to expect that the different items of the random element are positively or negatively correlated. It is random in nature and hence its items are more or less independent.

We assume that there is an additive connection between the random element and the mathematical expectation of our time series. Every item of the series consists of the mathematical expectation plus the random element.

Professor Anderson of Sofia was the first to use the variate difference method extensively for the analysis of economic data. He has devised an ingenious method for separating the two components which we just distinguished. He starts from the proposition that every smooth curve can be indefinitely reduced by the process of finite differencing. It is a well-known mathematical theorem that polynomials, for instance, can be entirely eliminated by finite differencing. We can simply assume that the smooth part of our time series, the mathematical expectation, can be approximated more or less closely by polynomials of the variable time which otherwise need not be specified. This approximation may hold true in a certain restricted neighborhood only. It is not necessary that the whole series can be represented by polynomials.

The same is certainly not true regarding the random element. The random element cannot be reduced by finite differencing, since it is not ordered in time. By its very nature, it cannot be significantly approximated by any kind of a function. If we, therefore, apply the process of successive finite differencing to our original time series we shall, in general, reduce the smooth element or the mathematical expectation more and more without significantly affecting the random element.

The variate difference method proposes to give a statistical method for answering this question: Beginning with which difference can we be reasonably sure that we have more or less eliminated the nonrandom element or mathematical expectation? The method consists in calculating the variances (squares of standard deviations) of the original series and of the series of successive finite differences. Then the difference between variances of two successive series of finite differences is compared with its standard error. If the difference is smaller than about three times its standard error, we can be reasonably certain, from the point of view of probabilities, that we have carried the finite differencing far enough (Chapters V to VII).

This method of Anderson's is based upon the idea of standard error. Strictly speaking therefore, it is applicable only for large samples, where the probability of a divergence between the sample variance and the true variance is not very great. But we propose, in Chapter VIII, a new method, in which we apply R. A. Fisher's z test or Snedecor's F table. In order to do this, we have to make the variances of two consecutive difference series independent, which can be done by the method of selection. We select the items from which the variances of two consecutive differences are calculated in such a way that they are really independent. This involves, of course, some sacrifice of information but gives us tests that are valid from the point of view of modern statistics.

Assume that we have established the order of the finite difference beginning from which we can be more or less sure that we have eliminated the mathematical expectation. We can then get an approximation to the latter by the use of certain smoothing formulae. These formulae eliminate the random element to a certain degree, but we can get only approximations to the mathematical expectation (Chapter IX). Our smoothed series will always contain a remainder of the original random element. This, however, may be very small and can be accurately estimated. It is then possible, also, to obtain an estimate of the random variation in the series and in some statistical series derived from the original or from the smoothed series. (See Appendix IV.)

We can also use the variate difference method in order to calculate the correlation between the random elements of two time series. Here again we shall calculate the covariances, i.e., the product moments of the differences, and try to find out at which difference they become stable. This can be done by formulae which involve the use of standard errors (Chapter X, Section A). This method, strictly speaking, is permitted only if the sample is large so that the empirical values are good estimates of the true population values which we do not know. Otherwise, it will be shown in Chapter X, Section B, that we can again apply the method of selection. That is to say, we can calculate the correlation coefficients of the differences of the two series in such a way that the coefficients become independent with two consecutive differences. We then can use Fisher's transformation in order to determine the difference beginning from which we can be reasonably sure that the correlations become stable.

If we make the further assumption that there is no correlation between the random and the nonrandom elements of either series, we can also calculate an approximation to the linear relationship between the mathematical expectations of the two time series. (Chapter X, Section C.)

CHAPTER II

HISTORY AND LITERATURE OF THE VARIATE DIFFERENCE METHOD

The following is based upon an account given by Professor Yule of the earlier history of the subject.[1] The first article using the variate difference method was by Professor J. H. Poynting.[2] It dealt with the analysis of wheat prices in several countries and cotton and silk imports into Great Britain. This paper was published in the year 1884. The next two papers were by Mr. R. H. Hooker[3] dealing, respectively, with the correlation of marriage with trade and the effect of the suspension of the Berlin Produce Exchange on prices. Both came out in 1901.

Miss F. E. Cave[4] published a paper on the correlation of barometric heights in the year 1904. There was a paper on corn prices by Mr. R. H. Hooker in 1905.[5] Dr. L. March,[6] the well-known French statistician, published in the same year an article on the numerical comparison between statistical curves.

These papers belong, according to Yule, to the first period of the development of the variate difference method. The later period is inaugurated by a paper by the famous "Student" (W. S. Gosset)[7] on the elimination of spurious correlation, which was published in *Biometrika* in the year 1914. He introduced the idea of correlating differences instead of the original series, in order to eliminate variability due to

[1] G. U. Yule, "On the Time Correlation Problem with Especial Reference to the Variate Difference Method," *Journal of the Royal Statistical Society*, Vol. 84, 1921, pp. 497 ff.

[2] J. H. Poynting, "A Comparison of the Fluctuations in the Price of Wheat and in the Cotton and Silk Imports into Great Britain," *Journal of the Royal Statistical Society*, Vol. 47, 1884, pp. 34 ff. See also his *Collected Scientific Papers*, Cambridge, 1920, pp. 506 ff.

[3] R. H. Hooker, "Correlation of the Marriage-Rate with Trade," *Journal of the Royal Statistical Society*, Vol. 64, 1901, pp. 485 ff.; "The Suspension of the Berlin Produce Exchange and its Effect upon Corn Prices," *Journal of the Royal Statistical Society*, Vol. 64, 1901, pp. 574 ff.

[4] F. E. Cave-Browne-Cave, "On the Influence of the Time Factor on the Correlation between the Barometric Heights at Stations More than 1000 Miles Apart," *Proceedings of the Royal Society of London*, Series A, Vol. 74, 1904, pp. 403 ff.

[5] R. H. Hooker, "On the Correlation of Successive Observations; Illustrated by Corn Prices," *Journal of the Royal Statistical Society*, Vol. 68, 1905, pp. 696 ff.

[6] L. March, "Comparaison numérique des courbes statistiques," *Journal de la Société de Statistique de Paris*, Vol. 46, 1905, pp. 255 ff., 306 ff.

[7] "Student," "The Elimination of Spurious Correlation Due to Position in Time or Space," *Biometrika*, Vol. 10, 1914, pp. 179 ff.

position in time and space. Professor O. Anderson[8] of Sofia also published the first of his papers on the variate difference method in the same year. He introduced the idea of a random series and, for the first time, that of standard errors, which enabled him to draw conclusions based on the theory of probability.

Miss B. M. Cave and Professor Karl Pearson[9] published a paper on numerical illustrations of the variate difference method in the same year. This is a very extensive study of 11 economic time series in Italy taken from a publication of Mortara. Miss Elderton and Karl Pearson,[10] in 1915, used the same method in a study on natural selection in man. The subject is mainly a correlation between sex and the death rate. Mr. A. Ritchie-Scott[11] discussed some correlation problems in 1915.

The first critical article on the variate difference method was published by Professor W. M. Persons[12] in 1917. He made a very extensive application of this analysis to 21 American economic time series. He came to a general conclusion that deviations from a calculated trend were more appropriate for his purposes than the study of differences. He also criticized the variate difference method as then used for not taking care of lag correlations.

The next critical approach came from Professor Yule[13] and was published in 1921. Whereas Yule agreed with the earlier writers on the variate difference method (Poynting, Hooker, Cave, and March), he disagreed with the subsequent contributors. The variate difference method is in his opinion fairly well fitted for the isolation of oscillations of different duration in order to study them separately. But it is not fitted for isolating random residuals and eliminating spurious correlations due to time. The variate difference method is also not able to deal with periodic oscillations, according to Professor Yule.

Miss Elderton and Professor Karl Pearson[14] answered the criti-

[8] O. Anderson, "Nochmals über 'The Elimination of Spurious Correlation Due to Position in Time or Space'," *Biometrika*, Vol. 10, 1914, pp. 269 ff.

[9] B. M. Cave and Karl Pearson, "Numerical Illustrations of the Variate Difference Correlation Method," *Biometrika*, Vol. 10, 1914, pp. 340 ff.

[10] E. M. Elderton and Karl Pearson, "Further Evidence of Natural Selection in Man," *Biometrika*, Vol. 10, 1915, pp. 488 ff.

[11] A. Ritchie-Scott, "Note on the Probable Error of the Coefficient of Correlation in the Variate Difference Correlation Method," *Biometrika*, Vol. 11, 1915, pp. 136 ff.

[12] W. M. Persons, "On the Variate Difference Correlation Method and Curve-Fitting," *Publications of the American Statistical Association*, Vol. 15, 1916-17, pp. 602 ff.

[13] G. U. Yule, *loc. cit.*

[14] Karl Pearson and E. M. Elderton, "On the Variate Difference Method," *Biometrika*, Vol. 14, 1922, pp. 281 ff. See also: K. Pearson, *Tables for Statisticians and Biometricians*, Vol. 2, London, 1931, pp. ccix ff., 235.

cisms of Persons and Yule in an article published in 1922. They conceded to Yule that very short periodic fluctuations are not eliminated by the process of finite differencing, but they stressed the point that long periodic fluctuations are eliminated. They also developed some methods to deal with Persons' criticism which was based on the existence of serial correlations between the subsequent items in the time series.

The second article of Professor Anderson appeared[15] in *Biometrika* in 1923. His most important contribution was the classification of time series as of three types: Zigzag or with strong negative correlation between consecutive items; Random with no correlation between consecutive items; and Smooth series with strong positive correlation between consecutive items. The variate difference method cannot deal with the first type, but proposes to separate the random part from the smooth part of a series. Professor Anderson also introduced serial correlations and studied the relations between the variances of the series of finite differences and the serial correlations of a random series. He gave certain criteria and evaluated standard errors. As an example Professor Anderson gave an experiment in random series (throwing of a coin).

Professor R. A. Fisher,[16] the eminent English statistician, dealt critically with the variate difference method in his celebrated paper on "The Influence of Rainfall on the Yield of Wheat at Rothamsted," published in 1925. His point of view is that, in general, curve fitting is to be preferred to the correlation of differences, since it introduces lower correlations between subsequent items. He very rightly points out the close connections between the variate difference method and Sheppard's smoothing formulae. He shows that the variate difference method, the application of Sheppard's smoothing formulae, and the fitting of polynomials are really based on the same fundamental idea.

The best mathematical exposition of the variate difference method is to be found in two articles published by Professor O. Anderson in *Biometrika* in 1926 and 1927.[17] Professor Anderson discussed the following problems: He first dealt with the general problem of the decomposition of a time series into its components. He further showed

[15] O. Anderson, "Ueber ein neues Verfahren bei Anwendung der 'Variate-Difference' Methode," *Biometrika,* Vol. 15, 1923, pp. 134 ff.

[16] R. A. Fisher, "The Influence of Rainfall on the Yield of Wheat at Rothamsted," *Philosophical Transactions of the Royal Society of London,* Series B, Vol. 213, 1925, pp. 89 ff., especially pp. 103 ff.

[17] O. Anderson, "Ueber die Anwendung der Differenzenmethode ('Variate Difference Method') bei Reihenausgleichungen, Stabilitätsuntersuchungen und Korrelationsmessungen," Part 1, *Biometrika,* Vol. 18, 1926, pp. 293 ff.; Part 2, *ibid.,* Vol. 19, 1927, pp. 53 ff.

the connection that exists between Sheppard's smoothing formulae and the fundamental ideas of the variate difference method. Standard errors were found for the variances of the series of finite differences. Professor Anderson also developed standard errors for the differences between the variances of two successive series of finite differences. These are very important for the practical application of his method. He dealt with the problem of serial correlations of differences and with product moments and their standard errors. He also introduced product moments of differences, which are very important for the study of correlations by the method of finite differencing. Professor Anderson dealt also with lag correlations and gave certain criteria for the different types of statistical series (*zigzag, random, smooth*) which he distinguishes. Those criteria were given in terms of the variances of the series of finite differences. He also investigated the dispersion and stability of series (Lexis) by the method of differences. The problem of serial correlations (Yule) also received some attention. He dealt with the problem of short periodic fluctuations and finally gave three numerical examples: a coin experiment as an example of random series, a study of the digits of the number π, and a correlation between wheat prices in Berlin, New York, and Chicago.

Professor A. L. Bowley[18] dealt with the variate difference method in his *Elements of Statistics*. He stressed the point that the data must be very precise in order to enable the calculation of higher differences. According to him, the variate difference method is too refined and too sensitive for ordinary statistical analysis.

Another article by Professor Anderson on the "Logic of Decomposition of Statistical Series into Separate Components"[19] was published in English in 1927. He gave an exposition of his hypotheses and methods and a criticism of the so-called Harvard method, which was designed for dealing with economic time series. He distinguished zigzag, random, and smooth series and stressed the connection of the variate difference method with Sheppard's smoothing formulae. He gave three numerical examples: one random series based on the frequency of certain letters in 36 printed lines; another dealing with indexes of market prices; and the third dealing with the Berlin and New York wheat quotations.

The most extensive exposition of the variate difference method is

[18] A. L. Bowley, *Elements of Statistics*, 4th ed., London, 1920, pp. 376 ff.

[19] O. Anderson, "On the Logic of Decomposition of Statistical Series into Separate Components," *Journal of the Royal Statistical Society*, Vol. 90, 1927, pp. 548 ff.

contained in Professor Anderson's German monograph,[20] which was published in 1929. He gave few mathematical proofs but he developed his argument at great length with special emphasis upon its economic applications. He gave a very thorough exposition of the difference analysis of time series and of the use of this method in correlation. He takes strictly the point of view of probability theory and presents the large-sample approach (standard errors). A mathematical appendix collects all the formulae which are pertinent to the argument. He also gave certain numerical examples: a coin experiment in order to show a random series, an analysis of egg prices, a correlation which is again based on a coin experiment, and the correlation between wheat prices in Berlin and New York. Another interesting and economically significant application of the variate difference method to economic time series was made by the same author in an article published in 1931.[21] It is an investigation of Bulgarian economic data and an attempt to verify the quantity theory of money.

The present author used the variate difference method rather extensively in his study, *Prices in the Trade Cycle*, which was published in 1935.[22] He did not apply it for correlation purposes but for the elimination of the random element in European prewar commodity prices and freight and interest rates. He treated 71 English prices, 83 German prices, 12 Dutch prices, 5 Austrian prices, and 3 Russian prices according to this method. The book also contains a short account and discussion of the principal assumptions and procedures of the variate difference method. A mathematical appendix gives a few formulae. The author also endeavored to make use of the results secured with the variate difference method when dealing with the further analysis of the nonrandom components of the time series.

Dr. A. Wald,[23] a Viennese mathematician (now at Columbia University), dealt with the variate difference method in his book on the calculation and elimination of seasonal variations (published in 1936). He claims that seasonal variations frequently show periodic movements of a short period. Hence the variate difference method could not be applied without modification to a series which contains considerable seasonal variation.

Dr. R. Zaycoff, a Bulgarian statistician, discussed Dr. Wald's

[20] O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, Veröffentlichungen der Frankfurter Gesellschaft für Konjunkturforschung, Heft 4, Bonn, 1929.

[21] O. Anderson, "Ist die Quantitätstheorie statistisch nachweisbar?" *Zeitschrift für Nationalökonomie*, Vol. 2, 1931, pp. 523 ff.

[22] G. Tintner, *Prices in the Trade Cycle*, Vienna, 1935.

[23] A. Wald, *Berechnung und Ausschaltung von Saisonschwankungen*, Vienna, 1936.

method in a number of publications and offered certain criticisms.[24] In an article published in 1937[25] he gave two criteria for deciding, by the variate difference method, at which difference the mathematical expectation is eliminated. He claimed that those criteria hold true even for short series and high differences. He also contended that it is sometimes necessary to subdivide an economic time series since it shows different behavior in its random element in different parts. He gave a criterion for dealing with short periodic fluctuations.

The English statistician M. S. Bartlett dealt with "Some Aspects of the Time Correlation Problem in Regard to Tests of Significance"[26] in 1935. He treated the problem of serial correlations and estimated the efficiency of the variate difference method. This seems to be the first attempt to use tests of significance in this subject.

The author proposed a new method of selection in a recent article[27] which gives exact tests of significance but sacrifices some available information.

[24] R. Zaycoff, "Ausschaltung der Saisonkomponente nach der Methode von Dr. A. Wald," *Publications of the Statistical Institute for Economic Research*, State University of Sofia, 1935, No. 2-3, pp. 263 ff., 274; "Ueber die Zerlegung statistischer Zeitreihen in drei Komponenten," *ibid.*, 1936, No. 4, pp. 141 ff.

[25] R. Zaycoff, "Ueber die Ausschaltung der zufälligen Komponente nach der 'Variate-Difference' Methode," *ibid.*, 1937, No. 1, pp. 75 ff.

[26] M. S. Bartlett, "Some Aspects of the Time-Correlation Problem in Regard to Tests of Significance," *Journal of the Royal Statistical Society*, Vol. 98, 1935, pp. 536 ff.

[27] G. Tintner, "On Tests of Significance in Time Series," *Annals of Mathematical Statistics*, Vol. 10, 1939, pp. 139 ff.

CHAPTER III

Criticism of the Variate Difference Method

It has been pointed out in the foregoing chapter that the criticism of the variate difference method centers more or less around the following points: (1) the accuracy of higher differences (A. Bowley); (2) serial correlations (W. M. Persons, R. A. Fisher, M. S. Bartlett); (3) short periodic oscillations (G. U. Yule, A. Wald). We propose to deal with these different points in turn.

A. The Accuracy of Higher Differences

This point was first raised by Professor Arthur L. Bowley in his *Elements of Statistics*.[1] His view is that errors that occur in the data are magnified in the differences and become larger with the higher orders of the differences. This is true but it is pertinent only in so far as the individual differences are concerned and not for all parameters or statistics derived from them. The variances (squares of the standard deviations) of a difference series should, for instance, be divided by a binomial coefficient in order to get an estimate of the true variance of the random element.[2] This divisor is so great that it would take care, for all practical purposes, of those small inaccuracies in the original data which, however, are magnified in the higher differences. This binomial coefficient is, for the kth difference, equal to the number of combinations of $2k$ things, taken k at a time. To give an example, the binomial coefficient for the fifth difference is 252, for the sixth difference 924, etc. (Table 9). It increases rapidly with the order of the difference. Professor Bowley is perfectly correct in stating that we should not trust our individual differences of higher order too much. But I do not think he is justified in so far as certain statistical parameters derived from higher differences are concerned.

B. Serial Correlation

The problem of serial correlation was brought into the discussion of the variate difference method by the late Professor W. M. Persons

[1] A. L. Bowley, *Elements of Statistics*, 4th ed., London, 1920, pp. 376 ff., 388 ff.
[2] See for instance O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, Bonn, 1929, pp. 54 ff., 111 ff. Also below, Chapter V.

in his critical article of 1917.[3] A similar argument was later put forward by Professor R. A. Fisher[4] and M. S. Bartlett.[5]

We understand by serial correlations the correlations between successive items of the same series, according to a definition given by Professor Yule[6] in his celebrated article "Why Do We Sometimes Get Nonsense-Correlations Between Time Series?" where he gave the first complete treatment of this problem. For instance, we have a series with the general term w_i and we correlate with every term w_i the one that immediately follows it, w_{i+1}—that is, we correlate the original series with the series we get by shifting it by one item. This would give us an example of a serial correlation.

Those correlations are very important in the treatment of time series.[7] It is quite evident that a series that shows very strong serial correlations cannot be treated as a random series. Its consecutive items are not independent. Hence, a great number of propositions of the theory of probability are certainly not applicable in such a case. For instance any correlation found between this series and any other series is very much devoid of meaning, in terms of probabilities. These difficulties, however, can sometimes be overcome by making a proper selection of the items from which the correlation is calculated. This involves some loss of information but will enable us to interpret the correlation coefficients statistically in terms of probabilities (Chapter X, Section B).

It is the contention of Professor Persons[8] that the variate difference method becomes nonapplicable in the case where there are serial correlations between two different items of the same series. He has, however, already been refuted by Miss Elderton and Professor Karl Pearson in 1922.[9] In their article these authors proposed a number of

[3] W. M. Persons, "On the Variate Difference Correlation Method and Curve-Fitting," *Publications of the American Statistical Association,* Vol. 15, 1916-17, pp. 602 ff.

[4] R. A. Fisher, "The Influence of Rainfall on the Yield of Wheat at Rothamsted," *Philosophical Transactions of the Royal Society of London,* Series B, Vol. 213, pp. 103 ff.

[5] M. S. Bartlett, "Some Aspects of the Time-Correlation Problem in Regard to Tests of Significance," *Journal of the Royal Statistical Society,* Vol. 98, 1935, pp. 536 ff.

[6] G. U. Yule, "Why Do We Sometimes Get Nonsense-Correlations between Time-Series?" *Journal of the Royal Statistical Society,* Vol. 89, 1926, pp. 1 ff.

[7] H. T. Davis, "The Econometric Problem" in Cowles Commission for Research in Economics, *Report of Third Annual Research Conference on Economics and Statistics,* Colorado Springs, 1937, pp. 11 ff.; H. Wold, *A Study in the Analysis of Stationary Time Series,* Uppsala, 1938; see also Appendix V, Section B.

[8] W. M. Persons, *loc. cit.*

[9] Karl Pearson and E. M. Elderton, "On the Variate Difference Method," *Biometrika,* Vol. 14, 1922, pp. 281 ff.; K. Pearson, *Tables for Statisticians and Biometricians,* Vol. 2, London, 1931, pp. ccix ff., 235.

new criteria to take care of serial correlations.

Professor Anderson has developed this theory most fully in his article in *Biometrika*[10] and in his monograph.[11] He has shown that there exists a very definite connection between the variances of the series of finite differences and the serial correlations of a statistical series. Professor Anderson, as well as Professor Karl Pearson and Miss Elderton, gives criteria in terms of finite differences by which we can readily detect and deal with serial correlations (see Appendix V, Section B).

A similar criticism was put forward by Professor R. A. Fisher[12] from an entirely different point of view. In his celebrated article on "The Influence of Rainfall on the Yield of Wheat at Rothamsted," in 1925, Professor Fisher also deals critically with the variate difference method. He contends that repeated finite differencing of a random series introduces artificial correlations between the consecutive items and hence diminishes the usefulness of the series of differences. This, of course, is a criticism which is entirely justified. If we have a random series in which every item is entirely independent of every other, the first differences of this series will introduce serial correlations between two consecutive items, the second differences will introduce correlations between three consecutive items, etc. This is certainly a great shortcoming of the variate difference method. But it is, in our opinion, more than balanced by advantages which this method possesses especially from an economic point of view. It is here more appropriate than the procedure of fitting curves, especially polynomials, which Professor Fisher proposes.[13] The latter method may be entirely justified and very well applicable in the natural sciences as, for instance, biology and meteorology with which Professor Fisher is dealing in the article quoted, but it is in our opinion too rigid for application to economic quantities.[13a]

The decision depends essentially on the point of view that one takes about the nature of economic data and economic phenomena in

[10] O. Anderson, "Ueber die Anwendung der Differenzenmethode ('Variate Difference Method') bei Reihenausgleichungen, Stabilitätsuntersuchungen, und Korrelationsmessungen," Part 1, *Biometrika*, Vol. 18, 1926, pp. 315 ff.

[11] O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, pp. 61 ff., 114 ff.

[12] R. A. Fisher, *loc. cit.*

[13] R. A. Fisher, *Statistical Methods for Research Workers*, 7th ed., London, 1938, pp. 148 ff., Chapter V, Section 27; see also M. Sasuly, *Trend Analysis of Statistics*, Washington, 1934.

[13a] The need for flexibility in the analysis of economic time series was also stressed by R. Frisch, "A Method of Decomposing an Empirical Series into Its Cyclical and Progressive Components," *Journal of the American Statistical Association*, March, 1931, Supplement, pp. 73 ff.

general. If we could contend that we find the same rigid and invariable laws in social phenomena which the astronomers and physicists or natural scientists in general have been able to detect, we should always fit polynomials or other appropriate functions. Unfortunately, the situation is quite different in economics. Economic "laws," whatever the definite meaning of this term may be, are not fixed in time but change in different periods of time.[14] So, for instance, a demand curve, or a demand surface, is hardly likely to stay the same during a number of years. To give another example which is even more important in connection with economic time series: the conditions for a specific time series during the boom are probably different from those that are prevailing in the slump.[15] It is more than likely that economic phenomena in an expanding economic system are entirely different from those in a stationary or declining system.[16] Hence, we need in economics a very flexible method for dealing with time series and we shall have to put up with the serial correlations that the variate difference method introduces into our series. Apart from the serial correlations, the variate difference method and the method of smoothing are much more flexible. Hence, in spite of this shortcoming, they are much less likely to introduce errors systematically than any other method, especially the fitting of polynomials. We may in some cases be able to eliminate the trend by the fitting of orthogonal polynomials. But this method does not seem to be applicable in series which contain business cycles.

The author proposes in a later part of this book[17] a method which eliminates, at least to a certain extent, those difficulties. It simply consists in making a selection of entirely uncorrelated differences from a time series. But it is evident that the number of uncorrelated first differences is only one-half of the number originally contained in the series, and the number of uncorrelated second differences is only one-third, etc. We shall reduce our time series very greatly by this method. Professor R. A. Fisher's and Mr. Bartlett's criticisms still hold true. There are, nevertheless, cases where it may be advantageous to make these selections. In any case, the method gives a significant estimate. It really enables us to make an exact test of significance and to use the methods of the analysis of variance that have been so well developed by Professor Fisher himself.

[14] L. Robbins, *An Essay on The Nature and Significance of Economic Science*, 2nd ed., London, 1937, pp. 54 ff., 79 ff.

[15] G. Haberler, *Prosperity and Depression*, New Edition, Geneva, 1939, pp. 257 ff.

[16] J. A. Schumpeter, *The Theory of Economic Development*, Cambridge, Mass., 1934.

[17] See Chapters VIII and X, Section B.

C. Periodic Oscillations

Professor Yule[18] and Dr. Wald[19] have objected that the occurrence of periodic fluctuations with a short period renders the application of the variate difference method impossible. It cannot be denied that this is true, since such periodic fluctuations are not "smooth" in the sense defined above, but are of the extreme zigzag nature which we mentioned before. They show strong negative correlations between neighbouring terms. Hence, they do not belong to the category of fluctuations that can be eliminated or at least reduced by finite differencing.

It is very unlikely, however, that fluctuations of this character occur frequently in economic time series. One such case has been described by Holbrook Working and Harold Hotelling.[19a] Professor Anderson[20] and Dr. Zaycoff[21] have more or less refuted Dr. Wald's assertions, which pertain especially to the seasonal fluctuations.

We have performed some experiments in order to test the validity of those objections (see Appendix III). Isolated examples can, of course, never be conclusive. The example given below is no argument for the absence in all series of these oscillatory fluctuations which may prevent the successful application of the variate difference method. Dr. Zaycoff gives a criterion[22] by which we can determine whether this is the case. The present author, however, in dealing with prices has never found an example in which these oscillatory fluctuations occurred.

It is nevertheless better, perhaps, to follow Dr. Wald's advice[23] and to eliminate the seasonal variation before starting the difference analysis. This has been done in the case of the commodity analyzed below. The simplest way is to approximate the seasonal by the average monthly values, as we have done (Appendix III). The seasonal

[18] G. U. Yule, "On the Time-Correlation Problem with Especial Reference to the Variate-Difference Method," *Journal of the Royal Statistical Society,* Vol. 84, 1921, pp. 497 ff.

[19] A. Wald, *Berechnung und Ausschaltung von Saisonschwankungen,* Vienna, 1936, pp. 37 ff. See also H. Mendershausen, "Annual Survey of Statistical Technique: Methods of Computing and Eliminating Changing Seasonal Fluctuations," *Econometrica,* Vol. 5, 1937, pp. 242 ff.

[19a] H. Working and H. Hotelling, "The Application of the Theory of Errors to the Interpretation of Trends," *Proceedings of the American Statistical Association,* March, 1929, pp. 73 ff. especially p. 77.

[20] See O. Anderson's review of Dr. Wald's book in *Zeitschrift für Nationalökonomie,* Vol. 8, 1937, pp. 251 ff.

[21] R. Zaycoff, "Ueber die Ausschaltung der zufälligen Komponente nach der 'Variate-Difference' Methode," *Publications of the Statistical Institute for Economic Research,* State University of Sofia, 1937, No. 1, pp. 100 ff.

[22] R. Zaycoff, *op. cit.,* pp. 104 ff.

[23] A. Wald, *op. cit.,* pp. 37 ff., 73.

is then differenced and by squaring the differences and summing them up we get an approximation to the portion of the variance of the difference that is due to the seasonal. Deducting this portion from the total variance we have an estimate of the variance due to the trend, cycle, and random element alone. The further analysis of the variances of finite differences corrected for the seasonal will, however, not differ very greatly from the ordinary procedure and will yield, generally, more or less the same results.

CHAPTER IV

Fundamental Concepts

A. The Definition of Probability

At this point a brief exposition of some of the more recent ideas on the definition of probability seems in order. We do not believe, however, that it makes a great deal of difference for the application in economic statistics from which definition of probability we start.

The older definition of probability, which may be called classical and which goes back to Laplace, was based on the concept of equal likelihood.[1] The probability of an event is defined as the ratio between the number of favorable cases and the number of all equally possible cases. But there are certain objections to the classical definition of probability. The most important one from our point of view is that it is in general impossible, in the applications of probability to economics, to know a priori which events are equally possible.[2]

There are two main tendencies which try to get away from the classical definition of probability. The first may be called the formalistic one.[3] It is entirely sufficient from the point of view of the pure mathematical theory of probability. The concept of probability is here introduced as an undefined fundamental concept and all the propositions of the classical theory of probability follow without the necessity of adopting the classical definition.

The third point of view, which more recently is especially associated with the name of von Mises, is the so-called frequency definition of probability.[4] It seems to us to be better suited for our specific pur-

[1] P. Laplace, *Essai philosophique sur les probabilitiés*, 4th ed., Paris, 1819; English edition, New York, 1902.

[2] R. von Mises and H. Pollaczek-Geiringer, "Probability," in *Encyclopaedia of the Social Sciences*, New York, 1935. See also O. Anderson, *Einführung in die Mathematische Statistik*, Vienna, 1935, pp. 19 ff.

[3] See especially A. Kolmogoroff, "Grundbegriffe der Wahrscheinlichkeitsrechnung" in *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Vol. 2, No. 3, Berlin, 1933; H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts in Mathematics and Mathematical Physics, No. 36, Cambridge, 1937.

[4] R. von Mises, *loc. cit.*; see also *Probability, Statistics and Truth*, London, 1939; *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und theoretischen Physik*, Vienna, 1931; H. Reichenbach, *Wahrscheinlichkeitslehre*, Leiden, 1935, pp. 79 ff.; A. Wald, "Die Widerspruchsfreiheit des Kollektivbegriffs der Wahrscheinlichkeitsrechnung" in *Ergebnisse eines mathematischen Kolloquiums*, ed. Karl Menger, Heft 8, 1935-36, pp. 38 ff.; A. Wald, "Die Widerspruchsfreiheit

poses, since here probability is an empirical concept. The frequency definition of probability is as follows:

We have first to assume a "collective," which is a well-defined set of individual events that have certain distinguishing characteristics. The probability is defined as the limit to which the relative frequency of the number of individuals possessing the characteristic in which we happen to be interested tends as the total number of individuals contained in the sample becomes larger and larger. Hence, probability is simply the limiting value of the relative frequency (an empirical concept) as our series of trials becomes infinite. (See also below, Appendix II, pp. 136 ff.)

One of the most important properties of the collective is the following: Suppose we make a certain selection of items from the original series, which selection, of course, should be independent of the characteristic. The new sample thus formed must show the same probability as the old one. To give an example: let us assume that our collective is the number of heads secured by tossing a coin, and that, the larger the number of trials, the nearer our relative frequency of heads will tend towards the probability ½. This will happen only if our coin is true or unbiased. If it should be biased, the limit will be different from ½. Suppose now that we number every one of our trials. We simply attach the number 1 to the first trial, 2 to the second, and so on. We make any selection from these numbers: for instance, we take all the even throws or we take all the throws whose number is a prime, or we take all the throws that come after a head has been thrown, or we take all the throws that come after a succession of two tails, etc. In all those new series formed from our old collective the relative frequency of heads must ultimately tend towards ½ if the number of trials increases indefinitely just as in the original collective.

An example given by von Mises[5] will help to clarify this concept. Suppose I am traveling along a road and I am counting the milestones. There is a big milestone at every mile and there are smaller markers at every quarter mile. If I consider the milestones as my collective, then the relative frequency of big milestones tends towards ¼ if I make my sample big enough. But this relative frequency certainly is not the same if I choose every fourth milestone as a new collective.

des Kollektivbegriffes," in M. Fréchet, *Colloque Consacré a la Théorie des Probabilités*, 2me partie, Paris, 1938 (*Actualités Scientifiques et Industrielles*, no. 735), pp. 79 ff. See also the excellent account of these ideas in E. Nagel, "Principles of the Theory of Probability" in *International Encyclopaedia of Unified Science*, Vol. 1, No. 6, Chicago, 1939.

[5] R. von Mises, *Probability, Statistics and Truth*, pp. 30 f.

Then the relative frequency of large milestones will easily be seen to be equal to 1 if I happen to start with a large milestone, or 0 if I happen to start with a small milestone. Its limit certainly will not be equal to ¼, as in the original collective. Next I take as my new collective all milestones that come after a big milestone. The relative frequency of large milestones in the new collective will be equal to 0 and again will not, in the limit, be equal to the relative frequency of large milestones in the original collective. Hence, my collection of milestones is not really a true collective according to von Mises and the concept of probability can certainly not be used in this connection.

If we start with this frequency definition of probability then all the classical propositions in the theory of probability can still be used. We propose, however, to make use in this connection of certain concepts which ultimately go back to the classical theory of probability but which have been especially developed by the Russian school of statisticians. In our opinion they are more suited to the treatment of our problems than are the other concepts. They are based on the ideas of the random or casual variable, of the distribution, and of the mathematical expectation.

B. Random Variable, Distribution, and Mathematical Expectation

It will be possible to give in this connection only a very short indication of some fundamental concepts. The reader can easily find more detailed information on this subject in other books.[6] Our point of view is essentially the one which has been called "stochastic."[7]

We understand a casual or random or stochastic variable[8] to be a magnitude that can have a series of different values with definite probabilities. Those probabilities, as we have indicated above, are simply limiting values of the relative frequencies of the variate in a collective. To give a simple example: Suppose that we throw a die and take as our characteristic the numbers 1 to 6. Those numbers are a casual or random variable if it is a true die or even if it is a biased die. Their probabilities are simply the limiting values of their relative frequencies. In this particular case the probabilities will be 1/6 for every number in a true die, and different from this in a biased die.

The values that the random or stochastic variable can take togeth-

[6] J. V. Uspensky, *Introduction to Mathematical Probability*, New York, 1937. See also below, Appendix II, pp. 136 ff.

[7] L. von Bortkiewicz, *Die Iterationen*, Berlin, 1917. See also O. Anderson, *op. cit.*, and J. V. Uspensky, *op. cit.*

[8] J. V. Uspensky, *op. cit.*, pp. 161 ff.; W. Winkler, *Theoretische Statistik*, Berlin, 1931, pp. 20 ff.

er with their probabilities we shall call the distribution of the casual or random variable. In our previous example, with the true die the distribution will consist of the numbers 1 to 6 where the probability of 1/6 is attached to every number.

The mathematical expectation,[9] finally, is simply the sum of the products of the values of the random variable and their respective probabilities. The sum of those probabilities is, of course, always equal to 1. In the case of our true die, again, the mathematical expectation is: $E(x) = 1 \cdot (1/6) + 2 \cdot (1/6) + 3 \cdot (1/6) + 4 \cdot (1/6) + 5 \cdot (1/6) + 6 \cdot (1/6) = 3.5$. The mathematical expectation is, in other words, the weighted arithmetic mean of the values of our casual variable with the probabilities as weights. This is also the reason why it has been called mean value (*valeur moyenne*) by some French statisticians.[10]

Another concept which is of importance in this connection is the population variance. It is the mathematical expectation of the square of the deviations from the mathematical expectation. We have $\sigma^2 = E[x - E(x)]^2$. In our example we have $E(x) = 3.5$ and $\sigma^2 = (1 - 3.5)^2(1/6) + (2 - 3.5)^2(1/6) + (3 - 3.5)^2(1/6) + (4 - 3.5)^2(1/6) + (5 - 3.5)^2(1/6) + (6 - 3.5)^2(1/6) = 2.9167$. Hence the population variance is 2.9167 and the population standard deviation, the square root of the variance, is 1.7078.

We should always distinguish the values of statistical parameters calculated for the (hypothetically infinite) population and their statistical estimates calculated from the sample. The theorems of probability are strictly true only for the former kind. We know, however, from Bernoulli's theorem[11] that we can expect (with a probability as near to certainty as we like) that the relative frequency of an event in a series of independent trials with constant probability will differ from this probability as little as we like, if the number of trials becomes sufficiently large. The law of large numbers states: We can expect with a probability as near to certainty as we please, that the arithmetic mean of a number of stochastic variables with the same mathematical expectation will differ from this mathematical expectation by less than any given number, however small, if the number of variables becomes large enough and another condition is fulfilled.[12]

C. Finite Differences

Only one other fundamental mathematical concept is necessary

[9] J. V. Uspensky, *op. cit.*, pp. 161 ff.
[10] G. Darmois, *Statistique Mathématique*, Paris, 1928, pp. 30 ff.
[11] J. V. Uspensky, *op. cit.*, pp. 96 ff.
[12] J. V. Uspensky, *op. cit.*, pp. 182 ff.

for the treatment of the variate difference method. This is the idea of finite differences.[13] We attach numbers to every item in our time series, for instance, the natural numbers 1, 2, 3, etc. Let us denote the ith item ($i = 1, 2, 3, \cdots, N$) by w_i. Then the first finite difference $\Delta^{(1)}w_i = w_{i+1} - w_i$. This will be the difference between the $(i+1)$th item and the ith item. We form the series of second differences by performing the same operation on the series of first differences and so on for any order of differences. This idea will be made clearer in Table 1.

TABLE 1

DIFFERENCES OF THE SQUARES OF NUMBERS

$w_i = i^2$

i	w_i	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$
1	1	3	2	0
2	4	5	2	0
3	9	7	2	0
4	16	9	2	0
5	25	11	2	0
6	36	13	2	0
7	49	15	2	0
8	64	17	2	0
9	81	19	2	0
10	100	21	2	......
11	121	23	......	......
12	144	......	......	......

Suppose those 12 items are our time series w. The number of every item is given in the first column and the item itself is given in the second column. The third column shows the series of first finite differences ($\Delta^{(1)}$) which is, of course, formed by subtracting from every item in the second column the one immediately above. In the fourth column is the series of second differences ($\Delta^{(2)}$). This series is again formed by subtracting from every item in the third column the one immediately above. In the fifth column we have the series of the third finite differences ($\Delta^{(3)}$). It is again formed by subtracting from every item in the fourth column the one immediately above.

There is another thing which can immediately be seen from this table. We have chosen in our particular example as our time series w_i the squares of the natural numbers as shown in the first column. We

[13] See, e.g., G. Boole, *A Treatise on the Calculus of Finite Differences*, 3rd ed., London, 1880; H. T. Davis and W. F. C. Nelson, *Elements of Statistics*, 2nd ed., Bloomington, Indiana, 1937, pp. 216 ff.; H. T. Davis, *Theory of Linear Operators*, Bloomington, Indiana, 1936, pp. 85 ff. See also below, Appendix II, pp. 137 ff.

note from our table in the fourth column that the second finite differences ($\Delta^{(2)}$) are stable (in our case equal to 2) and we see from the fifth column that the third finite differences ($\Delta^{(3)}$) are always equal to zero. The same would be true, of course, of the higher differences.

If we had chosen for our series w_i the cubes of the natural numbers, the third finite differences would have been constant and the fourth and all higher differences would have been equal to zero. Had we, for instance, chosen the fourth powers of the natural numbers as our series w_i, the fourth finite difference would have been constant and all differences higher than the fourth would have been equal to zero. Generalizing this idea we can say: If our series is equal to the nth power of the natural numbers, then the nth finite difference ($\Delta^{(n)}$) will be constant and all higher finite differences will be equal to zero.

TABLE 2

DIFFERENCES OF A POLYNOMIAL

$w_i = 1 + 2i^2 - i^3 + 3i^4$

i	w_i	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$	$\Delta^{(4)}$	$\Delta^{(5)}$
1	5	44	142	174	72	0
2	49	186	316	246	72	0
3	235	502	562	318	72	0
4	737	1,064	880	390	72	0
5	1,801	1,944	1,270	462	72	0
6	3,745	3,214	1,732	534	72	0
7	6,959	4,946	2,266	606	72	0
8	11,905	7,212	2,872	678	72	..
9	19,117	10,084	3,550	750	....	..
10	29,201	13,634	4,300	......	....	..
11	42,835	17,934	....	......	....	..
12	60,769	....	....	......	....	..

Let us now consider a polynomial of the degree n. As seen from Table 2, the series of the nth finite differences is constant and all differences of order higher than n will be zero. Hence, any polynomial of the degree n can be completely eliminated by the process of finite differencing if carried far enough.

A similar statement holds true for other "smooth" functions—that is to say, for functions which do not show an excessively zigzag nature.[14] Smooth functions have a strong positive correlation between consecutive items. It has been shown by Professor Anderson[15] that

[14] O. Anderson, "On the Logic of Decomposition of Statistical Series into Separate Components," *Journal of the Royal Statistical Society*, Vol. 90, 1927, pp. 548 ff.; *Die Korrelationsrechnung in der Konjunkturforschung*, Bonn, 1929, pp. 46 ff.

[15] O. Anderson, *op. cit.*, pp. 107 ff.

exponential functions or hyperbolae or trigonometric functions with a long period cannot be completely eliminated but only reduced to any desired degree by finite differencing. The only classes of functions for which this is not the case are those showing strong negative correlation between subsequent items or the extreme zigzag shape that we have mentioned above.

TABLE 3

DIFFERENCES OF AN EXPONENTIAL

$w_i = 2^i$

i	w_i	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$	$\Delta^{(4)}$	$\Delta^{(5)}$	$\Delta^{(6)}$	$\Delta^{(7)}$	$\Delta^{(8)}$	$\Delta^{(9)}$	$\Delta^{(10)}$	$\Delta^{(11)}$
1	2	2	2	2	2	2	2	2	2	2	2	2
2	4	4	4	4	4	4	4	4	4	4	4	......
3	8	8	8	8	8	8	8	8	8	8	......	......
4	16	16	16	16	16	16	16	16	16	......	......	......
5	32	32	32	32	32	32	32	32	......	......	......	......
6	64	64	64	64	64	64	64	......	......	......	......	......
7	128	128	128	128	128	128	......	......	......	......	......	......
8	256	256	256	256	256	......	......	......	......	......	......	......
9	512	512	512	512	......	......	......	......	......	......	......	......
10	1,024	1,024	1,024	......	......	......	......	......	......	......	......	......
11	2,048	2,048	......	......	......	......	......	......	......	......	......	......
12	4,096	......	......	......	......	......	......	......	......	......	......	......

We show a few examples of the process of elimination or reduction in the tables. Table 2 shows a polynomial of the fourth degree whose fifth differences become zero and all higher differences vanish. In Table 3 we have a simple exponential which cannot be completely reduced. Even in this disadvantageous case of an exponential whose base is positive and greater than 1, the highest values occurring in every series of differences decrease with higher order of the differences. The difference of a given order of any term is smaller than the difference of the preceding order of the term following. Take, for instance, the eighth difference of the third term, which is 8. It is smaller than the seventh difference of the fourth term, which is 16.

Table 4 shows reduction of a hyperbola by taking successive finite differences. It is seen that the absolute value of the differences (i.e., the value without regard to sign) decreases with the increasing order of the difference. The same phenomenon is exhibited in Table 5 for a trigonometric function of long period. The period, i.e., the distance from maximum to maximum or minimum to minimum, is here 12 units. Variations of all these "smooth" functions can hence be greatly reduced by taking finite differences.

TABLE 4

DIFFERENCES OF A HYPERBOLA

$$w_i = \frac{100}{i}$$

i	w_i	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$	$\Delta^{(4)}$	$\Delta^{(5)}$	$\Delta^{(6)}$	$\Delta^{(7)}$	$\Delta^{(8)}$	$\Delta^{(9)}$	$\Delta^{(10)}$	$\Delta^{(11)}$
1	100.000	–50.000	33.333	–24.999	19.998	–16.663	14.279	–12.488	11.091	–9.969	9.047	–8.276
2	50.000	–16.667	8.334	– 5.001	3.335	– 2.384	1.791	– 1.397	1.122	–0.922	0.771	....
3	33.333	– 8.333	3.333	– 1.666	0.951	– 0.593	0.394	– 0.275	0.200	–0.151	....	....
4	25.000	– 5.000	1.667	– 0.715	0.358	– 0.199	0.119	– 0.075	0.049	....	....	....
5	20.000	– 3.333	0.952	– 0.357	0.159	– 0.080	0.044	– 0.026	....	....	....	....
6	16.667	– 2.381	0.595	– 0.198	0.079	– 0.036	0.018	....	....	....	....	....
7	14.286	– 1.786	0.397	– 0.119	0.043	– 0.018	....	....	....	....	....	....
8	12.500	– 1.389	0.278	– 0.076	0.025	....	....	....	....	....	....	....
9	11.111	– 1.111	0.202	– 0.051	....	....	....	....	....	....	....	....
10	10.000	– 0.909	0.151	....	....	....	....	....	....	....	....	....
11	9.091	– 0.758	....	....	....	....	....	....	....	....	....	....
12	8.333	....	....	....	....	....	....	....	....	....	....	....

TABLE 5

DIFFERENCES OF A TRIGONOMETRIC FUNCTION WITH LONG PERIOD

$$w_i = \sin \frac{2\pi}{12} i$$

i	w_i	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$	$\Delta^{(4)}$	$\Delta^{(5)}$	$\Delta^{(6)}$	$\Delta^{(7)}$	$\Delta^{(8)}$	$\Delta^{(9)}$	$\Delta^{(10)}$	$\Delta^{(11)}$
1	0.500	0.366	–0.232	–0.036	0.072	–0.010	–0.016	0.006	0.004	–0.004	0.000	0.006
2	0.866	0.134	–0.268	0.036	0.062	–0.026	–0.010	0.010	0.000	–0.004	0.006	....
3	1.000	–0.134	–0.232	0.098	0.036	–0.036	0.000	0.010	–0.004	0.002	....	....
4	0.866	–0.366	–0.134	0.134	0.000	–0.036	0.010	0.006	–0.002	....	....	....
5	0.500	–0.500	0.000	0.134	–0.036	–0.026	0.016	0.004	....	....	....	....
6	0.000	–0.500	0.134	0.098	–0.062	–0.010	0.020	....	....	....	....	....
7	–0.500	–0.366	0.232	0.036	–0.072	0.010	....	....	....	....	....	....
8	–0.866	–0.134	0.268	–0.036	–0.062	....	....	....	....	....	....	....
9	–1.000	0.134	0.232	–0.098	....	....	....	....	....	....	....	....
10	–0.866	0.366	0.134	....	....	....	....	....	....	....	....	....
11	–0.500	0.500	....	....	....	....	....	....	....	....	....	....
12	0.000	....	....	....	....	....	....	....	....	....	....	....

TABLE 6

DIFFERENCES OF A TRIGONOMETRIC FUNCTION WITH SHORT PERIOD

$$w_i = \sin \frac{2\pi}{4} i$$

i	w_i	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$	$\Delta^{(4)}$	$\Delta^{(5)}$	$\Delta^{(6)}$	$\Delta^{(7)}$	$\Delta^{(8)}$	$\Delta^{(9)}$	$\Delta^{(10)}$	$\Delta^{(11)}$
1	1	–1	0	2	–4	4	0	–8	16	–16	0	32
2	0	–1	2	–2	0	4	–8	8	0	–16	32	....
3	–1	1	0	–2	4	–4	0	8	–16	16	....	....
4	0	1	–2	2	0	–4	8	–8	0	....	....	....
5	1	–1	0	2	–4	4	0	–8	....	....	....	....
6	0	–1	2	–2	0	4	–8	....	....	....	....	....
7	–1	1	0	–2	4	–4	....	....	....	....	....	....
8	0	1	–2	2	0	....	....	....	....	....	....	....
9	1	–1	0	2	....	....	....	....	....	....	....	....
10	0	–1	2	....	....	....	....	....	....	....	....	....
11	–1	1	....	....	....	....	....	....	....	....	....	....
12	0	....	....	....	....	....	....	....	....	....	....	....

But the same is not true with functions of an extreme zigzag nature, as for instance trigonometric functions with a small period. This is shown in Table 6, the period being 4 units. The differences are seen to increase in their numerical or absolute value, the higher the order of difference. Hence we can not hope to eliminate or even reduce this and similar functions by taking finite differences.

Whereas smooth functions can be eliminated or at least reduced indefinitely by finite differencing, finite differencing has, of course, no such effect at all on a random series. A true random series is not ordered in time and hence will not be affected by the process of finite differencing. This different behavior of smooth curves (which represent the mathematical expectation) and random series (which represent the random element of our time series) is the starting point for the variate difference method.

Finite differencing will not eliminate the random element but will affect it in some other way. The forming of differences introduces certain serial correlations into the series of differences even if the items in the original random series were entirely independent. But it is very easy to see what this effect is and to calculate it in advance. Our formulae for the standard errors and the estimates of the population variance of the random element from the variances of the differences take account of this fact. The correlations between the consecutive items of the series of finite differences necessitate our making selections from these series if we want to apply the modern tests of significance (Chapter VIII and Chapter X, Section B).

CHAPTER V

The Calculation of the Variances of the Finite Difference Series

To recapitulate briefly, the fundamental idea of the variate difference method is the following:

We begin with the assumption that the economic time series in question consists of two parts: a "smooth" part which we shall call the mathematical expectation and which we assume to be the result of the more permanent economic and social factors, and a random part which is random in the sense defined in Chapter I, Section C. It may be considered the result of the nonpermanent or less permanent factors in economic life. It is clear that the question of what is permanent and what is not depends on the time unit. The random element will be different in series of annual data from what it is, for instance, in a series of monthly data. We have to use a concept that is very similar to Alfred Marshall's[1] ideas of long run and short run in economic life (see also below, p. 72).

We assume that the mathematical expectation and the random element are connected by addition. The variance of the entire series can hence be split up into one part which comes from the mathematical expectation and another part which is the random variance. (The variance is the square of the standard deviation.) It has already been pointed out (see Chapter IV, Section C) that the smooth part or the mathematical expectation can be eliminated, or at least reduced to any desired degree, by successive finite differencing. Hence the part of the variance that results from the mathematical expectation can be reduced in the same fashion. This is, however, true only for a "smooth" and not for an excessively "zigzag" series.

The variance of the finite difference of any order is again equal to the sum of the variance of the difference of the mathematical expectation and the variance of the difference of the random element. The first component is reduced more and more in a smooth series by successive finite differencing, the second component is not changed at all apart from the multiplication by certain constants.

We ask ourselves the following question: Beginning from which finite difference k_0 can we assume that the mathematical expectation

[1] A. Marshall, *Principles of Economics*, 8th ed. London, 1920, pp. 378 ff.

has been eliminated to a considerable degree and that we are left approximately with only the random element?[2] The solution of this problem results from the following consideration: If we have a series which consists only of the random element, then the variances of the successive series of finite differences are equal, if corrected for the multiplication by a binomial coefficient. This is self-evident from the fact that the series is random, that is to say, is not ordered in time. Hence the variance of its first and higher differences must be the same as the variance of the original random series.[3]

This is the reason why we can make the following statement: If we find a certain finite difference of the order k_0 such that the variance of the k_0th difference is equal to the variance of the (k_0+1)th difference and equal to that of the (k_0+2)th difference, etc., then we shall be justified in assuming that we have eliminated the mathematical expectation to a reasonable degree by taking k_0 differences.[4] The equality will, of course, never hold exactly true, since there will always be a certain amount of random variation. We are essentially dealing with problems in the realm of probability. Hence, we shall show that it is required only that the difference between the variances of two successive series of finite differences be smaller than three times its standard error. This is sufficient from the point of view of the theory of probability in assuring us reasonably well that in this k_0th difference only traces of the mathematical expectation are left. This assumption is justified only if our series is long enough. (We shall present a procedure in Chapter VIII which holds true also for short series.) We can base this argument on Tchebycheff's inequality,[5] which holds true for all distributions. But in our case the distributions are likely to be nearly normal and so we shall rarely be wrong in taking three times the standard error as our "fiducial limit."[6]

[2] O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung,* Bonn, 1929, pp. 44 ff., 52 ff.

[3] O. Anderson, *op. cit.,* pp. 54 ff. See also below, Appendix II, pp. 138 ff.

[4] R. Zaycoff, "Ueber die Ausschaltung der zufälligen Komponente nach der 'Variate-Difference' Methode," *Publications of the Statistical Institute for Economic Research,* State University of Sofia, 1937, No. 1, pp. 78 ff.

[5] J. V. Uspensky, *Introduction to Mathematical Probability,* New York, 1937, pp. 182 ff. See also below, Appendix II, pp. 139 ff.

[6] R. A. Fisher, "The Logic of Inductive Inference," *Journal of the Royal Statistical Society,* Vol. 98, 1935, pp. 39 ff.; J. Neyman, "On the Two Different Aspects of the Representative Method," *ibid.,* Vol. 97, 1934, pp. 558 ff. See also R. A. Fisher, *Statistical Methods for Research Workers,* 7th ed., London, 1938, pp. 42 ff., Chapter III, Section 11; J. Neyman, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Philosophical Transactions of the Royal Society of London,* Series A., Vol. 236, 1937, pp. 333 ff.; S. S. Wilks, *Statistical Inference,* Princeton, 1937, pp. 71 ff.; R. A. Fisher, *The Design of Experiments,* London, 1935, pp. 15 ff.

This corresponds to a level of significance of at least 8/9 from Tchebycheff's inequality. This inequality states that the probability of a deviation, positive or negative, from the mathematical expectation by more than a times the standard deviation or standard error (where a may be any positive number) is less than or equal to $1/a^2$. In our case a equals 3. We get, of course, much better limits if we assume normal distributions, an assumption that is sometimes justified in economic data both from a priori considerations and from an empirical point of view (Appendix VII). Three times the standard deviation or standard error gives us here a level of significance of 0.9973. That is to say, we shall be disappointed only in about 3 cases in 1000 if we accept this level of significance.

The common levels of tests of significance, i.e., the 5-per-cent point, 1-per-cent point, and 0.1-per-cent point, give the following values in the case of Tchebycheff's inequality: 4.47 times, 10 times, and 31.6 times the standard deviation. This means, for instance, that the chances are not more than 1 in 100 that we may get by chance a deviation from the mathematical expectation greater than 10 times the population standard deviation or error. It should be pointed out again that Tchebycheff's inequality holds true for any distribution. The same levels of significance, i.e., 5 per cent, 1 per cent, and 0.1 per cent, give the following fiducial limits in case of a normal distribution, which we may sometimes expect in economic data: 1.960, 2.576, and 3.290 times the population standard deviation. That is to say, we can, for instance, expect that we shall be wrong only 1 time in 100 cases in rejecting the hypothesis if we have a normal distribution and we get a deviation from the hypothetical value greater than 2.576 times the population standard deviation or the standard error.[7]

Our hypothesis is that the variances are equal. We consider here only errors of Type I, i.e., errors committed by rejecting the hypothesis if it is really true. Besides, there are errors of Type II which arise if the hypothesis is not rejected if it is really false, i.e., if the variances are not in reality equal.

We have first to find the variances of our original series and of the series of finite differences in order to carry out this analysis. We give, as an example, the yearly series of prices of wheat flour in the United States between the years 1890 and 1937, that is, for 48 years. The data are taken from the publications of wholesale prices of the United States Bureau of Labor Statistics.

[7] G. U. Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, 11th ed., London, 1937, pp. 350 ff.

TABLE 7

ANNUAL AMERICAN WHEAT-FLOUR PRICES AND DIFFERENCES, 1890-1937
(United States Bureau of Labor Statistics)

Year	Price w	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$	$\Delta^{(4)}$	$\Delta^{(5)}$	$\Delta^{(6)}$	$\Delta^{(7)}$	$\Delta^{(8)}$	$\Delta^{(9)}$	$\Delta^{(10)}$
1890	5.185	0.120	-1.079	1.698	-2.389	3.613	-5.728	9.633	-18.070	36.491	-71.374
1891	5.305	-0.959	0.619	-0.691	1.224	-2.115	3.905	-8.437	18.421	-34.883	51.375
1892	4.346	-0.340	-0.072	0.533	-0.891	1.790	-4.532	9.984	-16.462	16.492	7.682
1893	4.006	-0.412	0.461	-0.358	0.899	-2.742	5.452	-6.478	0.030	24.174	-81.797
1894	3.594	0.049	0.103	0.541	-1.843	2.710	-1.026	-6.448	24.204	-57.623	112.119
1895	3.643	0.152	0.644	-1.302	0.867	1.684	-7.474	17.756	-33.419	54.496	-81.204
1896	3.795	0.796	-0.658	-0.435	2.551	-5.790	10.282	-15.663	21.077	-26.708	36.656
1897	4.591	0.138	-1.093	2.116	-3.239	4.492	-5.381	5.414	-5.631	9.948	-23.380
1898	4.729	-0.955	1.023	-1.123	1.253	-0.889	0.033	-0.217	4.317	-13.432	19.027
1899	3.774	0.068	-0.100	0.130	0.364	-0.856	-0.184	4.100	-9.115	5.595	30.339
1900	3.842	-0.032	0.030	0.494	-0.492	-1.040	3.916	-5.015	-3.520	35.934	-114.757
1901	3.810	-0.002	0.524	0.002	-1.532	2.876	-1.099	-8.535	32.414	-78.823	159.026
1902	3.808	0.522	0.526	-1.530	1.344	1.777	-9.634	23.879	-46.409	80.203	-129.836
1903	4.330	1.048	-1.004	-0.186	3.121	-7.857	14.245	-22.530	33.794	-49.633	69.538
1904	5.378	0.044	-1.190	2.935	-4.736	6.388	-8.285	11.264	-15.839	19.905	-12.713
1905	5.422	-1.146	1.745	-1.801	1.652	-1.897	2.979	-4.575	4.066	7.192	-50.060
1906	4.276	0.599	-0.056	-0.149	-0.245	1.082	-1.596	-0.509	11.258	-42.868	114.494
1907	4.875	0.543	-0.205	-0.394	0.837	-0.514	-2.105	10.749	-31.610	71.626	-130.285
1908	5.418	0.338	-0.599	0.443	0.323	-2.619	8.644	-20.861	40.016	-58.659	41.305
1909	5.756	-0.261	-0.156	0.766	-2.296	6.025	-12.217	19.155	-18.643	-17.354	167.257
1910	5.495	-0.417	0.610	-1.530	3.729	-6.192	6.938	0.512	-35.997	149.903	-452.791
1911	5.078	0.193	-0.920	2.199	-2.463	0.746	7.450	-35.485	113.906	-302.888	704.886
1912	5.271	-0.727	1.279	-0.264	-1.717	8.196	-28.035	78.421	-188.982	401.998	-760.376
1913	4.544	0.552	1.015	-1.981	6.479	-19.839	50.386	-110.561	213.016	-358.378	511.524
1914	5.096	1.567	-0.966	4.498	-13.360	30.547	-60.175	102.455	-145.362	153.146	-50.129
1915	6.663	0.601	3.532	-8.862	17.187	-29.628	42.280	-42.907	7.784	103.017	-352.769

TABLE 7 (concluded)

ANNUAL AMERICAN WHEAT-FLOUR PRICES AND DIFFERENCES, 1890-1937
(United States Bureau of Labor Statistics)

Year	Price w	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$	$\Delta^{(4)}$	$\Delta^{(5)}$	$\Delta^{(6)}$	$\Delta^{(7)}$	$\Delta^{(8)}$	$\Delta^{(9)}$	$\Delta^{(10)}$
1916	7.264	4.133	–5.330	8.325	–12.441	12.652	–0.627	–35.123	110.801	–249.752	482.893
1917	11.397	–1.197	2.995	–4.116	0.211	12.025	–35.750	75.678	–138.951	233.141	–359.934
1918	10.200	1.798	–1.121	–3.905	12.236	–23.725	39.928	–63.273	94.190	–126.793	144.470
1919	11.998	0.677	–5.026	8.331	–11.489	16.203	–23.345	30.917	–32.603	17.677	26.777
1920	12.675	–4.349	3.305	–3.158	4.714	–7.142	7.572	–1.686	–14.926	44.454	–84.150
1921	8.326	–1.044	0.147	1.556	–2.428	0.430	5.886	–16.612	29.528	–39.696	40.457
1922	7.282	–0.897	1.703	–0.872	–1.998	6.316	–10.726	12.916	–10.168	0.761	14.327
1923	6.385	0.806	0.831	–2.870	4.318	–4.410	2.190	2.748	–9.407	15.088	–14.442
1924	7.191	1.637	–2.039	1.448	–0.092	–2.220	4.938	–6.659	5.681	0.646	–20.038
1925	8.828	–0.402	–0.591	1.356	–2.312	2.718	–1.721	–0.978	6.327	–19.392	55.302
1926	8.426	–0.993	0.765	–0.956	0.406	0.997	–2.699	5.349	–13.065	35.910	–94.134
1927	7.433	–0.228	–0.191	–0.550	1.403	–1.702	2.650	–7.716	22.845	–58.224	134.067
1928	7.205	–0.419	–0.741	0.853	–0.299	0.948	–5.066	15.129	–35.379	75.843	....
1929	6.786	–1.160	0.112	0.554	0.649	–4.118	10.063	–20.250	40.464	....	....
1930	5.626	–1.048	0.666	1.203	–3.469	5.945	–10.187	20.214	....	....	....
1931	4.578	–0.382	1.869	–2.266	2.476	–4.242	10.027	....	....	....	....
1932	4.196	1.487	–0.397	0.210	–1.766	5.785	....	....	....	....	....
1933	5.683	1.090	–0.187	–1.556	4.019	....	....	....	....	....	....
1934	6.773	0.903	–1.743	2.463	....	....	....	....	....	....	....
1935	7.676	–0.840	0.720	....	....	....	....	....	....	....	....
1936	6.836	–0.120	....	....	....	....	....	....	....	....	....
1937	6.716	....	....	....	....	....	....	....	....	....	....

We show in Table 7 the process of finite differencing. The series of first finite differences is formed by subtracting the first item of the original series (5.185) from the second item (5.305), the result being +0.120, the first item in the series of first differences. Then the second item is subtracted from the third item, the third item from the fourth item, etc. In this way we get the series of first finite differences. In this series again we subtract the first item (+0.120) from the second item (—0.959), which gives —1.079, the first item in the series of second differences. The second item is subtracted from the third item, and the third item from the fourth item, etc., and we get the series of the second finite differences. By carrying out this process ten times, we finally have ten series of finite differences which are shown in Table 7.

A very useful check for this procedure consists in the fact that the difference between the nth item and the first item of the kth difference is equal to the sum of the first $n—1$ items of the $(k+1)$th finite difference. (We denote here for the sake of convenience our original series as the 0th difference.) Summing the ten items of the series of our first differences: +0.120 — 0.959 — 0.340 — 0.412 + 0.049 + 0.152 + 0.796 + 0.138 — 0.955 + 0.068 = —1.343. The difference between the eleventh and the first item of our original series (0th difference) is equal to this sum of the first ten items of the series of the first finite differences, i.e., 3.842 — 5.185 = —1.343. Similar checks, of course, hold true for the higher differences as well. For instance, the difference between the eleventh item in our series of first differences and the first item must be equal to the sum of the first ten items in our series of second differences, etc. Those checks should always be carried through, because otherwise some errors may creep into the calculations of the finite differences. It is desirable that those checks should be made at not too infrequent intervals, because this facilitates the detection of errors.

In calculating the finite differences we must bear in mind the rules for algebraic subtraction. That is to say, to subtract a negative number is the same as to add the absolute value, etc.

We notice, by the way, that every series of finite differences is one item shorter than the preceding one, so that by taking ten series of finite differences, we have finally reduced the number of items in our series from 48 to 38. This, of course, puts a very definite limitation on the number of finite differences that we can calculate. In a short series like the one we are treating as an example, the calculation of very high differences reduces the number of items in our series considerably. The same would not be true, however, in a series of monthly data where the reduction of the number of items would be almost negligible. As a general rule, the order of the highest difference calculated should never be greater than half the number of items of the original series. We could in our case go so far as the 24th difference.[8]

Having calculated the series of finite differences, we have now to find the variances and, also, for reasons given later on, the fourth moments.[9] The variances (squares of the standard deviations) of the finite differences are the statistical estimates or parameters that are most important for the carrying through of the process of analysis by the variate difference method.

The variance of the original series could, of course, be calculated in the ordinary manner. That is to say, we could take the deviations from the arithmetic mean and square them, sum them, and divide the sum by the number of items less one, in our case, by 47. This is the mean sum of squares and gives us the best estimate for the variance of our original series. We divide by $N—1$ instead of N, because this is the number of degrees of freedom left after we have calculated the mean. Let our original series be called the difference of order zero. We call $\bar{w} = S_1^{(0)}/N$ the arithmetic mean of our original series, where $S_1^{(0)} = \sum_{i=1}^{N} w_i$ is the sum of the original item in our series[10] (Table 8). The variance of the 0th difference or the series of original items is then:

$$V_0 = \frac{\sum_{i=1}^{N} (w_i - \bar{w})^2}{(N-1)} .$$

This can also be found by calculating the sum of the squares in the following manner, which is probably more convenient on a calculating machine.[11]

First, the arithmetic mean is found by summing up the items of the original series (giving in our example a sum $S_1^{(0)} = 291.510$) and dividing the total by the number of items, that is to say, 48. The arithmetic mean $\bar{w}$ is 6.0731; then the items in the original series are squared and we deduct from the sum of squares of the original items ($S_2^{(0)} = 1{,}995.816$) the square of the arithmetic mean ($\bar{w}^2 = 36.8825$)

[8] O. Anderson, *op. cit.*, p. 55.

[9] O. Anderson, *op. cit.*, pp. 54 ff.; R. Zaycoff, *op. cit.*, pp. 78 ff.; G. Tintner, *Prices in the Trade Cycle*, Vienna, 1935, pp. 11 ff. See also below, Appendix II, pp. 139 ff. For a summary of computations see Appendix I, Section A.

[10] The notation is taken from R. A. Fisher, *Statistical Methods for Research Workers*, 7th ed., London, 1938, p. 74, Chapter III, appendix.

[11] G. W. Snedecor, *Statistical Methods*, Ames, Iowa, 1937, pp. 134 ff.

multiplied by the number of items in the original series ($N = 48$). In this way we find the sums of the squares of deviations from the arithmetic mean: 1,995.816 — 1,770.360 = 225.456. By dividing this number by the number of items in the series minus 1 (that is to say, $N-1 = 47$) we get the best estimate of the variance of our original series: $V_0 = 225.456/47 = 4.7969$. The formula for the calculation of the variance of the original series is then:

$$V_0 = \frac{(S_2^{(0)} - N\bar{w}^2)}{(N-1)},$$

where $S_2^{(0)} = \sum_{i=1}^{N} w_i^2$ is the sum of the squares of the items of the original series[12] (Table 8, Table 17).

We do not regard the sum of the squares of the deviations from the arithmetic mean as the best basis for an estimate of variances of finite differences. We take, rather, the sum of squares of deviations *from zero* or simply the sum of the squares of the finite differences themselves.[13] The reason for this is the following: we know from our assumptions that the true mean or mathematical expectation of the random element is equal to zero. Hence, instead of taking the deviations from the arithmetic mean of the finite differences we take the deviations from the true mean, which, by definition, is equal to zero, since the mathematical expectation of the differences of the random element is equal to zero. (The arithmetic mean of any series of finite differences is the difference between the last and first item of the preceding series of finite differences divided by the number of items. This is the basis of the checks we have indicated above. There is, however, no use in calculating those means of the series of finite differences.)

All we have to do is to square the finite differences and sum them up. This can easily be done by the use of a table of squares, or with a calculating machine. The sums of the squares appear in Table 8 ($S_2^{(k)}$).

We have then to find the variance of every series of finite differences. This is done by dividing the sum of the squares by the number of degrees of freedom, i.e., by the number of items in the particular series of finite differences, and then by another constant. We divide the sum of the squares of the first finite differences by the number of items of our original series minus 1, that is, by 47. The sum of squares of the third finite differences, for instance, is divided by the sum of the items in our original series minus 3, that is to say by 45, etc. This gives the mean squares.

[12] R. A. Fisher, *loc. cit.*
[13] O. Anderson, *op. cit.*, pp. 54 ff., 112 ff.

Having found the mean squares, we have not yet the best estimate of the random variance of the finite differences. It appears that the mean squares increase very rapidly. Professor Anderson has shown that the original variance of a random series is multiplied by a certain binomial coefficient with every successive finite differencing.[14] This coefficient is for the kth difference equal to the number of combinations of $2k$ things taken k at a time, ${}_{2k}C_k$. For instance, the mean square of the first finite differences is equal to the variance of the ran-

TABLE 8

SUMMARY, ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Order of Difference	Sum	Sum of Squares	Sum of Cubes	Sum of Fourth Powers
k	$S_1^{(k)}$	$S_2^{(k)}$	$S_3^{(k)}$	$S_4^{(k)}$
0	291.510	1,995.816	15,497.211	135,336.771
1	1.531	65.445	–0.755	695.146
2	–0.240	120.427	–170.871	1,873.501
3	1.799	350.674	401.786	17,139.532
4	0.765	1,148.594	1,455.661	187,135.609
5	6.408	3,928.982	–10,188.148	2,250,883.230
6	2.172	13,645.040	–28,194.338	27,985,203.869
7	15.755	48,688.015	251,455.786	354,892,010.774
8	10.581	173,928.318	785,204.865	4,572,561,429.029
9	58.534	624,302.720	–5,415,515.798	59,505,060,360.278
10	39.352	2,251,828.836	–17,454,629.992	781,714,768,034.777

TABLE 8 (continued)

SUMMARY, ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Order of Difference	Sum of Fifth Powers	Sum of Sixth Powers
k	$S_5^{(k)}$	$S_6^{(k)}$
0	1,302,985.812	13,504,986.791
1	–311.003	11,847.364
2	–6,325.144	43,227.464
3	25,001.813	1,171,261.175
4	863,946.123	40,940,464.576
5	–5,114,183.497	1,694,476,476.759
6	–309,827,424.524	76,358,979,389.846
7	–1,053,502,071.712	3,476,905,425,869.029
8	124,020,260,640.821	160,375,263,473,164.003
9	1,896,199,979,918.160	7,544,921,827,024,200.395
10	–49,157,453,123,299.549	359,319,726,450,087,749.431

[14] O. Anderson, *op. cit.*, pp. 54 ff.

TABLE 8 (concluded)

SUMMARY, ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Order of Difference	Sum of Seventh Powers	Sum of Eighth Powers
k	$S_7^{(k)}$	$S_8^{(k)}$
0	147,489,750.629	1,669,633,466.799
1	–8,704.986	213,353.332
2	–189,923.522	1,104,199.000
3	1,265,224.788	84,645,926.793
4	336,063,393.199	10,012,592,708.261
5	–280,519,108.082	1,482,185,263,372.150
6	–1,721,183,508,245.942	233,273,978,783,075.287
7	–55,703,051,803,429.628	37,217,856,952,905,647.657
8	9,440,545,591,946,463.816	6,261,852,918,586,706,732.149
9	683,544,731,361,408,015.578	1,049,457,745,656,437,306,286.888
10	–50,567,669,300,918,999,870.114	182,625,520,067,971,771,696,401.337

Number of items in the original series: $N = 48$.

Mean of the original series: $\bar{w} = 6.0731$.

Fourth moment of the original series about the mean: $m_4 = 96.8859$.

dom series multiplied by the number of combinations of 2 things taken 1 at a time, which is equal to 2. The mean square of the second differences is the random variance multiplied by the number of combinations of 4 things taken 2 at a time, which is equal to 6. The same holds true for the higher differences. For instance, the mean square of the series of the tenth finite differences is equal to the variance of the original random series multiplied by the number of combinations of 20 things taken 10 at a time, which is equal to 184,756. Those coefficients are tabulated in Table 9, which we take from Mr. Zaycoff's publication.[15] In order to get the estimate for the variance of the random element, we have simply to divide the mean square by the binomial coefficient ${}_{2k}C_k$ given in Table 9.

In order to facilitate the numerical work we have calculated coefficients A_{kN} which are given in Table 10. They are equal to

$$\frac{1}{{}_{2k}C_k\,(N-k)}$$

where N is the total number of items in the original series, k the order of the difference, and ${}_{2k}C_k$ the binomial coefficient mentioned above. All we have to do is to enter our table under the right N and k and multiply the sum of squares of the differences in question by the num-

15 R. Zaycoff, *op. cit.*, p. 80, Table 1.

TABLE 9

BINOMIAL COEFFICIENTS ${}_{2k}C_k$

Order of Difference	Coefficients
k	${}_{2k}C_k$
0	1
1	2
2	6
3	20
4	70
5	252
6	924
7	3,432
8	12,870
9	48,620
10	184,756

ber given. Interpolation is necessary and should be done with the help of the (negative) divided differences given in the same table.

We enter the table, for instance, under $k=1$ and interpolate for $N=48$ in order to find the number by which we have to multiply the sum of the squares of the first finite differences to get the best estimate for their variance V_1. The same process is repeated for every difference.

The variances of the first and higher differences are calculated according to the formula $V_k = S_2^{(k)} A_{kN}$, where

$$S_2^{(k)} = \sum_{i=1}^{N-k} \left[\Delta^{(k)} w_i \right]^2,$$

i.e., the sum of the squares of the kth differences (Table 8). We find from Table 10, for $k=1$ and $N=48$, $A_{kN} = 0.010727$. From Table 8 we have, for $k=1$, $S_2^{(k)} = 65.445$, the sum of the squares of the first differences. The best estimate of the variance of the series of first differences, $V_1 = (65.445)\,(0.010\,727) = 0.7020$ (Table 17). The same result would have been reached approximately by dividing $S_2^{(k)}$ by $(N-1)\,{}_2C_1$, i.e., calculating $65.445/(47 \cdot 2) = 0.6962$. Similarly, for the fourth difference, we get from Table 10 for $k=4$, $N=48$, the coefficient $A_{kN} = 0.000328$. Multiplying by the value $S_2^{(k)}$ for $k=4$ in Table 8, which is 1,148.594 (sum of squares of fourth differences), we get for the best estimate of the variance of the fourth differences $V_4 = 0.3767$ (Table 17). This is simpler than dividing 1,148.594 by $(N-4)\,{}_8C_4$ or calculating $1{,}148.594/(44 \cdot 70) = 0.3729$.

The variances of the series of finite differences calculated in the manner indicated above form our estimate for the true variance of the

TABLE 10

COEFFICIENTS $A_{kN} = \frac{1}{{}_{2k}C_k (N - k)}$

(k = Order of Difference)

Number of Items in the Original Series	$k = 1$		$k = 2$	
N	A_{1N}	D.D.*	A_{2N}	D.D.*
10	0.055 556	0.002 9240	0.020 833	0.001 1573
20	0.026 316	0.000 9074	0.009 260	0.000 3308
30	0.017 242	0.000 4422	0.005 952	0.000 1566
40	0.012 820	0.000 2616	0.004 386	0.000 0914
50	0.010 204	0.000 1730	0.003 472	0.000 0598
60	0.008 474	0.000 1228	0.002 874	0.000 0423
70	0.007 246	0.000 0917	0.002 451	0.000 0314
80	0.006 329	0.000 0711	0.002 137	0.000 0243
90	0.005 618	0.000 0567	0.001 894	0.000 0193
100	0.005 051	0.000 03390	0.001 701	0.000 01150
150	0.003 356	0.000 01688	0.001 126	0.000 00568
200	0.002 512	0.000 01008	0.000 842	0.000 00340
250	0.002 008	0.000 00672	0.000 672	0.000 00226
300	0.001 672	0.000 00480	0.000 559	0.000 00160
350	0.001 432	0.000 00358	0.000 479	0.000 00120
400	0.001 253	0.000 00278	0.000 419	0.000 00094
450	0.001 114	0.000 00224	0.000 372	0.000 00074
500	0.001 002	0.000 00182	0.000 335	0.000 00062
550	0.000 911	0.000 00152	0.000 304	0.000 00050
600	0.000 835	0.000 00130	0.000 279	0.000 00044
650	0.000 770	0.000 00108	0.000 257	0.000 00036
700	0.000 716	0.000 00096	0.000 239	0.000 00032
750	0.000 668	0.000 00084	0.000 223	0.000 00028
800	0.000 626	0.000 00074	0.000 209	0.000 00024
850	0.000 589	0.000 00066	0.000 197	0.000 00022
900	0.000 556	0.000 00058	0.000 186	0.000 00020
950	0.000 527	0.000 00052	0.000 176	0.000 00018
1000	0.000 501	....	0.000 167	....

* Divided difference, negative.

TABLE 10 (continued)

COEFFICIENTS $A_{kN} = \frac{1}{{}_{2k}C_k(N-k)}$

(k = Order of Difference)

Number of Items in the Original Series	$k = 3$		$k = 4$	
N	A_{3N}	D.D.*	A_{4N}	D.D.*
10	0.007 143	0.000 420 2	0.002 381	0.000 148 8
20	0.002 941	0.000 108 9	0.000 893	0.000 034 3
30	0.001 852	0.000 050 1	0.000 550	0.000 015 3
40	0.001 351	0.000 028 7	0.000 397	0.000 008 6
50	0.001 064	0.000 018 7	0.000 311	0.000 005 6
60	0.000 877	0.000 013 1	0.000 255	0.000 003 9
70	0.000 746	0.000 009 7	0.000 216	0.000 002 8
80	0.000 649	0.000 007 4	0.000 188	0.000 002 2
90	0.000 575	0.000 006 0	0.000 166	0.000 001 7
100	0.000 515	0.000 003 5	0.000 149	0.000 001 024
150	0.000 340	0.000 001 72	0.000 0978	0.000 000 498
200	0.000 254	0.000 001 04	0.000 0729	0.000 000 296
250	0.000 202	0.000 000 68	0.000 0581	0.000 000 196
300	0.000 168	0.000 000 48	0.000 0483	0.000 000 140
350	0.000 144	0.000 000 36	0.000 0413	0.000 000 104
400	0.000 126	0.000 000 28	0.000 0361	0.000 000 082
450	0.000 112	0.000 000 22	0.000 0320	0.000 000 064
500	0.000 101	0.000 000 192	0.000 0288	0.000 000 052
550	0.000 0914	0.000 000 152	0.000 0262	0.000 000 044
600	0.000 0838	0.000 000 130	0.000 0240	0.000 000 038
650	0.000 0773	0.000 000 110	0.000 0221	0.000 000 032
700	0.000 0718	0.000 000 096	0.000 0205	0.000 000 028
750	0.000 0670	0.000 000 086	0.000 0191	0.000 000 024
800	0.000 0627	0.000 000 074	0.000 0179	0.000 000 020
850	0.000 0590	0.000 000 066	0.000 0169	0.000 000 020
900	0.000 0557	0.000 000 058	0.000 0159	0.000 000 016
950	0.000 0528	0.000 000 052	0.000 0151	0.000 000 016
1000	0.000 0502	....	0.000 0143	....

* Divided difference, negative.

TABLE 10 (continued)

COEFFICIENTS $A_{kN} = \frac{1}{{}_{2k}C_k(N-k)}$

(k = Order of Difference)

Number of Items in the Original Series	$k=5$		$k=6$	
N	A_{5N}	D.D.*	A_{6N}	D.D.*
10	0.000 794	0.000 052 9	0.000 271	0.000 019 37
20	0.000 265	0.000 010 6	0.000 0773	0.000 003 22
30	0.000 159	0.000 004 6	0.000 0451	0.000 001 33
40	0.000 113	0.000 002 48	0.000 0318	0.000 000 72
50	0.000 0882	0.000 001 60	0.000 0246	0.000 000 46
60	0.000 0722	0.000 001 11	0.000 0200	0.000 000 31
70	0.000 0611	0.000 000 82	0.000 0169	0.000 000 23
80	0.000 0529	0.000 000 62	0.000 0146	0.000 000 17
90	0.000 0467	0.000 000 49	0.000 0129	0.000 000 14
100	0.000 0418	0.000 000 288	0.000 0115	0.000 000 0796
150	0.000 0274	0.000 000 140	0.000 00752	0.000 000 0388
200	0.000 0204	0.000 000 084	0.000 00558	0.000 000 0228
250	0.000 0162	0.000 000 054	0.000 00444	0.000 000 0152
300	0.000 0135	0.000 000 040	0.000 00368	0.000 000 0106
350	0.000 0115	0.000 000 030	0.000 00315	0.000 000 0080
400	0.000 0100	0.000 000 0216	0.000 00275	0.000 000 0062
450	0.000 00892	0.000 000 0180	0.000 00244	0.000 000 0050
500	0.000 00802	0.000 000 0148	0.000 00219	0.000 000 0040
550	0.000 00728	0.000 000 0122	0.000 00199	0.000 000 0034
600	0.000 00667	0.000 000 0104	0.000 00182	0.000 000 0028
650	0.000 00615	0.000 000 0088	0.000 00168	0.000 000 0024
700	0.000 00571	0.000 000 0076	0.000 00156	0.000 000 0022
750	0.000 00533	0.000 000 0068	0.000 00145	0.000 000 0018
800	0.000 00499	0.000 000 0060	0.000 00136	0.000 000 0016
850	0.000 00469	0.000 000 0052	0.000 00128	0.000 000 0014
900	0.000 00443	0.000 000 0046	0.000 00121	0.000 000 0012
950	0.000 00420	0.000 000 0042	0.000 00115	0.000 000 0012
1000	0.000 00399	....	0.000 00109	....

* Divided difference, negative.

TABLE 10 (continued)

COEFFICIENTS $A_{kN} = \dfrac{1}{{}_{2k}C_k(N-k)}$

(k = Order of Difference)

Number of Items in the Original Series	$k=7$		$k=8$	
N	A_{7N}	D.D.*	A_{8N}	D.D*
10	0.000 097 1	0.000 007 47	0.000 038 9	0.000 003 243
20	0.000 022 4	0.000 000 97	0.000 006 47	0.000 000 294
30	0.000 012 7	0.000 000 387	0.000 003 53	0.000 000 110
40	0.000 008 83	0.000 000 205	0.000 002 43	0.000 000 058
50	0.000 006 78	0.000 000 128	0.000 001 85	0.000 000 036
60	0.000 005 50	0.000 000 087	0.000 001 49	0.000 000 024
70	0.000 004 63	0.000 000 064	0.000 001 25	0.000 000 017
80	0.000 003 99	0.000 000 048	0.000 001 08	0.000 000 013 2
90	0.000 003 51	0.000 000 038	0.000 000 948	0.000 000 010 3
100	0.000 003 13	0.000 000 0218	0.000 000 845	0.000 000 005 96
150	0.000 002 04	0.000 000 0106	0.000 000 547	0.000 000 002 84
200	0.000 001 51	0.000 000 0062	0.000 000 405	0.000 000 001 68
250	0.000 001 20	0.000 000 00412	0.000 000 321	0.000 000 001 10
300	0.000 000 994	0.000 000 00290	0.000 000 266	0.000 000 000 78
350	0.000 000 849	0.000 000 00214	0.000 000 227	0.000 000 000 58
400	0.000 000 742	0.000 000 00168	0.000 000 198	0.000 000 000 44
450	0.000 000 658	0.000 000 00134	0.000 000 176	0.000 000 000 36
500	0.000 000 591	0.000 000 00108	0.000 000 158	0.000 000 000 30
550	0.000 000 537	0.000 000 00092	0.000 000 143	0.000 000 000 24
600	0.000 000 491	0.000 000 00076	0.000 000 131	0.000 000 000 20
650	0.000 000 453	0.000 000 00066	0.000 000 121	0.000 000 000 18
700	0.000 000 420	0.000 000 00056	0.000 000 112	0.000 000 000 14
750	0.000 000 392	0.000 000 00050	0.000 000 105	0.000 000 000 138
800	0.000 000 367	0.000 000 00042	0.000 000 0981	0.000 000 000 116
850	0.000 000 346	0.000 000 00040	0.000 000 0923	0.000 000 000 104
900	0.000 000 326	0.000 000 00034	0.000 000 0871	0.000 000 000 092
950	0.000 000 309	0.000 000 00032	0.000 000 0825	0.000 000 000 084
1000	0.000 000 293	....	0.000 000 0783	....

* Divided difference, negative.

TABLE 10 (concluded)

COEFFICIENTS $A_{kN} = \frac{1}{{}_{2k}C_k (N - k)}$

(k = Order of Difference)

Number of Items in the Original Series	$k = 9$		$k = 10$	
N	A_{9N}	D.D.*	A_{10N}	D.D.*
10	0.000 020 6	0.000 001 873	....	....
20	0.000 001 87	0.000 000 089 1	0.000 000 541	0.000 000 027 0
30	0.000 000 979	0.000 000 031 6	0.000 000 271	0.000 000 009 1
40	0.000 000 663	0.000 000 016 1	0.000 000 180	0.000 000 004 5
50	0.000 000 502	0.000 000 009 9	0.000 000 135	0.000 000 002 7
60	0.000 000 403	0.000 000 006 6	0.000 000 108	0.000 000 001 78
70	0.000 000 337	0.000 000 004 7	0.000 000 0902	0.000 000 001 29
80	0.000 000 290	0.000 000 003 6	0.000 000 0773	0.000 000 000 96
90	0.000 000 254	0.000 000 002 8	0.000 000 0677	0.000 000 000 76
100	0.000 000 226	0.000 000 001 60	0.000 000 0601	0.000 000 000 428
150	0.000 000 146	0.000 000 000 76	0.000 000 0387	0.000 000 000 204
200	0.000 000 108	0.000 000 000 454	0.000 000 0285	0.000 000 000 118
250	0.000 000 0853	0.000 000 000 292	0.000 000 0226	0.000 000 000 078
300	0.000 000 0707	0.000 000 000 208	0.000 000 0187	0.000 000 000 056
350	0.000 000 0603	0.000 000 000 154	0.000 000 0159	0.000 000 000 040
400	0.000 000 0526	0.000 000 000 120	0.000 000 0139	0.000 000 000 032
450	0.000 000 0466	0.000 000 000 094	0.000 000 0123	0.000 000 000 026
500	0.000 000 0419	0.000 000 000 078	0.000 000 0110	0.000 000 000 020
550	0.000 000 0380	0.000 000 000 064	0.000 000 0100	0.000 000 000 0166
600	0.000 000 0348	0.000 000 000 054	0.000 000 00917	0.000 000 000 0142
650	0.000 000 0321	0.000 000 000 046	0.000 000 00846	0.000 000 000 0124
700	0.000 000 0298	0.000 000 000 040	0.000 000 00784	0.000 000 000 0106
750	0.000 000 0278	0.000 000 000 036	0.000 000 00731	0.000 000 000 0092
800	0.000 000 0260	0.000 000 000 030	0.000 000 00685	0.000 000 000 0082
850	0.000 000 0245	0.000 000 000 028	0.000 000 00644	0.000 000 000 0072
900	0.000 000 0231	0.000 000 000 024	0.000 000 00608	0.000 000 000 0064
950	0.000 000 0219	0.000 000 000 022	0.000 000 00576	0.000 000 000 0058
1000	0.000 000 0208	....	0.000 000 00547	....

* Divided difference, negative.

random element in our time series (Table 17). The k_0th difference has by assumption the property that we can be reasonably sure from the point of view of probability that we have eliminated almost everything but the random element. Hence the variance of the k_0th difference and of all higher differences must be nearly or really equal—at least the differences between them must be so small that we can expect that they would be equal if it were not for certain irregular random variations. We have to calculate the standard errors of the differences between the variances of two consecutive series of finite differences in order to have an objective test for this. This will be done in the following chapter.

This is, however, only one of the two possible approaches. The other will be taken up in Chapter VIII. There, we shall not use the crude test of significance with the standard error, which is applicable only in the case of large samples. We shall give an exact test which follows the modern ideas of statistics and can be applied even for short series. It involves, however, a certain sacrifice of the available information.

TABLE 11

SUMMARY, ANNUAL WOOL PRICES, 1890-1937

Order of Difference	Sum	Sum of Squares	Sum of Cubes	Sum of Fourth Powers
k	$S_1^{(k)}$	$S_2^{(k)}$	$S_3^{(k)}$	$S_4^{(k)}$
0	29.845	23.583	23.092	26.754
1	....	2.581	....	1.355
2	....	7.085	....	8.659
3	....	23.407	....	95.190
4	....	80.417	....	1,116.637
5	....	281.707	....	13,693.515
6	....	999.810	....	172,289.439
7	....	3,584.558	....	2,205,897.664
8	....	12,959.048	....	28,658,448.268
9	....	47,171.816	....	376,608,652.443
10	....	172,723.323	....	5,004,963,519.913

Number of items in the original series: $N = 48$.

Mean of the original series: $\overline{w} = 0.6218$.

Fourth moment of the original series about the mean: $m_4 = 0.0521$.

We show in Tables 11 to 13 the summaries of the sums, sums of squares, and higher powers for the original and differences of the following prices, which we shall use later: Annual Wool Prices, Monthly Wool Prices, Annual Raw-Silk Prices. In Tables 23 to 25 we list the corresponding variances of the original data and the differences. All prices are for American commodities and refer to the years 1890 to 1937. They are all taken from the wholesale price series of the United States Bureau of Labor Statistics.

In order to show an example for a monthly price, let us consider

TABLE 12

SUMMARY, MONTHLY WOOL PRICES, 1890-1937

Order of Difference	Sum	Sum of Squares	Sum of Cubes	Sum of Fourth Powers
k	$S_1^{(k)}$	$S_2^{(k)}$	$S_3^{(k)}$	$S_4^{(k)}$
0	360.37	289.890	296.425	374.627
1	....	2.358	....	1.076
2	....	4.250	....	2.084
3	....	12.543	....	18.720
4	....	41.001	....	170.277
5	....	140.617	....	1,856.234
6	....	497.297	....	21,466.284
7	....	1,797.518	....	262,229.740
8	....	6,605.214	....	3,325,571.725
9	....	24,581.156	....	43,490,554.286
10	....	92,386.963	....	583,154,213.983

Number of items in the original series: $N = 576$.

Mean of the original series: $\bar{w} = 0.6256$.

Fourth moment of the original series about the mean: $m_4 = 0.0849$.

TABLE 13

SUMMARY, ANNUAL RAW-SILK PRICES, 1890-1937

Order of Difference	Sum	Sum of Squares	Sum of Cubes	Sum of Fourth Powers
k	$S_1^{(k)}$	$S_2^{(k)}$	$S_3^{(k)}$	$S_4^{(k)}$
0	205.862	1,018.797	5,695.511	35,365.347
1	....	35.878	....	123.871
2	....	74.258	....	501.398
3	....	206.732	....	3,536.184
4	....	619.728	....	30,480.239
5	....	1,956.779	....	284,308.195
6	....	6,394.062	....	2,915,948.943
7	....	21,322.858	....	31,720,853.344
8	....	72,159.350	....	362,331,906.090
9	....	247,953.972	....	4,284,699,987.645
10	....	867,244.548	....	52,589,449,771.063

Number of items in the original series: $N = 48$.

Mean of the original series: $\bar{w} = 4.2888$.

Fourth moment of the original series about the mean: $m_4 = 28.6504$.

the calculation of V_6, i.e., the variance of the sixth finite differences, of the monthly wool prices. We get from Table 12 for $k=6$, $S_2^{(k)} = 497.297$ (the sum of squares of sixth differences). From Table 10 we get by interpolation for $k=6$ and $N = 576$ (since we have 576 items in the monthly series) $A_{kN} = 0.000\,0019$. The product of these two numbers gives us the best estimate of the true variance of the series of sixth differences of the monthly wool prices as $V_6 = 0.000\,9449$, which is reproduced in Table 24 together with the other variances.

We believe that the calculation of the variances of the differences is greatly simplified by the use of Table 10, which should give a sufficient accuracy for ordinary purposes. The squaring of the differences and original items can be done with the help of tables of squares or a calculating machine. A procedure which eliminates the influence of the seasonal is given in Appendix III.

CHAPTER VI

THE STANDARD ERROR OF THE DIFFERENCE BETWEEN THE VARIANCES OF TWO CONSECUTIVE SERIES OF FINITE DIFFERENCES

In what follows we propose to deal with our problem in the same way in which it was first treated by Professor Oscar Anderson,[1] and we shall also take into account certain refinements recently introduced by Dr. R. Zaycoff[2] of Sofia, Bulgaria. We shall not, however, reproduce here any of the mathematical formulations, but shall content ourselves with the presentation of the logic of the reasoning and also try to indicate the necessary calculations. Some formulae can be found in Appendix II, pp. 139 ff. Appendix I, Section D, gives a summary of computations.

The standard error of the difference between the variances of two consecutive series of finite differences as defined in the last chapter involves the square of the variance of the lower difference series itself, and also the kurtosis or excess of the distribution of finite differences. A distribution with a positive kurtosis has a greater concentration of items around the mean than a normal distribution with the same variance. The reverse is true for a negative kurtosis.[3] The calculations are, however, greatly simplified if we have little or no kurtosis. In this case we can use approximative formulas. This involves a normal or nearly normal distribution of the original series or the differences. (See Appendix VII.)

We have to calculate the fourth moments of our original series and of their finite differences in order to estimate the kurtosis. The fourth moment of the original series should again be calculated for the deviations from the arithmetic mean. We could find it by calculating the deviations of every item of our original series from the arithmetic mean, raising them to the fourth power, and dividing the sum by the number of items in our series, in our case, 48 (annual wheat-flour prices). A more convenient method will be indicated later.

But when we calculate the fourth moment of the first and higher

[1] O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, Bonn, 1929, pp. 57 ff., 113. See also G. Tintner, *Prices in the Trade Cycle*, Vienna, 1935, pp. 14 ff.

[2] R. Zaycoff, "Ueber die Ausschaltung der zufälligen Komponente nach der 'Variate-Difference' Methode," *Publications of the Statistical Institute for Economic Research*, State University of Sofia, 1937, No. 1, pp. 78 ff.

[3] H. T. Davis and W. F. C. Nelson, *Elements of Statistics*, 2nd ed., Bloomington, Indiana, 1937, pp. 317 ff.

differences, we do not take the deviations from the arithmetic mean, but the deviations from the true mean or mathematical expectation of the finite differences, which is by assumption equal to zero. The mathematical expectation or the mean value of the random element is equal to zero by hypothesis. Hence all we have to do is to raise the finite differences to the fourth power. Since we have already calculated the squares, it is very easy to get the fourth powers which are simply the squares of the squares. All we have to do is to take the squares of differences which have already been calculated, and then square them again with the help of a table of squares or a calculating machine. The sum of the fourth powers of the kth differences, $S_4^{(k)}$, is also shown in Table 8 for the annual wheat-flour prices, as well as the sums of the cubes of the original data (0th difference), $S_3^{(0)}$, which we shall need later.

The mean fourth power of a given finite difference is found by dividing the sums of the fourth powers by the number of items in the series. The sum of the fourth powers of the first finite differences is divided by the number of items in our original series minus 1, in our case, 47. The sum of the fourth powers of the second finite differences is divided by the number of items in our original series, minus 2, that is by 46, etc.

TABLE 14

B_k, MULTIPLIER FOR THE SQUARE OF THE VARIANCE

Order of Difference k	$B_k = 3\,({}_{2k}C_k)^2$	Order of Difference k	$B_k = 3\,({}_{2k}C_k)^2$
1	12	6	2,561,328
2	108	7	35,335,872
3	1,200	8	496,910,700
4	14,700	9	7,091,713,200
5	190,512	10	102,404,338,608

In order to find the kurtosis,[4] we still have to deduct from the fourth moment of any finite difference three times the square of the variance multiplied by the square of the binomial coefficient mentioned in Chapter V (Table 9).[5] Table 14 gives these numbers by which the square of the variance is to be multiplied.[6] We call them B_k. The difference between the fourth moment and the square of the

[4] R. Zaycoff, *op. cit.*, pp. 78 ff.
[5] R. A. Fisher, *Statistical Methods for Research Workers*, 7th ed., London, 1938, p. 74, Chapter III, appendix.
[6] R. Zaycoff, *op. cit.*, Table 1, p. 80.

TABLE 15

DIVISOR FOR THE CALCULATION OF THE KURTOSIS

Order of Difference k	$\sum_{i=0}^{k} ({}_kC_i)^4$	Order of Difference k	$\sum_{i=0}^{k} ({}_kC_i)^4$
1	2	6	263,844
2	18	7	3,395,016
3	164	8	44,916,498
4	1,810	9	607,041,380
5	21,252	10	8,345,319,268

TABLE 16

C_k, MULTIPLIER FOR THE DIFFERENCE BETWEEN THE FOURTH MOMENT AND B_k TIMES THE SQUARE OF THE VARIANCE

Order of Difference k	$C_k = \frac{1}{\sum_{i=0}^{k} ({}_kC_i)^4}$	Order of Difference k	$C_k = \frac{1}{\sum_{i=0}^{k} ({}_kC_i)^4}$
1	0.500 000 000 0	6	0.000 003 790 12
2	0.055 555 555 6	7	0.000 000 294 549
3	0.006 097 561 0	8	0.000 000 022 263 5
4	0.000 552 486 2	9	0.000 000 001 647 83
5	0.000 047 054 4	10	0.000 000 000 119 828

TABLE 17

CALCULATION OF KURTOSIS AND STANDARD ERRORS
ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Order of Difference	Variance	Kurtosis				Standard Error of $V_k - V_{k+1}$	Approximate Standard Error of $V_k - V_{k+1}$
k	V_k	D_k	G_k	Q_k	H_{kN}	e_k	e^o_k
0	4.7969	33.4472	1.45	6.73	6.780	0.71267	0.70751
1	0.7020	4.4384	9.01	12.30	13.545	0.05708	0.05183
2	0.4402	1.1000	5.68	16.34	18.108	0.02694	0.02431
3	0.3931	1.1920	7.72	18.30	21.455	0.02148	0.01832
4	0.3767	1.1974	8.44	19.70	23.928	0.01912	0.01574
5	0.3662	1.2609	9.40	20.49	25.741	0.01788	0.01423
6	0.3548	1.3041	10.36	20.87	27.045	0.01700	0.01312
7	0.3501	1.2733	10.39	21.37	27.953	0.01638	0.01252
8	0.3426	1.2460	10.61	21.60	28.549	0.01586	0.01200
9	0.3334	1.2141	10.92	21.64	28.895	0.01541	0.01154
10	0.3243	....	....	....	....	....	....

variance multiplied by B_k must be divided by the sum of the fourth powers of certain binomial coefficients,[7] in order to give us the best estimate of the kurtosis. The reason is the same as for the division of the mean square of the differences by a binomial coefficient (page 40). These divisors appear in Table 15. We show in Table 16 the reciprocals of these sums, or a coefficient C_k by which the difference between the fourth moment and the square of the variance multiplied by B_k must be multiplied. The pertinent calculations are shown in Table 17 for our annual wheat-flour prices. The result is a number D_k which is the best estimate of the kurtosis of the kth difference.

D_k, the best estimate of the kurtosis for the first and higher differences, is calculated in the following way:

$$D_k = \frac{\dfrac{S_4^{(k)}}{N-k} - 3\,({}_{2k}C_k)^2\,V_k^2}{\sum\limits_{i=0}^{k} ({}_kC_i)^4} = \left[\frac{S_4^{(k)}}{N-k} - B_k\,V_k^2\right]C_k\,, \qquad k = 1, 2, \cdots.$$

For instance, we get from Table 8 for the annual wheat-flour prices for $k = 2$ the sum of the fourth powers of the second differences $S_4^{(2)} = 1{,}873.501$. Table 17 gives us for the same price series for $k=2$ the variance of the second differences as $V_k = 0.4402$. The multiplier B_k is, from Table 14 for $k=2$, shown as 108. Hence we get for the denominator of the expression which gives us the estimate of D_2, the kurtosis of the second differences, with $N-2 = 46$: $(1{,}873.501/46) - (108)(0.4402)^2 = 19.8005$. This is to be divided by the factor given in Table 15, for $k=2$, which is 18. The division $19.8005/18 = 1.1000$ gives us the exact value of D_2, the best estimate of the kurtosis of the second differences. We could also have multiplied the value 19.8005 by the factor given in Table 16, C_k for $k=2$, which is 0.055 555 5556. This gives exactly the same result. The result is shown in Table 17, together with estimates D_k for the kurtosis of the other differences of the wheat-flour prices.

The calculation for the original series is somewhat different from that for the series of differences.[8] First we need also the sum of the third powers of the original series $S_3^{(0)}$, since that is necessary for the calculation of the fourth moment (Table 8 for annual wheat-flour prices). It appears in our case to be equal to 15,497.211. We designate by m_4 the fourth moment about the mean. We have the well-known formula[9]

[7] *Ibid.*
[8] *Ibid.*, p. 78.
[9] H. T. Davis and W. F. C. Nelson, *op. cit.*, p. 81.

TABLE 18

COEFFICIENTS $E_N = 3\left(\frac{N-1}{N}\right)^2$

Number of Items in the Original Series N	E_N	D.D.*	Number of Items in the Original Series N	E_N	D.D.*
10	2.430 000	0.027 7500	350	2.982 882	0.000 04274
20	2.707 500	0.009 5834	400	2.985 019	0.000 03326
30	2.803 331	0.004 8544	450	2.986 682	0.000 02660
40	2.851 875	0.002 9325	500	2.988 012	0.000 02178
50	2.881 200	0.001 9633	550	2.989 101	0.000 01814
60	2.900 833	0.001 4065	600	2.990 008	0.000 01536
70	2.914 898	0.001 0571	650	2.990 776	0.000 01318
80	2.925 469	0.000 8235	700	2.991 435	0.000 01140
90	2.933 704	0.000 6596	750	2.992 005	0.000 01000
100	2.940 300	0.000 39666			
			800	2.992 505	0.000 00880
150	2.960 133	0.000 19884	850	2.992 945	0.000 00784
200	2.970 075	0.000 11946	900	2.993 337	0.000 00702
250	2.976 048	0.000 07972	950	2.993 688	0.000 00630
300	2.980 034	0.000 05696	1000	2.994 003	....

* Divided difference, positive.

$$m_4 = S_4^{(0)}/N - 4S_3^{(0)}\bar{w}/N + 6S_2^{(0)}\bar{w}^2/N - 3\bar{w}^4,$$

where $S_4^{(0)} = 135{,}336.771$ is the sum of the fourth powers, $S_3^{(0)} = 15{,}497.211$ the sum of the cubes, $S_2^{(0)} = 1{,}995.816$ the sum of the squares, and $\bar{w} = 6.0731$ the mean of the original items of the wheat-flour series (i.e., $S_1^{(0)}/N = 291.510/48$), given in Table 8.[10] The calculation gives for the fourth moment about the mean: $m_4 = (135{,}336.771)/48 - (4)(15{,}497.211)(6.0731)/48 + (6)(1{,}995.816)(6.0731)^2/48 - (3)(6.0731)^4 = 96.8859$.

This gives in our case $m_4 = 96.8859$. The formula for the kurtosis of the original series, D_0, is: $D_0 = (m_4 - E_N V_0^2)F_N$. We deduct from the fourth moment m_4 three times the square of the variance of the original series multiplied by a number which depends on N, the number of items in the original series. The coefficient $3[(N-1)/N]^2$ by which we have to multiply the square of the variance of the original series (V_0) is E_N. This is tabulated for selected values of N; further values can be found by interpolation (Table 18). In our example we

[10] R. A. Fisher, *loc. cit.*

TABLE 19

COEFFICIENTS $F_N = \dfrac{1}{1 - \dfrac{4}{N} + \dfrac{6}{N^2} - \dfrac{3}{N^3}}$

Number of Items in the Original Series N	F_N	D.D.*	Number of Items in the Original Series N	F_N	D.D.*
10	1.522 070	0.029 4511	350	1.011 511	0.000 02896
20	1.227 559	0.008 2375	400	1.010 063	0.000 02250
30	1.145 184	0.003 8626	450	1.008 938	0.000 01796
40	1.106 558	0.002 2401	500	1.008 040	0.000 01468
50	1.084 157	0.001 4622	550	1.007 306	0.000 01222
60	1.069 535	0.001 0295	600	1.006 695	0.000 01034
70	1.059 240	0.000 7640	650	1.006 178	0.000 00886
80	1.051 600	0.000 5894	700	1.005 735	0.000 00768
90	1.045 706	0.000 4687	750	1.005 351	0.000 00670
100	1.041 019	0.000 27804	800	1.005 016	0.000 00592
150	1.027 117	0.000 13730	850	1.004 720	0.000 00526
200	1.020 252	0.000 08182	900	1.004 457	0.000 00472
250	1.016 161	0.000 05432	950	1.004 221	0.000 00422
300	1.013 445	0.000 03868	1000	1.004 010	....

* Divided difference, negative.

have to multiply the square of the variance of the original series of wheat-flour prices, $V_0^2 = (4.7969)^2 = 23.0102$ (from Table 17), by the value E_N for $N = 48$ from Table 18, which is 2.875 335. This gives $E_N V_0^2 = (2.875\,335)(23.0102) = 66.1620$. Deducting this from $m_4 = 96.8859$ we get $m_4 - E_N V_0^2 = 96.8859 - 66.1620 = 30.7239$.

In order to get the best estimate of the kurtosis of the original series D_0, we have to divide this result by $1 - 4/N + 6/N^2 - 3/N^3$ for $N = 48$. We give in Table 19 the reciprocals of this value which we designate by F_N. The previous result $m_4 - E_N V_0^2$ has to be multiplied by F_N from Table 19 for $N = 48$ which is 1.088 637. This gives the best estimate for D_0. The kurtosis of the original series of wheat-flour prices, is $D_0 = (30.7239)(1.088\,637) = 33.4472$ and has been tabulated in Table 17.

The standard error of the difference between the variances of two consecutive series of finite differences involves a few more calculations.[11] We have first to find certain magnitudes G_k, which are the

[11] R. Zaycoff, *op. cit.*, p. 79 ff.

TABLE 20

COEFFICIENTS H_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k = 0$		$k = 1$		$k = 2$		$k = 3$	
N	H_{0N}	D.D.*	H_{1N}	D.D.*	H_{2N}	D.D.*	H_{3N}	D.D.*
10	2.875	0.1384	4.748	0.3140	5.287	0.4483	5.263	0.5486
20	4.259	0.1042	7.888	0.2354	9.770	0.3458	10.749	0.4401
30	5.301	0.0869	10.242	0.1948	13.228	0.2878	15.150	0.3710
40	6.170	0.0762	12.190	0.1694	16.106	0.2502	18.860	0.3244
50	6.932	0.0687	13.884	0.1517	18.608	0.2237	22.104	0.2907
60	7.619	0.0630	15.401	0.1385	20.845	0.2039	25.011	0.2650
70	8.249	0.0585	16.786	0.1282	22.884	0.1883	27.661	0.2449
80	8.834	0.0549	18.068	0.1198	24.767	0.1758	30.110	0.2284
90	9.383	0.0518	19.266	0.1130	26.525	0.1654	32.394	0.2147
100	9.901	0.04530	20.396	0.09818	28.179	0.14322	34.541	0.18562
150	12.166	0.03812	25.305	0.08212	35.340	0.11924	43.822	0.15402
200	14.072	0.03352	29.411	0.07420	41.302	0.10426	51.523	0.13436
250	15.748	0.03030	33.121	0.06276	46.515	0.09378	58.241	0.12062
300	17.263	0.02784	36.259	0.05958	51.204	0.08594	64.272	0.11038
350	18.655	0.02590	39.238	0.05536	55.501	0.07978	69.791	0.10236
400	19.950	0.02432	42.006	0.05194	59.490	0.07476	74.909	0.09582
450	21.166	0.02300	44.603	0.04910	63.228	0.07060	79.700	0.09044
500	22.316	0.02188	47.058	0.04664	66.758	0.06706	84.222	0.08586
550	23.410	0.02088	49.390	0.04454	70.111	0.06400	88.515	0.08188
600	24.454	0.02004	51.617	0.04270	73.311	0.06134	92.609	0.07842
650	25.456	0.01928	53.752	0.04108	76.378	0.05896	96.530	0.07538
700	26.420	0.01860	55.806	0.03962	79.326	0.05684	100.299	0.07266
750	27.350	0.01798	57.787	0.03828	82.168	0.05496	103.932	0.07020
800	28.249	0.01742	59.701	0.03712	84.916	0.05322	107.442	0.06798
850	29.120	0.01694	61.557	0.03602	87.577	0.05166	110.841	0.06596
900	29.967	0.01646	63.358	0.03502	90.160	0.05022	114.139	0.06414
950	30.790	0.01602	65.109	0.03412	92.671	0.04890	117.346	0.06240
1000	31.591	....	66.815	....	95.116	....	120.466	....

* Divided difference, positive.

TABLE 20 (continued)

COEFFICIENTS H_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k=4$		$k=5$		$k=6$		$k=7$	
N	H_{4N}	D.D*	H_{5N}	D.D*	H_{6N}	D.D*	H_{7N}	D.D*
10	4.985	0.6179	4.675	0.6539	4.893	0.6128	....	....
20	11.164	0.5188	11.214	0.5833	11.021	0.6357	10.662	0.6780
30	16.352	0.4443	17.047	0.5078	17.378	0.5624	17.442	0.6092
40	20.795	0.3916	22.125	0.4520	23.002	0.5054	23.534	0.5524
50	24.711	0.3526	26.645	0.4091	28.056	0.4604	29.058	0.5065
60	28.237	0.3223	30.736	0.3755	32.660	0.4244	34.123	0.4691
70	31.460	0.2952	34.491	0.3482	36.904	0.3947	38.814	0.4378
80	34.412	0.2814	37.973	0.3255	40.851	0.3699	43.192	0.4113
90	37.226	0.2617	41.228	0.3065	44.550	0.3488	47.305	0.3886
100	39.843	0.22620	44.293	0.26514	48.038	0.30244	51.191	0.33804
150	51.153	0.18738	57.550	0.21962	63.160	0.25084	68.093	0.28104
200	60.522	0.16316	68.531	0.19106	75.702	0.21818	82.145	0.24456
250	68.680	0.14628	78.084	0.17112	86.611	0.19528	94.373	0.21888
300	75.994	0.13370	86.640	0.15624	96.375	0.17820	105.317	0.19962
350	82.679	0.12384	94.452	0.14462	105.285	0.16482	115.298	0.18456
400	88.871	0.11588	101.683	0.13520	113.526	0.15400	124.526	0.17240
450	94.665	0.10920	108.443	0.12742	121.226	0.14506	133.146	0.16228
500	100.125	0.10372	114.814	0.12080	128.479	0.13748	141.260	0.15376
550	105.311	0.09884	120.854	0.11514	135.353	0.13094	148.948	0.14640
600	110.253	0.09462	126.611	0.11016	141.900	0.12528	156.268	0.14000
650	114.984	0.09088	132.119	0.10582	148.164	0.12026	163.268	0.13436
700	119.528	0.08760	137.410	0.10190	154.177	0.11554	169.986	0.12934
750	123.908	0.08460	142.505	0.09842	159.954	0.11206	176.453	0.12484
800	128.138	0.08190	147.426	0.09526	165.557	0.10818	182.695	0.12076
850	132.233	0.07946	152.189	0.09240	170.966	0.10490	188.733	0.11708
900	136.206	0.07722	156.809	0.08976	176.211	0.10188	194.587	0.11370
950	140.067	0.07514	161.297	0.08734	181.305	0.09912	200.272	0.11060
1000	143.824	....	165.664	....	186.261	....	205.802	....

* Divided difference, positive.

TABLE 20 (concluded)

COEFFICIENTS H_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k = 8$		$k = 9$		$k = 10$	
N	H_{8N}	D.D.*	H_{9N}	D.D.*	H_{10N}	D.D.*
10	....	....	....	....	....	....
20	10.191	0.7117	9.647	0.7378	9.061	0.7566
30	17.308	0.6491	17.025	0.6830	16.627	0.7120
40	23.799	0.5937	23.855	0.6300	23.747	0.6619
50	29.736	0.5479	30.155	0.5849	30.366	0.6178
60	35.215	0.5097	36.004	0.5465	36.544	0.5799
70	40.312	0.4774	41.469	0.5138	42.343	0.5469
80	45.086	0.4498	46.607	0.4853	47.812	0.5183
90	49.584	0.4259	51.460	0.4607	52.995	0.4930
100	53.843	0.37192	56.067	0.40402	57.925	0.43436
150	72.439	0.31022	76.268	0.33832	79.643	0.36532
200	87.950	0.27022	93.184	0.29518	97.909	0.31938
250	101.461	0.24192	107.943	0.26440	113.878	0.28634
300	113.557	0.22062	121.163	0.24118	128.195	0.26128
350	124.588	0.20392	133.222	0.22290	141.259	0.24152
400	134.784	0.19040	144.367	0.20808	153.335	0.22548
450	144.304	0.17920	154.771	0.19580	164.609	0.21210
500	153.264	0.16968	164.561	0.18538	175.214	0.20080
550	161.748	0.16156	173.830	0.17642	185.254	0.19108
600	169.826	0.15444	182.651	0.16864	194.808	0.18258
650	177.548	0.14818	191.083	0.16174	203.937	0.17510
700	184.957	0.14260	199.170	0.15564	212.692	0.16846
750	192.087	0.13762	206.952	0.15014	221.115	0.16250
800	198.968	0.13310	214.459	0.14520	229.240	0.15710
850	205.623	0.12900	221.719	0.14072	237.095	0.15222
900	212.073	0.12526	228.755	0.13658	244.706	0.14774
950	218.336	0.12180	235.584	0.13282	252.093	0.14364
1000	224.426	....	242.225	....	259.275	....

* Divided difference, positive.

TABLE 21

COEFFICIENTS J_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k = 0$		$k = 1$	
N	J_{0N}	D.D.*	J_{1N}	D.D.*
10	0.040 816	0.001 8203	0.082 610	0.003 2906
20	0.022 613	0.000 7023	0.049 704	0.001 4157
30	0.015 590	0.000 3700	0.035 547	0.000 7875
40	0.011 890	0.000 2282	0.027 672	0.000 5018
50	0.009 608	0.000 1548	0.022 654	0.000 3476
60	0.008 060	0.000 1118	0.019 178	0.000 2552
70	0.006 942	0.000 0846	0.016 626	0.000 1952
80	0.006 096	0.000 0663	0.014 674	0.000 1542
90	0.005 433	0.000 0532	0.013 132	0.000 1249
100	0.004 901	0.000 03224	0.011 883	0.000 07656
150	0.003 289	0.000 01628	0.008 055	0.000 03926
200	0.002 475	0.000 00984	0.006 092	0.000 02388
250	0.001 983	0.000 00654	0.004 898	0.000 01604
300	0.001 656	0.000 00472	0.004 096	0.000 01154
350	0.001 420	0.000 00352	0.003 519	0.000 00868
400	0.001 244	0.000 00276	0.003 085	0.000 00678
450	0.001 106	0.000 00220	0.002 746	0.000 00544
500	0.000 996	0.000 00180	0.002 474	0.000 00446
550	0.000 906	0.000 00150	0.002 251	0.000 00372
600	0.000 831	0.000 00128	0.002 065	0.000 00316
650	0.000 767	0.000 00110	0.001 907	0.000 00268
700	0.000 712	0.000 00094	0.001 773	0.000 00236
750	0.000 665	0.000 00084	0.001 655	0.000 00206
800	0.000 623	0.000 00072	0.001 552	0.000 00180
850	0.000 587	0.000 00066	0.001 462	0.000 00162
900	0.000 554	0.000 00058	0.001 381	0.000 00144
950	0.000 525	0.000 00052	0.001 309	0.000 00130
1000	0.000 499	....	0.001 244	....

* Divided difference, negative.

TABLE 21 (continued)

COEFFICIENTS J_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k = 2$		$k = 3$	
N	J_{2N}	D.D.*	J_{3N}	D.D.*
10	0.112 453	0.003 6115	0.116 206	0.003 0959
20	0.076 338	0.001 8766	0.085 247	0.001 8321
30	0.057 572	0.001 1374	0.066 926	0.001 1849
40	0.046 198	0.000 7623	0.055 077	0.000 8284
50	0.038 575	0.000 5465	0.046 793	0.000 6118
60	0.033 110	0.000 4109	0.040 675	0.000 4703
70	0.029 001	0.000 3202	0.035 972	0.000 3728
80	0.025 799	0.000 2564	0.032 244	0.000 3027
90	0.023 235	0.000 2102	0.029 217	0.000 2508
100	0.021 133	0.000 13162	0.026 709	0.000 16040
150	0.014 552	0.000 06910	0.018 689	0.000 08632
200	0.011 097	0.000 04258	0.014 373	0.000 05392
250	0.008 968	0.000 02888	0.011 677	0.000 03690
300	0.007 524	0.000 02086	0.009 832	0.000 02682
350	0.006 481	0.000 01580	0.008 491	0.000 02038
400	0.005 691	0.000 01234	0.007 472	0.000 01602
450	0.005 074	0.000 00994	0.006 671	0.000 01292
500	0.004 577	0.000 00818	0.006 025	0.000 01064
550	0.004 168	0.000 00682	0.005 493	0.000 00890
600	0.003 827	0.000 00580	0.005 048	0.000 00758
650	0.003 537	0.000 00496	0.004 669	0.000 00652
700	0.003 289	0.000 00434	0.004 343	0.000 00566
750	0.003 072	0.000 00378	0.004 060	0.000 00498
800	0.002 883	0.000 00336	0.003 811	0.000 00440
850	0.002 715	0.000 00298	0.003 591	0.000 00390
900	0.002 566	0.000 00266	0.003 396	0.000 00352
950	0.002 433	0.000 00242	0.003 220	0.000 00316
1000	0.002 312	....	0.003 062	....

* Divided difference, negative.

TABLE 21 (continued)

COEFFICIENTS J_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k = 4$		$k = 5$	
N	J_{4N}	D.D.*	J_{5N}	D.D.*
10	0.117 720	0.002 5410	0.113 492	0.001 8215
20	0.092 310	0.001 7308	0.095 277	0.001 5676
30	0.075 002	0.001 1856	0.079 601	0.001 1260
40	0.063 146	0.000 8617	0.068 341	0.000 8466
50	0.054 529	0.000 6547	0.059 875	0.000 6596
60	0.047 982	0.000 5143	0.053 279	0.000 5285
70	0.042 839	0.000 4146	0.047 994	0.000 4331
80	0.038 693	0.000 3415	0.043 663	0.000 3613
90	0.035 278	0.000 2860	0.040 050	0.000 3061
100	0.032 418	0.000 18702	0.036 989	0.000 20448
150	0.023 067	0.000 10328	0.026 765	0.000 11592
200	0.017 903	0.000 06548	0.020 969	0.000 07464
250	0.014 629	0.000 04524	0.017 237	0.000 05208
300	0.012 367	0.000 03312	0.014 633	0.000 03842
350	0.010 711	0.000 02530	0.012 712	0.000 02950
400	0.009 446	0.000 01996	0.011 237	0.000 02336
450	0.008 448	0.000 01614	0.010 069	0.000 01896
500	0.007 641	0.000 01320	0.009 121	0.000 01570
550	0.006 981	0.000 01132	0.008 336	0.000 01322
600	0.006 415	0.000 00952	0.007 675	0.000 01126
650	0.005 939	0.000 00822	0.007 112	0.000 00974
700	0.005 528	0.000 00714	0.006 625	0.000 00848
750	0.005 171	0.000 00628	0.006 201	0.000 00746
800	0.004 857	0.000 00556	0.005 828	0.000 00662
850	0.004 579	0.000 00496	0.005 497	0.000 00590
900	0.004 331	0.000 00444	0.005 202	0.000 00530
950	0.004 109	0.000 00402	0.004 937	0.000 00480
1000	0.003 908	....	0.004 697	....

* Divided difference, negative.

TABLE 21 (continued)

COEFFICIENTS J_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k = 6$		$k = 7$	
N	J_{6N}	D.D.*	J_{7N}	D.D.*
10	0.097 924	0.000 1355	0.169 866	0.007 3269
20	0.096 569	0.001 4017	0.096 597	0.001 2439
30	0.082 552	0.001 0483	0.084 158	0.000 9635
40	0.072 069	0.000 8113	0.074 523	0.000 7644
50	0.063 956	0.000 6467	0.066 879	0.000 6215
60	0.057 489	0.000 5276	0.060 664	0.000 5154
70	0.052 213	0.000 4388	0.055 510	0.000 4345
80	0.047 825	0.000 3707	0.051 165	0.000 3713
90	0.044 118	0.000 3173	0.047 452	0.000 3210
100	0.040 945	0.000 21654	0.044 242	0.000 22356
150	0.030 118	0.000 12594	0.033 064	0.000 13334
200	0.023 821	0.000 08238	0.026 397	0.000 08858
250	0.019 702	0.000 05810	0.021 968	0.000 06312
300	0.016 797	0.000 04316	0.018 812	0.000 04728
350	0.014 639	0.000 03332	0.016 448	0.000 03670
400	0.012 973	0.000 02652	0.014 613	0.000 02934
450	0.011 647	0.000 02160	0.013 146	0.000 02400
500	0.010 567	0.000 01794	0.011 946	0.000 01996
550	0.009 670	0.000 01512	0.010 948	0.000 01690
600	0.008 914	0.000 01294	0.010 103	0.000 01448
650	0.008 267	0.000 01120	0.009 379	0.000 01254
700	0.007 707	0.000 00976	0.008 752	0.000 01096
750	0.007 219	0.000 00860	0.008 204	0.000 00968
800	0.006 789	0.000 00762	0.007 720	0.000 00858
850	0.006 408	0.000 00684	0.007 291	0.000 00770
900	0.006 066	0.000 00612	0.006 906	0.000 00692
950	0.005 760	0.000 00554	0.006 560	0.000 00626
1000	0.005 483	....	0.006 247	....

* Divided difference, negative.

TABLE 21 (continued)

COEFFICIENTS J_{kN}
(k = Order of Difference)

Number of Items in the Original Series	$k = 8$		$k = 9$	
N	J_{8N}	D.D.*	J_{9N}	D.D.*
10	0.131 016	0.003 5118	....	....
20	0.095 898	0.001 1005	0.094 720	0.000 9709
30	0.084 893	0.000 8802	0.085 011	0.000 8018
40	0.076 091	0.000 7132	0.076 993	0.000 6618
50	0.068 959	0.000 5901	0.070 375	0.000 5559
60	0.063 058	0.000 4966	0.064 816	0.000 4739
70	0.058 092	0.000 4238	0.060 077	0.000 4089
80	0.053 854	0.000 3660	0.055 988	0.000 3567
90	0.050 194	0.000 3193	0.052 421	0.000 3137
100	0.047 001	0.000 22676	0.049 284	0.000 22692
150	0.035 663	0.000 13858	0.037 938	0.000 14194
200	0.028 734	0.000 09346	0.030 841	0.000 09718
250	0.024 061	0.000 06732	0.025 982	0.000 07072
300	0.020 695	0.000 05080	0.022 446	0.000 05376
350	0.018 155	0.000 03968	0.019 758	0.000 04228
400	0.016 171	0.000 03186	0.017 644	0.000 03410
450	0.014 578	0.000 02616	0.015 939	0.000 02808
500	0.013 270	0.000 02184	0.014 535	0.000 02354
550	0.012 178	0.000 01852	0.013 358	0.000 02002
600	0.011 252	0.000 01590	0.012 357	0.000 01722
650	0.010 457	0.000 01382	0.011 496	0.000 01498
700	0.009 766	0.000 01208	0.010 747	0.000 01314
750	0.009 162	0.000 01068	0.010 090	0.000 01164
800	0.008 628	0.000 00952	0.009 508	0.000 01036
850	0.008 152	0.000 00850	0.008 990	0.000 00930
900	0.007 727	0.000 00768	0.008 525	0.000 00838
950	0.007 343	0.000 00694	0.008 106	0.000 00758
1000	0.006 996	....	0.007 727	....

* Divided difference, negative.

TABLE 21 (concluded)

COEFFICIENTS J_{kN}
(k = Order of Difference)

Series of Items in the Original Number	$k = 10$		Number of Items in the Original Series	$k = 10$	
N	J_{10N}	D.D.*	N	J_{10N}	D.D.*
10	....	....	350	0.021 259	0.000 04448
20	0.093 223	0.000 8515	400	0.019 035	0.000 03606
30	0.084 708	0.000 7266	450	0.017 232	0.000 02982
40	0.077 442	0.000 6157	500	0.015 741	0.000 02506
50	0.071 285	0.000 5211			
			550	0.014 488	0.000 02138
60	0.066 074	0.000 4494	600	0.013 419	0.000 01844
70	0.061 580	0.000 3916	650	0.012 497	0.000 01606
80	0.057 664	0.000 3446	700	0.011 694	0.000 01412
90	0.054 218	0.000 3056	750	0.010 988	0.000 01252
100	0.051 162	0.000 22480			
			800	0.010 362	0.000 01116
150	0.039 922	0.000 14374	850	0.009 804	0.000 01002
200	0.032 735	0.000 09986	900	0.009 303	0.000 00906
250	0.027 742	0.000 07342	950	0.008 850	0.000 00820
300	0.024 071	0.000 05624	1000	0.008 440	....

* Divided difference, negative.

kurtosis D_k divided by the square of the variance of the series of finite differences in question V_k^2. We give the quantities G_k also in Table 17 for our example: $G_k = D_k/V_k^2$. We have, for instance, for wheat-flour prices in Table 17, the variance and kurtosis of the fifth difference $V_5 = 0.3662$ and $D_5 = 1.2609$. We hence get for the $G_5 = D_5/V_5^2 = 1.2609/(0.3662)^2 = 9.40$. This value and the other values of the G_k for the annual wheat-flour prices are given in Table 17.

We next form quantities Q_k which are defined by the formula: $Q_k = H_{kN}/\sqrt{1 + J_{kN}G_k}$. The magnitudes H_{kN} and J_{kN} have been tabulated in Tables 20 and 21. They are given for every difference from $k = 0$ to $k = 10$ and for selected values of N, the number of items in the original series. The tables can also be interpolated and divided differences are provided for this purpose. The calculation method for our example, wheat-flour prices, is shown in Table 17.

To show the process of calculation, we shall indicate the operations for the calculation of the value Q_6 of our wheat-flour prices. We have $G_6 = D_6/V_6^2 = 1.3041/(0.3548)^2 = 10.36$ from Table 17 for $k = 6$. By interpolation we find for H_{kN} in Table 20 the value 27.045

for $N = 48$ and $k = 6$ and similarly from Table 21 the value 0.065 579 for J_{kN}, $k=6$, $N=48$. By the formula given above, we then calculate $Q_6 = 27.045/\sqrt{1 + (0.065\,579)\,(10.36)} = 20.87$. This value and other values of the Q_k for the wheat-flour prices are shown in Table 17.

The standard error e_k of the difference between the variance of the kth difference and the $(k+1)$th difference series, $V_k - V_{k+1}$, is equal to the variance of the lower series of differences V_k divided by Q_k or $e_k = V_k/Q_k$. Take for instance again our annual wheat-flour prices. We get from Table 17 for $k=7$, $Q_7 = 21.37$, and the variance of the seventh differences, $V_7 = 0.3501$. Hence the standard error for the difference between the variances of the seventh and eighth differences, $V_7 - V_8$, is $e_7 = V_7/Q_7 = 0.3501/21.37 = 0.01638$. This is given in Table 17 together with other e_k.

If the kurtosis is small and we think we are near enough to a normal distribution in order to be able to neglect it, we can use an approximate error instead of the correct value of the standard error indicated above. We substitute the numerator of Q_k, that is, H_{kN}, for Q_k itself and get for the approximate standard error $e_k^0 = V_k/H_{kN}$. For instance, again for our annual wheat-flour prices we have from Table 17: $V_8 = 0.3426$; and by interpolation from Table 20 for $N = 48$, $k = 8$: $H_{8N} = 28.549$. This gives us for the approximate standard error of the difference between the variances of the eighth and ninth finite difference series: $e_8^0 = V_8/H_{8N} = 0.3426/28.549 = 0.01200$. This value and other values of e_k^0 are also given in Table 17.

CHAPTER VII

Criteria for the Stability of the Variances of the Series of Finite Differences

We have shown in the last chapter how it is possible to find the standard error of the difference between the variances of two consecutive series of finite differences. From the point of view of probability, it appears that these variances become stable in long series if consecutive variances do not differ from each other by more than about three times the standard error of the difference (see above, pages 33 f.). Criteria have been given for this by Professor O. Anderson[1] of Sofia, Bulgaria, and they have been recently improved by Dr. R. Zaycoff[2] of Sofia. We are going to use in this connection Dr. Zaycoff's first criterion, which he himself considers the better one.

The comparison is made in the following way. We form the criterion (standard-error ratio):

$$R_k = \frac{(V_k - V_{k+1})}{V_k} Q_k .$$

This is the difference between two consecutive variances divided by its standard error. The value of the difference between two consecutive variances of the series of finite differences (for instance the difference between the variances of the series of first finite differences and the series of second finite differences, $V_1 - V_2$) is divided by the variance of the difference of lower order (in our example by the variance of the series of the first differences, V_1). The result is multiplied by the quantity Q_k with the index which corresponds to the lower of the series of finite differences (in our case by Q_1). We shall use the quantities Q_k rather than the approximations H_{kN} if we wish to be quite accurate and if our distributions show considerable kurtosis. We do this in our case for all the variances of the series of finite differences of wheat-flour prices up to the tenth. The last criterion R_9, for instance, is the value of the difference between the variances of the ninth and tenth finite differences, $V_9 - V_{10}$, divided by the vari-

[1] O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, Bonn, 1929, pp. 56 ff., 112 ff. See also below, Appendix II, pp. 139 ff., and Appendix I, Sections C and D, for a summary of computations.

[2] R. Zaycoff, "Ueber die Ausschaltung der Zufälligen Komponente nach der 'Variate-Difference' Methode," *Publications of the Statistical Institute for Economic Research*, State University of Sofia, 1937, No. 1, pp. 78 ff.

ance of the ninth finite difference, V_9, and multiplied by the quantity Q_9 as described in the previous chapter (Table 22).

We have again in our example of annual wheat-flour prices, from Table 17: $V_9 = 0.3334$ and $V_{10} = 0.3243$. We get from Table 22: $Q_9 = 21.64$. Hence the criterion R_9 for the difference between the variances of the ninth and tenth differences is: $R_9 = (V_9 - V_{10})Q_9/V_9 = (0.3334 - 0.3243)(21.64)/0.3334 = 0.5906$. This value and all other values for the exact criterion R_k for the annual wheat-flour prices are tabulated in Table 22.

TABLE 22

DIFFERENCE ANALYSIS
ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Order of Difference	Variance	Kurtosis		Standard Error Ratio	Approximate Standard Error Ratio
k	V_k	D_k	Q_k	R_k	R^0_k
0	4.7969	33.4473	6.73	5.7459	5.7878
1	0.7020	4.4384	12.30	4.5863	5.0514
2	0.4402	1.1000	16.34	1.7488	1.9368
3	0.3931	1.1920	18.30	0.7637	0.8952
4	0.3767	1.1974	19.70	0.5492	0.6668
5	0.3662	1.2609	20.49	0.6376	0.8012
6	0.3548	1.3041	20.87	0.2764	0.3582
7	0.3501	1.2733	21.37	0.4579	0.5987
8	0.3426	1.2460	21.60	0.5800	0.7668
9	0.3334	1.2141	21.64	0.5906	0.7885
10	0.3243	....	....	....	....

Those criteria R_0, R_1, R_2, etc., are arranged in a series. We are going to consider as reasonably stable the variance beginning from which the R_k becomes numerically smaller than 3 and stays more or less so. The R's for our annual wheat-flour prices are given in Table 22, where also some of the calculations are exhibited. Whereas R_0 and R_1 are greater than 4, R_2 and all the following parameters are certainly smaller than 2 and become very small indeed for higher differences. Hence we can conclude that we have eliminated the mathematical expectation or the nonrandom element of our wheat-flour prices with reasonable accuracy in the second or third difference in the particular case under consideration. We can say that in all probability beginning with the second or third difference there are left only some remainders of the nonrandom element which we can neglect for our purposes (see above, pages 33 f.). An alternative procedure is given in Chapter VIII.

We shall now try to find the best approximation to the true random variance of the original series. In order to do this we take the variance of the series of finite differences beginning from which we can be reasonably sure that we have eliminated the nonrandom element to the desired degree. In our case, for example, we should take the variance of the second finite difference, which is equal to 0.4402 (Table 22). We could say that this probably represents a good approximation to the true variance of the pure random element in our time series, as defined above (Chapter I, Section C).

If we desire to avoid a great deal of calculation and if we have reason to think that the kurtosis of our series is not very great, we may use the approximations H_{kN} instead of the better estimates Q_k. We shall then form a criterion (approximate standard-error ratio):

$$R^0_k = \frac{(V_k - V_{k+1})}{V_k} H_{kN} .$$

The procedure is also shown in Table 22. It is the same as the former, except that we use the approximate instead of the more accurate estimates of the standard errors.

In order to show the calculations, let us use, for instance, our previous example and calculate the approximate criterion, R^0_4, for the difference between the variances of the fourth and fifth difference series of annual wheat-flour prices. We have from Table 22: $V_4 = 0.3767$ and $V_5 = 0.3662$ as estimates for the two variances. Table 20 gives for $k = 4$ and $N = 48$, by interpolation, $H_{4N} = 23.928$. Hence the approximate criterion $R^0_4 = (0.3767 - 0.3662)(23.928)/0.3767 = 0.6668$. The values of this approximate criterion R^0_k are also given in Table 22. They yield the same result as before. It appears that we probably eliminate our mathematical expectation or nonrandom element in the second or third difference, since R^0_2 and the following values are all smaller than 2.

We have accomplished our goal, for many purposes, if we reach this stage of our analysis. If we want, for instance, to make a comparison of our annual wheat-flour prices with some other statistical data, we must know the limits of errors and inaccuracies involved in this statistical comparison. We have seen that it is very improbable that we shall get a deviation from the true value that is greater than three times the standard deviation of the random element, which is approximately 0.66. We are going to be wrong in only about three cases out of a thousand if we indicate the probable limits of a deviation due to the random variation from the mathematical expectation by $\pm 3\sqrt{V_2} = \pm 1.99$. (See above, pages 33 f.)

This is extremely important if we wish to compare two statistical series and if we desire to know whether differences can be explained by the random variation. The same can easily be done for certain derived statistical series which we calculate from our original series. We reproduce, in Appendix IV, the standard errors of some statistical series, some of which have been given previously by the author.[3] They should facilitate the comparison of certain parameters calculated from the original series by giving their standard errors. It is, for instance, possible to indicate the standard error resulting from the random variation that is retained after calculating a moving average, a seasonal index, a trend, etc. All those statistical estimates are most important because they give the desired statistical measurements for definite economic purposes. They should always be treated from the point of view of probability. The variate difference method can give us a reasonably good estimate of the random variation in those parameters which is due to the erratic element as defined above. It enables us also to make statistical tests of hypotheses.

The same type of analysis is shown in Tables 23 to 25 for the other prices. We show the difference analysis for annual wool prices in Table 23. Both the accurate criterion R_k and the approximate criterion R_k^0 indicate that we have already eliminated the nonrandom ele-

TABLE 23

DIFFERENCE ANALYSIS
ANNUAL WOOL PRICES, 1890-1937

Order of Difference	Variance	Kurtosis		Standard Error Ratio	Approximate Standard Error Ratio
k	V_k	D_k	Q_k	R_k	R_k^0
0	0.1069	0.020 93	6.72	4.9775	5.0234
1	0.02769	0.009 810	11.87	0.7671	0.8758
2	0.02590	0.006 434	15.39	–0.2020	–0.2377
3	0.02624	0.007 861	17.22	–0.0918	–0.1145
4	0.02638	0.008 374	18.47	0.0840	0.1088
5	0.02626	0.008 802	19.26	0.1981	0.2647
6	0.02600	0.008 990	19.76	0.1673	0.2289
7	0.02577	0.008 934	20.17	0.1879	0.2603
8	0.02553	0.008 739	20.48	0.2727	0.3802
9	0.02519	0.008 493	20.64	0.2622	0.3671
10	0.02487	....	......	....	....

[3] G. Tintner, *Prices in the Trade Cycle*, Vienna, 1935, pp. 81 ff.

ment in the first or second finite differences, since both become, and stay, smaller than 1 in absolute value, i.e., without regard to sign, for $k = 2, 3$, etc.

TABLE 24

DIFFERENCE ANALYSIS
MONTHLY WOOL PRICES, 1890-1937

Order of Difference	Variance	Kurtosis		Standard Error Ratio	Approximate Standard Error Ratio
k	V_k	D_k	Q_k	R_k	R^0_k
0	0.112 1	0.047 6626	23.87	23.4370	23.5141
1	0.002 054	0.000 9100	41.75	16.6066	20.1061
2	0.001 237	0.000 1925	58.56	6.6750	8.1811
3	0.001 096	0.000 1904	66.93	4.0912	5.5411
4	0.001 029	0.000 1559	76.57	3.7434	5.2731
5	0.000 9787	0.000 1444	83.41	2.8804	4.2771
6	0.000 9449	0.000 1341	89.69	2.1642	3.3485
7	0.000 9221	0.000 1269	95.30	1.7775	2.8489
8	0.000 9049	0.000 1213	100.38	1.3979	2.3108
9	0.000 8923	0.000 1170	105.01	0.9650	1.6396
10	0.000 8841	....	....	....	....

Table 24 gives the same type of analysis for monthly wool prices. The accurate criterion would indicate that the mathematical expectation has been eliminated in the fifth difference, whereas the series

TABLE 25

DIFFERENCE ANALYSIS
ANNUAL RAW-SILK PRICES, 1890-1937

Order of Difference	Variance	Kurtosis		Standard Error Ratio	Approximate Standard Error Ratio
k	V_k	D_k	Q_k	R_k	R^0_k
0	2.8914	5.020 9	6.76	5.8598	5.8775
1	0.3849	0.429 2	13.10	3.8639	3.9943
2	0.2714	0.163 4	17.35	2.5383	2.6489
3	0.2317	0.086 24	20.67	2.5330	2.6297
4	0.2033	0.047 31	23.19	2.3842	2.4599
5	0.1824	0.012 60	25.45	2.2599	2.2862
6	0.1662	–0.004 80	27.20	2.1113	2.0993
7	0.1533	–0.016 70	28.66	2.0750	2.0241
8	0.1422	–0.021 80	29.70	2.0464	1.9677
9	0.1324	–0.023 45	30.39	1.7213	1.6367
10	0.1249	....	....	....	....

of the R_k^0 becomes numerically smaller than 3 only for $k = 7$ and higher, indicating that we succeed only in the seventh difference in getting rid of the nonrandom element.

Table 25 gives the difference analysis for the annual raw-silk prices. Both the exact criterion R_k and the approximation R_k^0 point to the second difference as the difference in which only traces of the nonrandom or smooth element remain.

Table 24 gives as an estimate of the true random variance of the monthly wool prices: $V_6 = 0.000\,9449$. From Table 23 for the yearly wool prices: $V_1 = 0.02769$. The random variance of the annual prices is here almost 30 times as great as that of the monthly prices.

The explanation of this may be attempted in the following way: some economic causes, which may appear permanent from the point of view of the shorter run, i.e., in the monthly series, become nonpermanent and hence part of the random element from the point of view of a longer run, i.e., in the annual series. (See also above, Chapter I, Section C.)

CHAPTER VIII

A Test of Significance for the Stability of Variances of the Series of Finite Differences[1]

The variate difference method as developed by Anderson[2] gives only asymptotic formulae for the estimated error of the difference between the variances of two consecutive finite differences of the original series (see Chapter VI). Those formulae hold true only for large samples.[3] It should be remembered that Anderson's work in this field dates back as far as 1914 and hence has not taken into account the modern ideas of exact tests of significance as developed by R. A. Fisher and his school. In the following we shall make an attempt to apply the Fisher approach to our problem and meet in this way some of the criticism which has been leveled against the variate difference method.

The asymptotic formulae given by Anderson (Chapters VI, VII) permit us a direct comparison of the variances of two consecutive differences and they utilize all the differences which can be calculated. We now make the assumption that the random element in the original series is normally distributed. We could find an exact test of significance for the ratio of the variances of two consecutive series of differences, if we knew the distribution of this ratio for correlated observations.[4] It is clear that even if the items of the original random element are normally distributed and mutually independent we have introduced serial correlations between several consecutive items of the series of differences by forming finite differences. Unfortunately, this

[1] The author wants to express his thanks to Professor Harold Hotelling, Columbia University, Professor S. S. Wilks, Princeton University, Professor G. W. Snedecor, Iowa State College, Mr. W. G. Cochran, Rothamsted, and Dr. W. G. Madow, Columbia University for their kind help and valuable suggestions in connection with the subject treated in this chapter. See, on the general content of this chapter, G. Tintner, "On Tests of Significance in Time Series," *Annals of Mathematical Statistics*, Vol. 10, 1939, pp. 139 ff. See also below, Appendix II, pp. 142 ff., and Appendix I, Section E, for a summary of computations.

[2] O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, Bonn, 1929, pp. 54 ff.

[3] R. A. Fisher, *Statistical Methods for Research Workers*, 7th ed., London, 1938, pp. 42 ff., Chapter III, Section 11.

[4] See M. S. Bartlett and J. Wishart, "The Distribution of Second Order Moment Statistics in a Normal System," *Proceedings of the Cambridge Philosophical Society*, Vol. 28, 1932, pp. 455 ff.; W. G. Cochran, "The Distribution of Quadratic Forms in a Normal System with Applications to the Analysis of Covariance," *ibid.*, Vol. 30, 1934, pp. 178 ff.

problem has not yet been solved and hence we cannot apply an exact test of significance which would utilize all our material.

The problem can be solved, however, by a method which is not efficient[5] in the sense that it utilizes all the information available, but which gives an unbiased estimate. It is consistent.[6] We will call this method the method of selection. We select from our difference series the items which are independent and utilize only the uncorrelated differences for the calculation of the variances and for their comparison.

To make this clear let us consider an example. Suppose we want to compare the variance of our original series with the variance of the first finite differences in order to see whether the "smooth" element has already been eliminated by taking the first differences. We cannot use all the original items and all the first finite differences because this would mean comparing the variances of elements which are not independent. By calculating the first finite difference we have introduced a serial correlation between two consecutive items of the series of first differences.

But we can do something else. We can take the first item, the fourth, the seventh, the tenth, etc. of our original series and compare the variance of those selected items with the variance of the second item, the fifth, eighth, and the eleventh item, etc. of the series of the first differences. By making these two selections we are really comparing the variances of independent elements. We have made our selection in such a way as to make the elements which enter into the variances independent. But we have by this method reduced the number of items from which we estimate our variances to one-third of the number originally contained in the series. The number of selected items should be the same in each series of differences.

We will call this selection "selection number 0-A" because we compare the original series, that is, the 0th difference, with the first differences. Instead of making this particular selection, we could also have selected the items number 2, 5, 8, etc. of the original series and the items 3, 6, 9, etc. of the series of the first differences. This we will call the selection 0-B. Finally, we could have taken the items number 3, 6, 9, etc. of the original series and the items number 4, 7, 10, etc. of the series of first differences, which would have given us the selection 0-C. Zero is the "order" of the selection (Table 26).

If we want to get an unbiased comparison of the variances of the first differences and of the second differences we can make five selec-

[5] R. A. Fisher, *op. cit.*, pp. 14 ff., Chapter I, Section 3.
[6] *Ibid.*, pp. 12 ff., Chapter I, Section 3.

TABLE 26

SELECTIONS FOR COMPARISON OF THE ORIGINAL AND TEN DIFFERENCES

Selection	0th Difference	1st Difference
0-A	1-4-7-10	2-5-8-11
0-B	2-5-8-11	3-6-9-12
0-C	3-6-9-12	4-7-10-13

Selection	1st Difference	2nd Difference
1-A	1-6-11-16	3-8-13-18
1-B	2-7-12-17	4-9-14-19
1-C	3-8-13-18	5-10-15-20
1-D	4-9-14-19	6-11-16-21
1-E	5-10-15-20	7-12-17-22

Selection	2nd Difference	3rd Difference
2-A	1-8-15-22	4-11-18-25
2-B	2-9-16-23	5-12-19-26
2-C	3-10-17-24	6-13-20-27
2-D	4-11-18-25	7-14-21-28
2-E	5-12-19-26	8-15-22-29
2-F	6-13-20-27	9-16-23-30
2-G	7-14-21-28	10-17-24-31

Selection	3rd Difference	4th Difference
3-A	1-10-19-28	5-14-23-32
3-B	2-11-20-29	6-15-24-33
3-C	3-12-21-30	7-16-25-34
3-D	4-13-22-31	8-17-26-35
3-E	5-14-23-32	9-18-27-36
3-F	6-15-24-33	10-19-28-37
3-G	7-16-25-34	11-20-29-38
3-H	8-17-26-35	12-21-30-39
3-I	9-18-27-36	13-22-31-40

Selection	4th Difference	5th Difference
4-A	1-12-23-34	6-17-28-39
4-B	2-13-24-35	7-18-29-40
4-C	3-14-25-36	8-19-30-41
4-D	4-15-26-37	9-20-31-42
4-E	5-16-27-38	10-21-32-43
4-F	6-17-28-39	11-22-33-44
4-G	7-18-29-40	12-23-34-45
4-H	8-19-30-41	13-24-35-46
4-I	9-20-31-42	14-25-36-47
4-J	10-21-32-43	15-26-37-48
4-K	11-22-33-44	16-27-38-49

Selection	5th Difference	6th Difference
5-A	1-14-27-40	7-20-33-46
5-B	2-15-28-41	8-21-34-47
5-C	3-16-29-42	9-22-35-48
5-D	4-17-30-43	10-23-36-49
5-E	5-18-31-44	11-24-37-50
5-F	6-19-32-45	12-25-38-51
5-G	7-20-33-46	13-26-39-52
5-H	8-21-34-47	14-27-40-53
5-I	9-22-35-48	15-28-41-54
5-J	10-23-36-49	16-29-42-55
5-K	11-24-37-50	17-30-43-56
5-L	12-25-38-51	18-31-44-57
5-M	13-26-39-52	19-32-45-58

Selection	6th Difference	7th Difference
6-A	1-16-31-46	8-23-38-53
6-B	2-17-32-47	9-24-39-54
6-C	3-18-33-48	10-25-40-55
6-D	4-19-34-49	11-26-41-56
6-E	5-20-35-50	12-27-42-57
6-F	6-21-36-51	13-28-43-58
6-G	7-22-37-52	14-29-44-59
6-H	8-23-38-53	15-30-45-60
6-I	9-24-39-54	16-31-46-61
6-J	10-25-40-55	17-32-47-62
6-K	11-26-41-56	18-33-48-63
6-L	12-27-42-57	19-34-49-64
6-M	13-28-43-58	20-35-50-65
6-N	14-29-44-59	21-36-51-66
6-O	15-30-45-60	22-37-52-67

Selection	7th Difference	8th Difference
7-A	1-18-35-52	9-26-43-60
7-B	2-19-36-53	10-27-44-61
7-C	3-20-37-54	11-28-45-62
7-D	4-21-38-55	12-29-46-63
7-E	5-22-39-56	13-30-47-64
7-F	6-23-40-57	14-31-48-65
7-G	7-24-41-58	15-32-49-66
7-H	8-25-42-59	16-33-50-67
7-I	9-26-43-60	17-34-51-68
7-J	10-27-44-61	18-35-52-69
7-K	11-28-45-62	19-36-53-70

TABLE 26 (concluded)

SELECTIONS FOR COMPARISON OF THE ORIGINAL AND TEN DIFFERENCES

Selection	7th Difference	8th Difference
7-L	12-29-46-63	20-37-54-71
7-M	13-30-47-64	21-38-55-72
7-N	14-31-48-65	22-39-56-73
7-O	15-32-49-66	23-40-57-74
7-P	16-33-50-67	24-41-58-75
7-Q	17-34-51-68	25-42-59-76
Selection	**8th Difference**	**9th Difference**
8-A	1-20-39-58	10-29-48-67
8-B	2-21-40-59	11-30-49-68
8-C	3-22-41-60	12-31-50-69
8-D	4-23-42 61	13-32-51-70
8-E	5-24-43-62	14-33-52-71
8-F	6-25-44-63	15-34-53-72
8-G	7-26-45-64	16-35-54-73
8-H	8-27-46-65	17-36-55-74
8-I	9-28-47-66	18-37-56-75
8-J	10-29-48-67	19-38-57-76
8-K	11-30-49-68	20-39-58-77
8-L	12-31-50-69	21-40-59-78
8-M	13-32-51-70	22-41-60-79
8-N	14-33-52-71	23-42-61-80
8-O	15-34-53-72	24-43-62-81
8-P	16-35-54-73	25-44-63-82
8-Q	17-36-55-74	26-45-64-83

Selection	8th Difference	9th Difference
8-R	18-37-56-75	27-46-65-84
8-S	19-38-57-76	28-47-66-85
Selection	**9th Difference**	**10th Difference**
9-A	1-22-43-64	11-32-53-74
9-B	2-23-44-65	12-33-54-75
9-C	3-24-45-66	13-34-55-76
9-D	4-25-46-67	14-35-56-77
9-E	5-26-47-68	15-36-57-78
9-F	6-27-48-69	16-37-58-79
9-G	7-28-49-70	17-38-59-80
9-H	8-29-50-71	18-39-60-81
9-I	9-30-51-72	19-40-61-82
9-J	10-31-52-73	20-41-62-83
9-K	11-32-53-74	21-42-63-84
9-L	12-33-54-75	22-43-64-85
9-M	13-34-55-76	23-44-65-86
9-N	14-35-56-77	24-45-66-87
9-O	15-36-57-78	25-46-67-88
9-P	16-37-58-79	26-47-68-89
9-Q	17-38-59-80	27-48-69-90
9-R	18-39-60-81	28-49-70-91
9-S	19-40-61-82	29-50-71-92
9-T	20-41-62-83	30-51-72-93
9-U	21-42-63-84	31-52-73-94

tions. Selection 1-A takes items number 1, 6, 11, etc. of the first differences and the items 3, 8, 13, etc. of the second differences. Selection 1-B has items number 2, 7, 12, etc. of the series of first differences and items number 4, 9, 14, etc. of the series of second differences. There are 5 possible selections of order one which are tabulated in Table 26. This table gives all selections which will give independent estimates of the variances of two consecutive differences. But two selections of the same order, say 1-A and 1-B, are not independent.

We use Fisher's z test[7] to compare the variances of selected items of two consecutive differences. We have, however, facilitated the comparison somewhat by using the variance ratio or Snedecor's F[8] instead of the z which would necessitate a logarithmic transformation. Further, we have adjusted our comparison to the testing of the ratio

[7] *Ibid.*, pp. 232 ff., Chapter VII, Section 41; G. W. Snedecor, *Statistical Methods*, Ames, Iowa, 1938, pp. 182 ff.

[8] G. W. Snedecor, *loc. cit.* See also R. A. Fisher and F. Yates, *Statistical Tables*, London, 1938, pp. 28 ff.

TABLE 27

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 5%
Order of Difference: k

Number of Items in the Original Series	$k = 0$				$k = 1$			
N	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	0.052	0.0053	4.792	0.2413	0.0150	0.00271	7.399	0.4757
20	0.105	0.0037	2.379	0.0616	0.0421	0.00213	2.642	0.0889
30	0.142	0.0026	1.763	0.0276	0.0634	0.00172	1.753	0.0374
40	0.168	0.0022	1.487	0.0168	0.0806	0.00140	1.379	0.0204
50	0.190	0.0017	1.319	0.0114	0.0946	0.00114	1.175	0.0122
60	0.207	0.0013	1.205	0.0070	0.106	0.0010	1.053	0.0091
70	0.220	0.0014	1.135	0.0066	0.116	0.0008	0.962	0.0065
80	0.234	0.0009	1.069	0.0042	0.124	0.0008	0.897	0.0052
90	0.243	0.0010	1.027	0.0040	0.132	0.0006	0.845	0.0041
100	0.253	0.00066	0.987	0.00224	0.138	0.00048	0.804	0.00238
150	0.286	0.00046	0.875	0.00134	0.162	0.00034	0.685	0.00130
200	0.309	0.00032	0.808	0.00078	0.179	0.00022	0.620	0.00072
250	0.325	0.00028	0.769	0.00062	0.190	0.00020	0.584	0.00058
300	0.339	0.00020	0.738	0.00042	0.200	0.00016	0.555	0.00044
350	0.349	0.00014	0.717	0.00030	0.208	0.00014	0.533	0.00030
400	0.356	0.00014	0.702	0.00026	0.215	0.00012	0.518	0.00032
450	0.363	0.00014	0.689	0.00028	0.221	0.00010	0.502	0.00016
500	0.370	0.00008	0.675	0.00014	0.226	0.00008	0.494	0.00018
550	0.374	0.00008	0.668	0.00012	0.230	0.00006	0.485	0.00016
600	0.378	0.00008	0.662	0.00014	0.233	0.00008	0.477	0.00014
650	0.382	0.00008	0.655	0.00014	0.237	0.00006	0.470	0.00012
700	0.386	0.00006	0.648	0.00012	0.240	0.00004	0.464	0.00010
750	0.389	0.00008	0.642	0.00012	0.242	0.00004	0.459	0.00010
800	0.393	0.00006	0.636	0.00010	0.244	0.00006	0.454	0.00008
850	0.396	0.00006	0.631	0.00008	0.247	0.00004	0.450	0.00010
900	0.399	0.00004	0.627	0.00008	0.249	0.00006	0.445	0.00008
950	0.401	0.00004	0.623	0.00006	0.252	0.00000	0.441	0.00000
1000	0.403	....	0.620	....	0.252	....	0.441	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 27 (continued)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 5%
Order of Difference: k

Number of Items in the Original Series N	$k=2$				$k=3$			
	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	0.0060	0.00181	15.120	1.1391	....	....	....	....
20	0.0241	0.00165	3.729	0.1512	0.0151	0.00133	5.404	0.2526
30	0.0406	0.00142	2.217	0.0575	0.0284	0.00118	2.878	0.0850
40	0.0548	0.00115	1.642	0.0284	0.0402	0.00105	2.028	0.0416
50	0.0663	0.00099	1.358	0.0177	0.0507	0.00087	1.612	0.0239
60	0.0762	0.00080	1.181	0.0113	0.0594	0.00076	1.373	0.0155
70	0.0842	0.00080	1.068	0.0092	0.0670	0.00071	1.218	0.0116
80	0.0922	0.00067	0.976	0.0066	0.0741	0.00061	1.102	0.0085
90	0.0989	0.00061	0.910	0.0053	0.0802	0.00059	1.017	0.0068
100	0.105	0.00044	0.857	0.00296	0.0861	0.00044	0.949	0.00390
150	0.127	0.00032	0.709	0.00160	0.108	0.00030	0.754	0.00184
200	0.143	0.00024	0.629	0.00098	0.123	0.00024	0.662	0.00114
250	0.155	0.00020	0.580	0.00066	0.135	0.00020	0.605	0.00082
300	0.165	0.00012	0.547	0.00044	0.145	0.00014	0.564	0.00056
350	0.171	0.00014	0.525	0.00040	0.152	0.00012	0.536	0.00042
400	0.178	0.00012	0.505	0.00030	0.158	0.00010	0.515	0.00030
450	0.184	0.00008	0.490	0.00024	0.163	0.00010	0.500	0.00030
500	0.188	0.00010	0.478	0.00022	0.168	0.00010	0.485	0.00024
550	0.193	0.00006	0.467	0.00016	0.173	0.00008	0.473	0.00020
600	0.196	0.00008	0.459	0.00016	0.177	0.00006	0.463	0.00018
650	0.200	0.00006	0.451	0.00014	0.180	0.00006	0.454	0.00016
700	0.203	0.00004	0.444	0.00010	0.183	0.00006	0.446	0.00014
750	0.205	0.00004	0.439	0.00010	0.186	0.00004	0.439	0.00012
800	0.207	0.00004	0.434	0.00008	0.188	0.00006	0.433	0.00010
850	0.209	0.00004	0.430	0.00008	0.191	0.00004	0.428	0.00012
900	0.211	0.00006	0.426	0.00010	0.193	0.00004	0.422	0.00008
950	0.214	0.00004	0.421	0.00008	0.195	0.00004	0.418	0.00008
1000	0.216	....	0.417	....	0.197	....	0.414	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 27 (continued)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 5%
Order of Difference: k

Number of Items in the Original Series	$k=4$				$k=5$			
N	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	....	....	....	....	....	....	....	....
20	0.0096	0.00110	7.997	0.4257	0.0062	0.00091	12.171	0.7313
30	0.0206	0.00102	3.740	0.1233	0.0153	0.00089	4.858	0.1791
40	0.0308	0.00091	2.507	0.0574	0.0242	0.00082	3.067	0.0772
50	0.0399	0.00083	1.933	0.0335	0.0324	0.00076	2.295	0.0435
60	0.0482	0.00073	1.598	0.0208	0.0400	0.00065	1.860	0.0259
70	0.0555	0.00064	1.390	0.0145	0.0465	0.00064	1.601	0.0195
80	0.0619	0.00059	1.245	0.0107	0.0529	0.00056	1.406	0.0134
90	0.0678	0.00057	1.138	0.0088	0.0585	0.00048	1.272	0.0098
100	0.0735	0.00042	1.050	0.00464	0.0633	0.00041	1.174	0.00572
150	0.0943	0.00031	0.818	0.00228	0.0838	0.00031	0.888	0.00278
200	0.110	0.00022	0.704	0.00134	0.0993	0.00023	0.749	0.00156
250	0.121	0.00018	0.637	0.00090	0.111	0.00018	0.671	0.00104
300	0.130	0.00016	0.592	0.00066	0.120	0.00016	0.619	0.00072
350	0.138	0.00012	0.559	0.00050	0.128	0.00012	0.583	0.00056
400	0.144	0.00012	0.534	0.00040	0.134	0.00012	0.555	0.00044
450	0.150	0.00010	0.514	0.00032	0.140	0.00008	0.533	0.00032
500	0.155	0.00008	0.498	0.00026	0.144	0.00008	0.517	0.00030
550	0.159	0.00008	0.485	0.00024	0.148	0.00010	0.502	0.00030
600	0.163	0.00006	0.473	0.00020	0.153	0.00006	0.487	0.00020
650	0.166	0.00008	0.463	0.00016	0.156	0.00006	0.477	0.00018
700	0.170	0.00006	0.455	0.00016	0.159	0.00006	0.468	0.00020
750	0.173	0.00004	0.447	0.00012	0.162	0.00006	0.458	0.00016
800	0.175	0.00006	0.441	0.00014	0.165	0.00004	0.450	0.00012
850	0.178	0.00004	0.434	0.00010	0.167	0.00006	0.444	0.00012
900	0.180	0.00004	0.429	0.00010	0.170	0.00004	0.438	0.00012
950	0.182	0.00004	0.424	0.00010	0.172	0.00004	0.432	0.00008
1000	0.184	....	0.419	....	0.174	....	0.428	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 27 (continued)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 5%
Order of Difference: k

Number of Items in the Original Series	$k=6$				$k=7$			
N	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	....	....	....	....	....	....	....	....
20	....	....	....	....	....	....	....	....
30	0.0113	0.00079	6.410	0.2637	0.0085	0.00068	8.400	0.3743
40	0.0192	0.00072	3.773	0.1033	0.0153	0.00068	4.657	0.1441
50	0.0264	0.00072	2.740	0.0585	0.0221	0.00063	3.216	0.0711
60	0.0336	0.00063	2.155	0.0337	0.0284	0.00059	2.505	0.0434
70	0.0399	0.00055	1.818	0.0222	0.0343	0.00056	2.071	0.0288
80	0.0454	0.00053	1.596	0.0166	0.0399	0.00051	1.783	0.0202
90	0.0507	0.00053	1.430	0.0136	0.0450	0.00047	1.581	0.0150
100	0.0560	0.00039	1.294	0.00670	0.0497	0.00037	1.431	0.00784
150	0.0757	0.00030	0.959	0.00316	0.0684	0.00029	1.039	0.00360
200	0.0905	0.00023	0.801	0.00182	0.0828	0.00023	0.859	0.00210
250	0.102	0.00018	0.710	0.00108	0.0943	0.00017	0.754	0.00128
300	0.111	0.00016	0.656	0.00090	0.103	0.00018	0.690	0.00108
350	0.119	0.00012	0.611	0.00060	0.112	0.00010	0.636	0.00062
400	0.125	0.00012	0.581	0.00056	0.117	0.00012	0.605	0.00058
450	0.131	0.00010	0.553	0.00044	0.123	0.00010	0.576	0.00046
500	0.136	0.00008	0.531	0.00030	0.128	0.00012	0.553	0.00038
550	0.140	0.00008	0.516	0.00028	0.134	0.00008	0.534	0.00032
600	0.144	0.00008	0.502	0.00024	0.138	0.00006	0.518	0.00026
650	0.148	0.00006	0.490	0.00022	0.141	0.00006	0.505	0.00024
700	0.151	0.00006	0.479	0.00018	0.144	0.00006	0.493	0.00020
750	0.154	0.00006	0.470	0.00016	0.147	0.00006	0.483	0.00018
800	0.157	0.00006	0.462	0.00016	0.150	0.00006	0.474	0.00018
850	0.160	0.00004	0.454	0.00014	0.153	0.00004	0.465	0.00014
900	0.162	0.00006	0.447	0.00012	0.155	0.00004	0.458	0.00012
950	0.165	0.00004	0.441	0.00008	0.157	0.00006	0.452	0.00012
1000	0.167	....	0.435	....	0.160	....	0.446	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 27 (concluded)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 5%
Order of Difference: k

Number of Items in the Original Series	$k=8$				$k=9$			
N	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	....	....	....	....	....	....	....	....
20	....	....	....	....	....	....	....	....
30	0.0064	0.00059	11.033	0.5331	0.0047	0.000 52	14.658	0.7664
40	0.0123	0.00060	5.702	0.1880	0.0099	0.000 55	6.994	0.2490
50	0.0183	0.00060	3.822	0.0933	0.0154	0.000 54	4.504	0.1167
60	0.0243	0.00053	2.889	0.0524	0.0208	0.000 51	3.337	0.0659
70	0.0296	0.00052	2.365	0.0350	0.0259	0.000 48	2.678	0.0419
80	0.0348	0.00048	2.015	0.0245	0.0307	0.000 46	2.259	0.0295
90	0.0396	0.00046	1.770	0.0185	0.0353	0.000 45	1.964	0.0222
100	0.0442	0.00037	1.585	0.00936	0.0398	0.000 356	1.742	0.01078
150	0.0627	0.00028	1.117	0.00404	0.0576	0.000 282	1.203	0.00474
200	0.0766	0.00023	0.915	0.00240	0.0717	0.000 216	0.966	0.00254
250	0.0881	0.00019	0.795	0.00150	0.0825	0.000 192	0.839	0.00174
300	0.0974	0.00015	0.720	0.00112	0.0921	0.000 154	0.752	0.00116
350	0.105	0.00014	0.664	0.00072	0.0998	0.000 124	0.694	0.00080
400	0.112	0.00010	0.628	0.00062	0.106	0.000 10	0.654	0.00064
450	0.117	0.00010	0.597	0.00050	0.111	0.000 12	0.622	0.00060
500	0.122	0.00010	0.572	0.00040	0.117	0.000 10	0.592	0.00044
550	0.127	0.00008	0.552	0.00034	0.122	0.000 08	0.570	0.00038
600	0.131	0.00008	0.535	0.00030	0.126	0.000 06	0.551	0.00032
650	0.135	0.00006	0.520	0.00026	0.129	0.000 08	0.535	0.00032
700	0.138	0.00006	0.507	0.00020	0.133	0.000 06	0.519	0.00020
750	0.141	0.00006	0.497	0.00020	0.136	0.000 06	0.509	0.00020
800	0.144	0.00006	0.487	0.00020	0.139	0.000 06	0.499	0.00020
850	0.147	0.00004	0.477	0.00016	0.142	0.000 04	0.489	0.00016
900	0.149	0.00006	0.469	0.00014	0.144	0.000 04	0.481	0.00016
950	0.152	0.00004	0.462	0.00012	0.146	0.000 06	0.473	0.00014
1000	0.154	....	0.456	....	0.149	....	0.466	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 28

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 1%
Order of Difference: k

Number of Items in the Original Series	$k=0$				$k=1$			
N	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	0.025	0.0039	9.844	0.5960	0.0057	0.00163	19.519	1.4459
20	0.064	0.0031	3.884	0.1254	0.0220	0.00157	5.060	0.2111
30	0.095	0.0025	2.630	0.0541	0.0377	0.00137	2.949	0.0786
40	0.120	0.0019	2.089	0.0291	0.0514	0.00120	2.163	0.0410
50	0.139	0.0018	1.798	0.0203	0.0634	0.00102	1.753	0.0244
60	0.157	0.0015	1.595	0.0137	0.0736	0.00086	1.509	0.0157
70	0.172	0.0012	1.458	0.0099	0.0822	0.00086	1.352	0.0129
80	0.184	0.0009	1.359	0.0066	0.0908	0.00066	1.223	0.0083
90	0.193	0.0010	1.293	0.0063	0.0974	0.00066	1.140	0.0077
100	0.203	0.00076	1.230	0.00384	0.104	0.00050	1.063	0.00402
150	0.241	0.00050	1.038	0.00198	0.129	0.00036	0.862	0.00210
200	0.266	0.00034	0.939	0.00110	0.147	0.00028	0.757	0.00130
250	0.283	0.00028	0.884	0.00086	0.161	0.00020	0.692	0.00082
300	0.297	0.00024	0.841	0.00066	0.171	0.00018	0.651	0.00066
350	0.309	0.00020	0.808	0.00048	0.180	0.00014	0.618	0.00048
400	0.319	0.00020	0.784	0.00046	0.187	0.00012	0.594	0.00040
450	0.329	0.00012	0.761	0.00030	0.193	0.00012	0.574	0.00030
500	0.335	0.00014	0.746	0.00030	0.199	0.00010	0.559	0.00028
550	0.342	0.00014	0.731	0.00028	0.204	0.00008	0.545	0.00022
600	0.349	0.00006	0.717	0.00014	0.208	0.00008	0.534	0.00020
650	0.352	0.00008	0.710	0.00016	0.212	0.00008	0.524	0.00016
700	0.356	0.00006	0.702	0.00012	0.216	0.00006	0.516	0.00016
750	0.359	0.00008	0.696	0.00014	0.219	0.00006	0.508	0.00014
800	0.363	0.00008	0.689	0.00014	0.222	0.00004	0.501	0.00012
850	0.367	0.00006	0.682	0.00014	0.224	0.00006	0.495	0.00010
900	0.370	0.00008	0.675	0.00014	0.227	0.00004	0.490	0.00010
950	0.374	0.00008	0.668	0.00012	0.229	0.00004	0.485	0.00010
1000	0.378	....	0.662	....	0.231	....	0.480	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 28 (continued)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 1%
Order of Difference: k

Number of Items in the Original Series	$k=2$				$k=3$			
N	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	0.0017	0.00093	51.730	4.3514	....	....	....	....
20	0.0110	0.00108	8.216	0.4095	0.0060	0.00078	13.561	0.7648
30	0.0218	0.00101	4.121	0.1303	0.0138	0.00078	5.913	0.2142
40	0.0319	0.00091	2.818	0.0623	0.0216	0.00076	3.771	0.0978
50	0.0410	0.00086	2.195	0.0380	0.0292	0.00072	2.793	0.0551
60	0.0496	0.00074	1.815	0.0237	0.0364	0.00063	2.242	0.0332
70	0.0570	0.00067	1.578	0.0165	0.0427	0.00060	1.910	0.0233
80	0.0637	0.00060	1.413	0.0121	0.0487	0.00051	1.677	0.0159
90	0.0697	0.00058	1.292	0.0100	0.0538	0.00050	1.518	0.0131
100	0.0755	0.00045	1.192	0.00546	0.0588	0.00041	1.387	0.00718
150	0.0979	0.00032	0.919	0.00256	0.0794	0.00031	1.028	0.00340
200	0.114	0.00024	0.791	0.00150	0.0951	0.00024	0.858	0.00194
250	0.126	0.00020	0.716	0.00110	0.107	0.00018	0.761	0.00118
300	0.136	0.00018	0.661	0.00076	0.116	0.00018	0.702	0.00092
350	0.145	0.00014	0.623	0.00062	0.125	0.00012	0.656	0.00070
400	0.152	0.00012	0.592	0.00046	0.131	0.00014	0.621	0.00054
450	0.158	0.00010	0.569	0.00034	0.138	0.00010	0.594	0.00044
500	0.163	0.00010	0.552	0.00032	0.143	0.00010	0.572	0.00038
550	0.168	0.00006	0.536	0.00022	0.148	0.00008	0.553	0.00034
600	0.171	0.00008	0.525	0.00020	0.152	0.00006	0.536	0.00020
650	0.175	0.00006	0.515	0.00020	0.155	0.00006	0.526	0.00022
700	0.178	0.00008	0.505	0.00022	0.158	0.00008	0.515	0.00024
750	0.182	0.00006	0.494	0.00016	0.162	0.00006	0.503	0.00018
800	0.185	0.00006	0.486	0.00014	0.165	0.00006	0.494	0.00016
850	0.188	0.00004	0.479	0.00012	0.168	0.00004	0.486	0.00014
900	0.190	0.00006	0.473	0.00012	0.170	0.00006	0.479	0.00014
950	0.193	0.00004	0.467	0.00010	0.173	0.00004	0.472	0.00012
1000	0.195	....	0.462	....	0.175	....	0.466	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 28 (continued)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 1%
Order of Difference: k

Number of Items in the Original Series N	$k=4$ Lower limit	D.D.*	Upper limit	D.D.*	$k=5$ Lower limit	D.D.*	Upper limit	D.D.*
10	....	....	....	....	....	....	....	....
20	0.0034	0.00057	22.852	1.4361	0.0019	0.00043	38.888	2.6818
30	0.0091	0.00063	8.491	0.3493	0.0062	0.00051	12.070	0.5446
40	0.0154	0.00063	4.998	0.1440	0.0113	0.00053	6.624	0.2139
50	0.0217	0.00061	3.558	0.0787	0.0166	0.00052	4.485	0.1061
60	0.0278	0.00055	2.771	0.0457	0.0218	0.00050	3.424	0.0597
70	0.0333	0.00054	2.314	0.0322	0.0268	0.00046	2.827	0.0462
80	0.0387	0.00050	1.992	0.0225	0.0314	0.00044	2.365	0.0288
90	0.0437	0.00046	1.767	0.0169	0.0358	0.00042	2.077	0.0217
100	0.0483	0.00038	1.598	0.00898	0.0400	0.00036	1.860	0.0115
150	0.0671	0.00030	1.149	0.00416	0.0579	0.00028	1.285	0.00508
200	0.0820	0.00023	0.941	0.00230	0.0721	0.00022	1.031	0.00270
250	0.0934	0.00017	0.826	0.00142	0.0830	0.00019	0.896	0.00186
300	0.102	0.00018	0.755	0.00116	0.0926	0.00015	0.803	0.00124
350	0.111	0.00014	0.697	0.00082	0.100	0.00012	0.741	0.00086
400	0.118	0.00012	0.656	0.00064	0.106	0.00014	0.698	0.00080
450	0.124	0.00010	0.624	0.00048	0.113	0.00010	0.658	0.00052
500	0.129	0.00010	0.600	0.00048	0.118	0.00010	0.632	0.00050
550	0.134	0.00008	0.576	0.00034	0.123	0.00010	0.607	0.00048
600	0.138	0.00008	0.559	0.00032	0.128	0.00006	0.583	0.00034
650	0.142	0.00006	0.543	0.00022	0.131	0.00008	0.566	0.00034
700	0.145	0.00006	0.532	0.00020	0.135	0.00006	0.549	0.00022
750	0.148	0.00006	0.522	0.00026	0.138	0.00006	0.538	0.00020
800	0.151	0.00006	0.509	0.00018	0.141	0.00006	0.528	0.00022
850	0.154	0.00006	0.500	0.00016	0.144	0.00006	0.517	0.00020
900	0.157	0.00004	0.492	0.00016	0.147	0.00006	0.507	0.00020
950	0.159	0.00006	0.484	0.00014	0.150	0.00002	0.497	0.00010
1000	0.162	....	0.477	....	0.151	....	0.492	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 28 (continued)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 1%
Order of Difference: k

Number of Items in the Original Series N	$k=6$				$k=7$			
	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	....	....	....	....	....	....	....	....
20	....	....	....	....	....	....	....	....
30	0.0042	0.000 42	17.250	0.8598	0.0029	0.000 33	24.736	1.3283
40	0.0084	0.000 44	8.652	0.3024	0.0062	0.000 38	11.453	0.4366
50	0.0128	0.000 46	5.628	0.1458	0.0100	0.000 40	7.087	0.1992
60	0.0174	0.000 45	4.170	0.0857	0.0140	0.000 39	5.095	0.1127
70	0.0219	0.000 40	3.313	0.0518	0.0179	0.000 40	3.968	0.0719
80	0.0259	0.000 42	2.795	0.0389	0.0219	0.000 35	3.249	0.0453
90	0.0301	0.000 39	2.406	0.0272	0.0254	0.000 36	2.796	0.0341
100	0.0340	0.000 334	2.134	0.01408	0.0290	0.000 320	2.455	0.01748
150	0.0507	0.000 262	1.430	0.00588	0.0450	0.000 254	1.581	0.00698
200	0.0638	0.000 220	1.136	0.00336	0.0577	0.000 202	1.232	0.00366
250	0.0748	0.000 176	0.968	0.00202	0.0678	0.000 188	1.049	0.00256
300	0.0836	0.000 156	0.867	0.00148	0.0772	0.000 144	0.921	0.00158
350	0.0914	0.000 132	0.793	0.00108	0.0844	0.000 142	0.842	0.00130
400	0.0980	0.000 120	0.739	0.00086	0.0915	0.000 112	0.777	0.00090
450	0.104	0.000 100	0.696	0.00068	0.0971	0.000 098	0.732	0.00072
500	0.109	0.000 10	0.662	0.00052	0.102	0.000 10	0.696	0.00068
550	0.114	0.000 10	0.636	0.00050	0.107	0.000 10	0.662	0.00052
600	0.119	0.000 06	0.611	0.00036	0.112	0.000 06	0.636	0.00038
650	0.122	0.000 08	0.593	0.00034	0.115	0.000 08	0.617	0.00038
700	0.126	0.000 08	0.576	0.00034	0.119	0.000 06	0.598	0.00032
750	0.130	0.000 04	0.559	0.00022	0.122	0.000 06	0.582	0.00030
800	0.132	0.000 06	0.548	0.00022	0.125	0.000 06	0.567	0.00024
850	0.135	0.000 06	0.537	0.00022	0.128	0.000 06	0.555	0.00024
900	0.138	0.000 06	0.526	0.00020	0.131	0.000 06	0.543	0.00020
950	0.141	0.000 04	0.516	0.00020	0.134	0.000 04	0.533	0.00020
1000	0.143	....	0.506	....	0.136	....	0.523	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 28 (concluded)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 1%
Order of Difference: k

Number of Items in the Original Series N	$k=8$ Lower limit	D.D.*	Upper limit	D.D.*	$k=9$ Lower limit	D.D.*	Upper limit	D.D.*
10	....	....	....	....	....	....	....	....
20	....	....	....	....	....	....	....	....
30	0.0020	0.000 27	35.547	2.0654	0.0013	0.000 22	51.676	3.2087
40	0.0047	0.000 32	14.893	0.6039	0.0035	0.000 28	19.589	0.8621
50	0.0079	0.000 35	8.854	0.2738	0.0063	0.000 30	10.968	0.3542
60	0.0114	0.000 35	6.116	0.1401	0.0093	0.000 32	7.426	0.1869
70	0.0149	0.000 35	4.715	0.0931	0.0125	0.000 32	5.557	0.1142
80	0.0184	0.000 34	3.784	0.0559	0.0157	0.000 31	4.415	0.0727
90	0.0218	0.000 33	3.225	0.0449	0.0188	0.000 30	3.688	0.0514
100	0.0251	0.000 298	2.776	0.02048	0.0218	0.000 276	3.174	0.02460
150	0.0400	0.000 248	1.752	0.00828	0.0356	0.000 240	1.944	0.00978
200	0.0524	0.000 194	1.338	0.00420	0.0476	0.000 200	1.455	0.00504
250	0.0621	0.000 172	1.128	0.00274	0.0576	0.000 158	1.203	0.00292
300	0.0707	0.000 148	0.991	0.00188	0.0655	0.000 154	1.057	0.00222
350	0.0781	0.000 130	0.897	0.00138	0.0732	0.000 122	0.946	0.00144
400	0.0846	0.000 124	0.828	0.00112	0.0793	0.000 114	0.874	0.00118
450	0.0908	0.000 112	0.772	0.00090	0.0850	0.000 106	0.815	0.00096
500	0.0964	0.000 092	0.727	0.00072	0.0903	0.000 092	0.767	0.00074
550	0.101	0.000 08	0.691	0.00054	0.0949	0.000 098	0.730	0.00072
600	0.105	0.000 08	0.664	0.00042	0.0998	0.000 084	0.694	0.00054
650	0.109	0.000 08	0.643	0.00042	0.104	0.000 06	0.667	0.00040
700	0.113	0.000 06	0.622	0.00034	0.107	0.000 06	0.647	0.00038
750	0.116	0.000 06	0.605	0.00034	0.110	0.000 08	0.628	0.00036
800	0.119	0.000 06	0.588	0.00028	0.114	0.000 04	0.610	0.00024
850	0.122	0.000 06	0.574	0.00028	0.116	0.000 06	0.598	0.00036
900	0.125	0.000 06	0.560	0.00022	0.119	0.000 06	0.580	0.00024
950	0.128	0.000 04	0.549	0.00022	0.122	0.000 04	0.568	0.00022
1000	0.130	....	0.538	....	0.124	....	0.557	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 29

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 0.1%
Order of Difference: k

Number of Items in the Original Series N	$k = 0$				$k = 1$			
	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	0.0112	0.00252	22.350	1.5482	0.0018	0.00086	60.423	4.9711
20	0.0364	0.00236	6.868	0.2702	0.0104	0.00103	10.712	0.5339
30	0.0600	0.00210	4.166	0.1080	0.0207	0.00098	5.373	0.1735
40	0.0810	0.00170	3.086	0.0534	0.0305	0.00095	3.638	0.0861
50	0.0980	0.0016	2.552	0.0356	0.0400	0.00084	2.777	0.0481
60	0.114	0.0013	2.196	0.0228	0.0484	0.00078	2.296	0.0319
70	0.127	0.0012	1.968	0.0170	0.0562	0.00065	1.977	0.0206
80	0.139	0.0010	1.798	0.0121	0.0627	0.00066	1.771	0.0169
90	0.149	0.0011	1.677	0.0114	0.0693	0.00058	1.602	0.0123
100	0.160	0.00074	1.563	0.00592	0.0751	0.00049	1.479	0.00722
150	0.197	0.00050	1.267	0.00286	0.0994	0.00037	1.118	0.00350
200	0.222	0.00042	1.124	0.00194	0.118	0.00028	0.943	0.00196
250	0.243	0.00030	1.027	0.00120	0.132	0.00020	0.845	0.00130
300	0.258	0.00028	0.967	0.00094	0.142	0.00018	0.780	0.00092
350	0.272	0.00022	0.920	0.00072	0.151	0.00016	0.734	0.00070
400	0.283	0.00016	0.884	0.00052	0.159	0.00014	0.699	0.00056
450	0.291	0.00018	0.858	0.00050	0.166	0.00012	0.671	0.00052
500	0.300	0.00012	0.833	0.00034	0.172	0.00012	0.645	0.00038
550	0.306	0.00016	0.816	0.00040	0.178	0.00010	0.626	0.00038
600	0.314	0.00012	0.796	0.00028	0.183	0.00008	0.607	0.00024
650	0.320	0.00010	0.782	0.00024	0.187	0.00006	0.595	0.00022
700	0.325	0.00010	0.770	0.00024	0.190	0.00008	0.584	0.00024
750	0.330	0.00008	0.758	0.00020	0.194	0.00008	0.572	0.00022
800	0.334	0.00010	0.748	0.00020	0.198	0.00008	0.561	0.00022
850	0.339	0.00006	0.738	0.00014	0.202	0.00004	0.550	0.00012
900	0.342	0.00006	0.731	0.00014	0.204	0.00004	0.544	0.00010
950	0.345	0.00008	0.724	0.00014	0.206	0.00004	0.539	0.00012
1000	0.349	....	0.717	....	0.208	....	0.533	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 29 (continued)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 0.1%
Order of Difference: k

Number of Items in the Original Series	$k=2$				$k=3$			
N	Lower limit	D.D.*	Upper limit	D.D.*	Lower limit	D.D.*	Upper limit	D.D.*
10	0.0004	0.00040	216.122	19.5712	....	....	....	....
20	0.0044	0.00061	20.410	1.1859	0.0020	0.00040	39.934	2.6236
30	0.0105	0.00067	8.551	0.3312	0.0060	0.00046	13.698	0.6029
40	0.0172	0.00065	5.239	0.1435	0.0106	0.00050	7.669	0.2424
50	0.0237	0.00061	3.804	0.0802	0.0156	0.00048	5.245	0.1241
60	0.0298	0.00059	3.002	0.0478	0.0204	0.00048	4.004	0.0759
70	0.0357	0.00057	2.524	0.0351	0.0252	0.00046	3.245	0.0507
80	0.0414	0.00048	2.173	0.0227	0.0298	0.00042	2.738	0.0334
90	0.0462	0.00049	1.946	0.0185	0.0340	0.00043	2.404	0.0272
100	0.0511	0.00040	1.761	0.0099	0.0383	0.00035	2.132	0.01348
150	0.0711	0.00031	1.266	0.00458	0.0560	0.00028	1.458	0.00576
200	0.0868	0.00024	1.037	0.00254	0.0698	0.00022	1.170	0.00326
250	0.0989	0.00020	0.910	0.00172	0.0810	0.00019	1.007	0.00210
300	0.109	0.00018	0.824	0.00128	0.0905	0.00017	0.902	0.00154
350	0.118	0.00016	0.760	0.00088	0.099	0.00014	0.825	0.00112
400	0.126	0.00012	0.716	0.00070	0.106	0.00012	0.769	0.00076
450	0.132	0.00012	0.681	0.00054	0.112	0.00010	0.731	0.00070
500	0.138	0.00010	0.654	0.00050	0.117	0.00012	0.696	0.00068
550	0.143	0.00008	0.629	0.00038	0.123	0.00008	0.662	0.00040
600	0.147	0.00010	0.610	0.00036	0.127	0.00008	0.642	0.00038
650	0.152	0.00006	0.592	0.00024	0.131	0.00008	0.623	0.00036
700	0.155	0.00006	0.580	0.00022	0.135	0.00008	0.605	0.00036
750	0.158	0.00006	0.569	0.00026	0.139	0.00006	0.587	0.00024
800	0.161	0.00008	0.556	0.00020	0.142	0.00006	0.575	0.00022
850	0.165	0.00006	0.546	0.00020	0.145	0.00006	0.564	0.00022
900	0.168	0.00004	0.536	0.00016	0.148	0.00006	0.553	0.00022
950	0.170	0.00006	0.528	0.00014	0.151	0.00002	0.542	0.00012
1000	0.173	....	0.521	....	0.152	....	0.536	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 29 (continued)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 0.1%
Order of Difference: k

Number of Items in the Original Series	$k=4$				$k=5$			
N	Lower Limit	D.D.*	Upper Limit	D.D.*	Lower Limit	D.D.*	Upper Limit	D.D.*
10	....	....	....	....	....	....	....	....
20	0.00099	0.000 256	77.778	5.6040	0.00048	0.000 168	154.717	12.0225
30	0.00355	0.000 339	21.738	1.0614	0.00216	0.000 250	34.492	1.8522
40	0.00694	0.000 386	11.124	0.3960	0.00466	0.000 294	15.970	0.6186
50	0.0108	0.000 390	7.164	0.1910	0.00760	0.000 32	9.784	0.2890
60	0.0147	0.000 380	5.254	0.1079	0.0108	0.000 32	6.894	0.1578
70	0.0185	0.000 380	4.175	0.0723	0.0140	0.000 33	5.316	0.1007
80	0.0223	0.000 370	3.452	0.0481	0.0173	0.000 32	4.309	0.0674
90	0.0260	0.000 360	2.971	0.0362	0.0205	0.000 30	3.635	0.0475
100	0.0296	0.000 318	2.609	0.01824	0.0235	0.000 284	3.160	0.02370
150	0.0455	0.000 258	1.697	0.00750	0.0377	0.000 242	1.975	0.00964
200	0.0584	0.000 216	1.322	0.00414	0.0498	0.000 196	1.493	0.00492
250	0.0692	0.000 176	1.115	0.00252	0.0596	0.000 180	1.247	0.00326
300	0.0780	0.000 148	0.989	0.00170	0.0686	0.000 144	1.0841	0.00206
350	0.0854	0.000 142	0.904	0.00142	0.0758	0.000 126	0.981	0.00152
400	0.0925	0.000 124	0.833	0.00102	0.0821	0.000 120	0.905	0.00122
450	0.0987	0.000 106	0.782	0.00082	0.0881	0.000 108	0.844	0.00102
500	0.104	0.000 10	0.741	0.00068	0.0935	0.000 102	0.793	0.00078
550	0.109	0.000 10	0.707	0.00054	0.0986	0.000 088	0.754	0.00064
600	0.114	0.000 08	0.680	0.00048	0.103	0.000 08	0.722	0.00054
650	0.118	0.000 06	0.656	0.00038	0.107	0.000 08	0.695	0.00048
700	0.121	0.000 08	0.637	0.00038	0.111	0.000 06	0.671	0.00040
750	0.125	0.000 08	0.618	0.00036	0.114	0.000 08	0.651	0.00036
800	0.129	0.000 04	0.600	0.00024	0.118	0.000 06	0.633	0.00032
850	0.131	0.000 06	0.588	0.00024	0.121	0.000 04	0.617	0.00028
900	0.134	0.000 06	0.576	0.00022	0.123	0.000 06	0.603	0.00028
950	0.137	0.000 04	0.565	0.00022	0.126	0.000 06	0.589	0.00024
1000	0.139	....	0.554	....	0.129	....	0.577	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 29 (continued)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 0.1%
Order of Difference: k

Number of Items in the Original Series	$k = 6$				$k = 7$			
N	Lower Limit	D.D.*	Upper Limit	D.D.*	Lower Limit	D.D.*	Upper Limit	D.D.*
10	....	....	....	....	....	....	....	....
20	....	....	....	....	....	....	....	....
30	0.00133	0.000 185	54.479	3.1655	0.00082	0.000 137	86.667	5.4264
40	0.00318	0.000 232	22.824	0.9656	0.00219	0.000 185	32.403	1.4797
50	0.00550	0.000 263	13.168	0.4252	0.00404	0.000 217	17.606	0.6153
60	0.00813	0.000 277	8.916	0.2245	0.00621	0.000 234	11.453	0.3136
70	0.0109	0.000 280	6.671	0.1370	0.00855	0.000 235	8.317	0.1775
80	0.0137	0.000 270	5.301	0.0874	0.0109	0.000 240	6.542	0.1186
90	0.0164	0.000 280	4.427	0.0654	0.0133	0.000 240	5.356	0.0837
100	0.0192	0.000 256	3.773	0.03014	0.0157	0.000 232	4.519	0.03824
150	0.0320	0.000 224	2.266	0.01176	0.0273	0.000 206	2.607	0.01428
200	0.0432	0.000 180	1.678	0.00580	0.0376	0.000 184	1.893	0.00748
250	0.0522	0.000 170	1.388	0.00386	0.0468	0.000 152	1.519	0.00422
300	0.0607	0.000 146	1.195	0.00250	0.0544	0.000 138	1.308	0.00296
350	0.0680	0.000 128	1.070	0.00184	0.0613	0.000 130	1.160	0.00222
400	0.0744	0.000 116	0.978	0.00148	0.0678	0.000 112	1.049	0.00160
450	0.0802	0.000 104	0.904	0.00112	0.0734	0.000 102	0.969	0.00126
500	0.0854	0.000 078	0.848	0.00072	0.0785	0.000 094	0.906	0.00102
550	0.0893	0.000 104	0.812	0.00090	0.0832	0.000 084	0.855	0.00084
600	0.0945	0.000 082	0.767	0.00062	0.0874	0.000 080	0.813	0.00070
650	0.0986	0.000 07	0.736	0.00054	0.0914	0.000 074	0.778	0.00060
700	0.102	0.000 06	0.709	0.00046	0.0951	0.000 068	0.748	0.00052
750	0.105	0.000 08	0.686	0.00040	0.0985	0.000 070	0.722	0.00046
800	0.109	0.000 06	0.666	0.00036	0.102	0.000 06	0.699	0.00040
850	0.112	0.000 06	0.648	0.00032	0.105	0.000 06	0.679	0.00036
900	0.115	0.000 04	0.632	0.00030	0.108	0.000 04	0.661	0.00036
950	0.117	0.000 06	0.617	0.00024	0.110	0.000 06	0.643	0.00026
1000	0.120	....	0.605	....	0.113	....	0.630	....

* Divided difference, lower limit, positive; upper limit, negative.

TABLE 29 (concluded)

LIMITS FOR THE RATIOS OF SUMS OF SQUARES OF SELECTED COMPARISONS OF DIFFERENCES
Level of Significance: 0.1%
Order of Difference: k

Number of Items in the Original Series N	$k=8$				$k=9$			
	Lower Limit	D.D.*	Upper Limit	D.D.*	Lower Limit	D.D.*	Upper Limit	D.D.*
10	....	....	....	....	....	....	....	....
20	....	....	....	....	....	....	....	....
30	0.00051	0.000 103	138.334	9.2691	....	....	....	....
40	0.00154	0.000 146	45.643	2.2287	0.00107	0.000 116	64.737	3.3705
50	0.00300	0.000 175	23.356	0.8612	0.00223	0.000 149	31.032	1.2398
60	0.00475	0.000 199	14.744	0.4354	0.00372	0.000 166	18.634	0.5763
70	0.00674	0.000 209	10.390	0.2458	0.00538	0 000 181	12.871	0.3183
80	0.00883	0.000 207	7.932	0.1503	0.00719	0.000 186	9.688	0.2036
90	0.0109	0.000 210	6.429	0.1059	0.00905	0.000 185	7.652	0.1324
100	0.0130	0.000 210	5.370	0.04786	0.0109	0.000 192	6.328	0.05916
150	0.0235	0.000 192	2.977	0.01716	0.0205	0.000 180	3.370	0.02038
200	0.0331	0.000 170	2.119	0.00872	0.0295	0.000 158	2.351	0.01002
250	0.0416	0.000 154	1.683	0.00526	0.0374	0.000 148	1.850	0.00610
300	0.0493	0.000 138	1.420	0.00346	0.0448	0.000 124	1.545	0.00376
350	0.0562	0.000 118	1.247	0.00238	0.0510	0.000 124	1.357	0.00284
400	0.0621	0.000 108	1.128	0.00172	0.0572	0.000 106	1.215	0.00208
450	0.0675	0.000 100	1.042	0.00160	0.0625	0.000 098	1.111	0.00171
500	0.0725	0.000 092	0.962	0.00112	0.0674	0.000 090	1.025	0.00118
550	0.0771	0.000 086	0.906	0.00088	0.0719	0.000 082	0.966	0.00114
600	0.0814	0.000 078	0.862	0.00086	0.0760	0.000 078	0.909	0.00088
650	0.0853	0.000 072	0.819	0.00064	0.0799	0.000 072	0.865	0.00068
700	0.0889	0.000 068	0.787	0.00062	0.0835	0.000 068	0.831	0.00064
750	0.0923	0.000 064	0.756	0.00044	0.0869	0.000 062	0.799	0.00064
800	0.0955	0.000 060	0.734	0.00044	0.0900	0.000 060	0.767	0.00044
850	0.0985	0.000 050	0.712	0.00040	0.0930	0.000 056	0.745	0.00044
900	0.101	0.000 06	0.692	0.00036	0.0958	0.000 052	0.723	0.00044
950	0.104	0.000 04	0.674	0.00032	0.0984	0.000 052	0.701	0.00028
1000	0.106	....	0.658	....	0.101	....	0.687	....

* Divided difference, lower limit, positive; upper limit, negative.

of the sums of squares of selected differences (i.e., $\bar{S}_2^{(k)}/\bar{S}_2^{(k+1)}$) instead of the variances themselves in order to avoid unnecessary computations. We denote throughout by $\bar{S}_2^{(k)}$ the sum of squares of selected items of the lower difference, the kth, and by $\bar{S}_2^{(k+1)}$ the sum of squares of selected items of the higher difference, the $(k+1)$th; k is the order of selection.

We give in Tables 27, 28, 29 the ratios for the following levels of significance: 5% (Table 27), 1% (Table 28), 0.1% (Table 29). In calculating these tables we have made use of a property of z indicated by Fisher. If the estimates of the variances are based on the same number of degrees of freedom, the distribution of the z is nearly normal with variance $1/n$, where n is the number of degrees of freedom.[9] Tables 27, 28, and 29 give for selected values of N (number of items in the original series) the limits within which the ratios of the sums of squares of selected items of the kth and $(k+1)$th differences, $\bar{S}_2^{(k)}/\bar{S}_2^{(k+1)}$, must fall in order that we may be reasonably sure from the point of view of the given level of significance that they may still be considered equal.

To give an example we have tabulated in Table 30 the sums of squares of selected items of the differences for the annual American wheat-flour prices, 1890 to 1937. For instance, for the selection 0-A (Table 26), this is done by taking only items 1, 4, 7, 10, etc. of our original series and items 2, 5, 8, 11, etc. of the series of the first differences. The sum of the squares of the selected original items is $\bar{S}_2^{(k)} =$ 396.04 and of the selected items of first differences is $\bar{S}_2^{(k+1)} = 12.713$ (Table 30). For selection 0-B (Table 26), that is to say, including only items 2, 5, 8, 11, etc. of the original series and items 3, 6, 9, 12, etc. of the series of first differences, we get for the sum of squares of the selected original items $\bar{S}_2^{(k)} = 350.36$ and for the sum of squares of the selected first differences $\bar{S}_2^{(k+1)} = 21.471$ (Table 30). It should be mentioned that the originals are corrected for the arithmetic mean, but none of the differences are corrected since the a priori or true mean of the differences is equal to zero.

The same method of procedure has been followed for the selection 0-C and also for the comparison of higher differences. We have for instance, in Table 30, for the selection 1-A (Table 26) (that is to say, by taking only the items number 1, 6, 11, 16, etc. of the series of first differences and the items number 3, 8, 13, 18, etc. of the series of sec-

[9] R. A. Fisher, *op. cit.*, p. 233, Chapter VII, Section 41; G. Tintner, *op. cit.*, p. 141.

TABLE 30

SUMS OF SQUARES OF SELECTED DIFFERENCES, ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Selection	Lower Difference $\bar{S}_2^{(k)}$	Higher Difference $\bar{S}_2^{(k+1)}$	Selection	Lower Difference $\bar{S}_2^{(k)}$	Higher Difference $\bar{S}_2^{(k+1)}$
0-A	396.04	12.713	5-A	234.859	265.980
0-B	350.36	21.471	5-B	189.8810	81.887
0-C	363.09	31.246	5-C	606.887	79.888
			5-D	271.227	788.957
1-A	22.061	15.220	5-E	58.616	2,561.369
1-B	21.254	6.189	5-F	9.880	3,629.261
1-C	5.732	34.863	5-G	109.717	1,906.076
1-D	7.846	28.542	5-H	77.967	303.582
1-E	7.016	34.066	5-I	6.275	1,450.478
			5-J	75.295	1,703.660
2-A	6.227	23.908	5-K	394.668	547.536
2-B	31.957	80.091	5-L	941.390	61.766
2-C	12.008	83.613	5-M	880.976	109.364
2-D	1.760	26.738			
2-E	16.032	35.059	6-A	99.019	6,205.863
2-F	30.260	77.316	6-B	52.441	12,452.669
2-G	18.434	20.305	6-C	35.975	10,923.900
			6-D	109.239	2,274.767
3-A	20.951	34.015	6-E	150.308	1,306.471
3-B	16.860	75.282	6-F	103.997	6,297.367
3-C	72.759	209.491	6-G	161.223	4,511.074
3-D	17.585	322.104	6-H	878.775	1,082.739
3-E	2.818	162.394	6-I	2,538.750	23.774
3-F	14.993	0.446	6-J	3,621.064	276.218
3-G	31.902	157.202	6-K	1,802.933	282.364
3-H	85.131	148.338	6-L	1.606	442.733
3-I	72.561	30.515	6-M	1,370.876	411.256
			6-N	1,797.165	1.218
4-A	29.647	149.507	6-O	613.630	1,287.797
4-B	45.289	613.622			
4-C	194.370	324.917	7-A	208.336	79.230
4-D	318.796	106.094	7-B	506.364	12,359.940
4-E	162.872	72.725	7-C	466.594	19,319.770
4-F	0.857	41.531	7-D	42.226	9,922.430
4-G	156.929	94.893	7-E	1,300.762	3,216.760
4-H	142.592	401.672	7-F	6,432.291	1,364.830
4-I	29.064	1,002.239	7-G	12,469.065	1,122.780
4-J	19.932	919.619	7-H	10,526.338	119.440
4-K	10.300	166.569	7-I	1,841.058	215.230

TABLE 30 (concluded)

SUMS OF SQUARES OF SELECTED DIFFERENCES, ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Selection	Lower Difference $\bar{S}_2^{(k)}$	Higher Difference $\bar{S}_2^{(k+1)}$	Selection	Lower Difference $\bar{S}_2^{(k)}$	Higher Difference $\bar{S}_2^{(k+1)}$
7-J	1,250.435	1,031.470	8-Q	126.740	10,612.500
7-K	5,752.310	1,641.310	8-R	999.190	62,376.060
7-L	4,076.319	518.260	8-S	1,601.280	54,354.730
7-M	1,526.068	1,817.190			
7-N	510.444	14,226.260	9-A	93,072.730	14,805.938
7-O	402.837	37,351.540	9-B	162,819.210	25,494.532
7-P	20.931	45,375.820	9-C	128,706.780	17,065.958
7-Q	0.259	21,130.110	9-D	24,038.080	5,237.054
			9-E	13,932.910	3,219.931
8-A	674.085	16,107.780	9 F	65,345.870	11,367.214
8-B	1,634.623	1,603.730	9-G	55,068.050	31,082.836
8-C	13,245.577	8,189.230	9-H	98.960	16,974.181
8-D	35,714.201	8,008.290	9-I	180.420	1,706.103
8-E	45,961.650	2,466.020	9-J	31.300	27,974.904
8-F	22,246.940	623.850	9-K	1,291.250	205,019.690
8-G	504.830	52.140	9-L	6,213.070	496,864.273
8-H	12,308.570	2,213.720	9-M	6,432.520	578,171.661
8-I	19,326.020	6,419.820	9-N	2,465.440	261,656.803
8-J	8,954.840	6,830.910	9-O	396.210	2,512.917
8-K	1,075.350	6,053.320	9-P	51.720	124,445.967
8-L	1,050.670	22,470.910	9-Q	1,837.670	233,185.649
8-M	2,153.800	91,741.140	9-R	5,130.290	129,552.484
8-N	1,142.040	161,602.390	9-S	3,440.880	20,871.581
8-O	250.880	128,434.790	9-T	301.160	717.008
8-P	16.532	23,453.700	9-U	22,470.910	7,081.223

ond differences), the following sums of squares: $\bar{S}_2^{(k)} = 22.061$ for the selected items of the first differences, and $\bar{S}_2^{(k+1)} = 15.220$ for the selected items of the second differences. This whole process has been carried out in Table 30 for all selections up to the comparison of the ninth and the tenth differences, i.e., selections of order 9.

We have made the assumption that our original series consists of a "smooth" component which can be eliminated by successive differencing (Chapter IV, Section C) and a random component which is distributed normally about the mean zero with unknown variance. We proceed to test the hypothesis that we have eliminated the nonrandom or smooth element in the kth difference as follows: The ratio of the sum of squares of selected items of the kth difference, $\bar{S}_2^{(k)}$, to the sum of squares of the corresponding selected items of the $(k+1)$th difference, $\bar{S}_2^{(k+1)}$ (the selections being given in Table 26), must fall within the limits indicated in Tables 27, 28, and 29 if we want to be sure from the point of view of specific levels of significance that the two variances are really equal and that the difference arises by chance fluctuations.

This test does not involve the true variances since it is nothing but a modification of Fisher's z test. We also know that, generally speaking, we shall hardly ever have eliminated completely the "smooth" nonrandom component by taking finite differences. Frequently there will be some remainder besides the true random component. Hence, we shall be satisfied if our test yields satisfactory results in most cases in a comparison of the variances (or sums of squares) of the kth and $(k+1)$th differences and shall not insist that it should always hold true.

We give in Table 31 the ratios of the sums of squares for selected items of the differences of annual American wheat-flour prices, the selections being given in Table 26. For selection 0-A, for instance, the ratio of the sums of squares of the selected original items and of the first differences is equal to $\bar{S}_2^{(k)}/\bar{S}_2^{(k+1)} = 396.04/12.713$ (Table 30) which is approximately 31.15236 (Table 31). The ratio of the sums of squares for the selection 0-B is equal to $\bar{S}_2^{(k)}/\bar{S}_2^{(k+1)} = 350.36/21.471$ (Table 30) which is approximately 16.31782 (Table 31). Finally, the ratio for the selection 0-C is equal to 11.62037 (Table 31).

We proceed in the same way for the higher differences. For the selection 1-A (Table 26), the ratio of the sums of squares of the selected first differences and the selected second differences is $\bar{S}_2^{(k)}/\bar{S}_2^{(k+1)} = 22.061/15.220$ (Table 30) which is approximately 1.44947 (Table 31). In the same way, for selection 1-B the ratio is 3.43416; for selection 1-C, 0.16441; 1-D, 0.27489; 1-E, 0.20595; etc. All these ratios are given in Table 31.

In order to test the above hypothesis we have now to turn to Tables 27, 28, and 29. Can we say that the ratio resulting from selection 0-A is consistent with our hypothesis? Are the variances of the series of the original items and of the series of first differences approximately equal? Have we eliminated the nonrandom element by taking one difference? The answer to this question can only be given from the point of view of an arbitrarily chosen level of significance. If we take, for instance, the 5% level, this has the following meaning: A true hypothesis will be rejected, in the long run, in only 5% of the cases, if we apply our criterion. (See above, pages 33 f.)

TABLE 31

RATIOS OF SUMS OF SQUARES FOR SELECTED ITEMS
ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Selection	Ratio of Sums of Squares $\frac{\bar{S}_2^{(k)}}{\bar{S}_2^{(k+1)}}$	Selection	Ratio of Sums of Squares $\frac{\bar{S}_2^{(k)}}{\bar{S}_2^{(k+1)}}$	Selection	Ratio of Sums of Squares $\frac{\bar{S}_2^{(k)}}{\bar{S}_2^{(k+1)}}$
0-A	31.15236***	5-D	0.34378	7-Q	0.00001***
0-B	16.31782***	5-E	0.02288*		
0-C	11.62037***	5-F	0.00272***	8-A	0.04185
		5-G	0.05756	8-B	1.01926
1-A	1.44947*	5-H	0.25682	8-C	1.61744
1-B	3.43416***	5-I	0.00433***	8-D	4.45965*
1-C	0.16441	5-J	0.04420	8-E	18.63799**
1-D	0.27489	5-K	0.72081	8-F	35.66072***
1-E	0.20595	5-L	15.24123***	8-G	9.68220*
		5-M	8.05545**	8-H	5.56013*
2-A	0.26046			8-I	3.01037
2-B	0.39901	6-A	0.01596**	8-J	1.31093
2-C	0.14361	6-B	0.00421***	8-K	0.17765
2-D	0.06582	6-C	0.00329***	8-L	0.04676
2-E	0.45729	6-D	0.04802	8-M	0.02348
2-F	0.39138	6-E	0.11505	8-N	0.00707*
2-G	0.90786	6-F	0.01651**	8-O	0.00195***
		6-G	0.03574	8-P	0.00070***
3-A	0.61593	6-H	0.81162	8-Q	0.01194*
3-B	0.22396	6-I	106.78683***	8-R	0.01602*
3-C	0.34731	6-J	13.10944**	8-S	0.02946
3-D	0.05459	6-K	6.38514**		
3-E	0.01735**	6-L	0.00363***	9-A	6.28618*
3-F	33.61659**	6-M	3.33339*	9-B	6.38644*
3-G	0.20294	6-N	1475.50493***	9-C	7.54173*
3-H	0.57390	6-O	0.47650	9-D	4.59000
3-I	2.37788*			9-E	4.32708
		7-A	2.62951	9-F	5.74863*
4-A	0.19830	7-B	0.04097	9-G	1.77167
4-B	0.07381	7-C	0.02415	9-H	0.00583*
4-C	0.59821	7-D	0.00426**	9-I	0.10575
4-D	3.00484*	7-E	0.40437	9-J	0.00112***
4-E	2.23956*	7-F	4.71289*	9-K	0.00630*
4-F	0.02064**	7-G	11.10553*	9-L	0.01250*
4-G	1.65375	7-H	88.13076***	9-M	0.01113*
4-H	0.35500	7-I	8.55391**	9-N	0.00942*
4-I	0.02900*	7-J	1.21228	9-O	0.15767
4-J	0.02167*	7-K	3.50471*	9-P	0.00042***
4-K	0.06184	7-L	7.86539**	9-Q	0.00788*
		7-M	0.83980	9-R	0.03960
5-A	0.88299	7-N	0.03588	9-S	0.16486
5-B	2.31882	7-O	0.01079*	9-T	0.42002
5-C	7.59672**	7-P	0.00046***	9-U	3.17331

Note: * ratio outside the limit, level of significance 5.0%
** ratio outside the limit, level of significance 5.0% and 1.0%
*** ratio outside the limit, level of significance 5.0%, 1.0%, and 0.1%

If we consider it from the point of view of a level of significance of 5% we can test the hypothesis by entering Table 27, interpolating in it for the value of N equal to 48 and the value of k equal to 0. The ratio of the sums of squares $(\bar{S}_2^{(k)}/\bar{S}_2^{(k+1)})$ should fall within the limits 0.186 and 1.353, if the variances were equal. The value of the ratio for selection 0-A, 31.15236 (Table 31), evidently falls outside these limits. So does the ratio for selection 0-B which is 16.31782 and for selection 0-C which is 11.62037.

Table 28 gives the following limits for a level of significance 1% for $N=48$ and $k=0$; 0.135 and 1.856. All three ratios for the selections 0-A, 0-B, and 0-C, i.e., 31.15236, 16.31782, and 11.62037 (Table 31) lie outside these limits. We conclude, therefore, that our hypothesis is unjustified from the point of view of this level of significance.

Finally, turning to Table 29 we are dealing with the level of sigsignificance of 0.1%. By entering Table 29 (for N equal to 48 and k equal to 0) we find that the limits from this point of view are 0.0946 and 2.659. Again we see from Table 31 that all the ratios of sums of squares for selections 0-A, 0-B, and 0-C, that is to say 31.15236, 16.31782, and 11.62037, fall outside the limits permitted by the theory of probability. Hence, our hypothesis that the variance of the original series is equal to the variance of the series of first differences is probably not justified since all three possible selections give ratios of the sums of squares that fall outside the permitted limits for all levels of significance which we consider.

In Table 31 we have put one star after the values of ratios that fall outside the limits based upon the level of significance 5% but inside the limits based upon the levels of significance 1% and 0.1%. Ratios that fall outside the limits given by the levels of significance of 5% and 1%, but within the limits given by the level of significance 0.1%, are designated by two stars. Finally, the ratios that fall outside the limits on all three levels of significance have three stars. On the other hand, ratios that fall inside all three limits given by the levels of significance 5%, 1%, and 0.1% have no stars at all. The ratios for selections 0-A, 0-B, and 0-C have three stars because they do not fall within any of the limits given by the three levels of significance.

The same procedure is followed for comparison of higher differences. For instance, the ratios for selections 1-A, 1-B, 1-C, 1-D, and 1-E, are respectively, 1.44947, 3.43416, 0.16441, 0.27489, and 0.20595 (Table 31). Entering Table 27 under $N = 48$, $k = 1$, we see that the limits from the point of view of the level of significance of 5% are 0.0918 and 1.216. Entering Table 28, for $N = 48$ and $k = 1$, we find that the limits, from the point of view of a level of significance of 1%, are 0.0610 and 1.835. Finally we enter Table 29 which gives us the limits for a level of significance of 0.1%; for $N = 48$, and $k = 1$, we get the following limits: 0.0381 and 2.949.

We see from Table 31 that the first two of our five ratios for the sums of squares of selected values of the first differences and second differences, that is to say, 1-A with the ratio 1.44947 and 1-B with the ratio 3.43416, fall definitely outside the limits given by the level of significance 5%. The ratio 1-A falls inside the limits given by the levels of significance 1% and 0.1% (one star). The ratio 1-B falls outside all limits (three stars). The majority of our ratios, that is to say 1-C, 1-D, and 1-E, come within the limits given by all levels of significance and hence have no stars at all. But this may not be sufficient. We conclude, therefore, that the hypothesis that the variances of the first and of the second differences are equal is probably not yet justified. This result, like the previous one, is the same as that reached by the cruder comparisons of the variate difference method as given by Anderson, based upon standard errors of the differences between the variances of two consecutive series of finite differences without making any selections (Table 22).

In Table 31 also appear the ratios for higher selections. All comparisons of the ratios of the sums of squares of selected values of the second and third differences, selections 2-A, 2-B, etc. fall within the limits given by all levels of significance. This also agrees with the results established previously in Chapter VII by the method of standard errors (Table 22). We should, therefore, conclude that the variances of the series of the second and of the third differences are probably equal or that we have already eliminated in the second or third difference to a considerable (and possibly sufficient) degree the smooth non-random element.

A glance over Table 31 confirms this conclusion. But we must keep in mind the fact that the number of items in our higher selections becomes rather small. Originally we had 48 items in our series. Selections of order 0, that is, 0-A, 0-B, and 0-C, bring this number down to about 15 items. Selections of order 1, that is, 1-A, 1-B, etc., bring it down to as low as 8 items. Some of the comparisons based on the selections of order 2, that is 2-A, 2-B, etc., are based on not more than 6 items. Generally, a selection of order k is based only upon $(N-k-1)/(2k+3)$ items, if the original series contained N items.

Hence, in our case, some selections of order 3 contain no more than 4 items, some selections of order 5 only 2 items, and selections of higher orders also 2 items, until, in order 7, 8, or 9, we occasionally get down to 1 item. This is, of course, a very meager basis of comparison and we shall expect only most, not all, of our ratios to fall within the preassigned limits. Our differences still contain some remainder of the original nonrandom element which can never be completely eliminated. The condition of a normal distribution of the random element is also only partly fulfilled (Appendix VII).

These are probably only slight inconsistencies, which we should expect in a short series such as the series of annual wheat-flour prices which contains only 48 items to start with. The assumption of normality is also not strictly fulfilled. But we should not be wrong in thinking that the hypothesis is more or less justified, that we have eliminated to a considerable extent our nonrandom component in the second or third differences, and, hence, that the variances of the third, fourth, fifth, and higher differences are more or less equal within the limits of probability. It is interesting to note that we have established the same results in the previous chapter from the point of view of Anderson's procedure by using the cruder methods of the standard error of the difference between the variances of two consecutive series of finite differences, applicable only in large samples (Table 22).

We propose to call our procedure the method of selection. It consists in making comparisons between independent items by selecting them out of our total material in such a way that they become independent. This method will later be used again for the study of correlation between time series (Chapter X, Section B). It is the only method with which we are acquainted that is available for this and similar purposes. We have pointed out before that it reduces the number of degrees of freedom considerably by reducing the number of items contained in an individual comparison. But this loss of information may be compensated for by the fact that we get an exact test of significance for the hypothesis that the variances of two consecutive difference series are equal. This test is correct and unbiased from the point of view of probability and modern statistics.

CHAPTER IX

REDUCTION OF THE RANDOM VARIATION BY SHEPPARD'S SMOOTHING FORMULAE

We may be interested in the extent of the random element in our time series as represented for instance by a coefficient of variability. This coefficient of random variability ($v = \sqrt{V_{k_0}}/\bar{w}$) is formed by dividing the random standard deviation, which is the square root of the random variance [in our case of the annual wheat-flour prices, $\sqrt{V_2} = \sqrt{0.4402} = 0.6635$ (Table 17)], by the arithmetic mean [which is in our case $\bar{w} = 6.0731$ (Table 8)]. Hence we get for the coefficient of random variability of the annual wheat-flour prices: $v = 0.6635/6.0731 = 0.1093$, if we assume that the nonrandom element is eliminated in the second differences (the difference beginning from which the variances become more or less stable, $k_0 = 2$). We may want to improve our statistical series and to eliminate the random element at least as far as possible.

It follows from the very definition of the random or erratic element that we can never find a true approximation to it as a function of time and then subtract it from our original series in order to get the mathematical expectation or the pure result of the permanent causes which are effective in our time series.[1] But it has been shown by Professor Anderson that we can eliminate the random element to any desired degree by using the so-called Sheppard's smoothing formulae.[2] (See also Appendix II, pp. 144 ff., and Appendix I, Section F, for a summary of computations.)

The main idea of Sheppard's smoothing formulae, which were

[1] O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, Bonn, 1929, pp. 72 ff.

[2] O. Anderson, *op. cit.*, pp. 74 ff., pp. 117 ff. See also E. T. Whittaker and G. Robinson, *The Calculus of Observations*, London, 1924, pp. 291 ff.; W. F. Sheppard, "Reduction of Errors by Means of Negligible Differences," *Proceedings of the Fifth International Congress of Mathematicians*, Cambridge, 1912, Vol. 2, pp. 348 ff.; W. F. Sheppard, "Fitting of Polynomial by Method of Least Squares," *Proceedings of the London Mathematical Society*, Second Series, Vol. 13, 1914, pp. 97 ff.; W. F. Sheppard, "Graduation by Reduction of Mean Square of Error," *Journal of the Institute of Actuaries*, Vol. 48, 1914, pp. 171 ff., 390 ff.; *ibid.*, Vol. 49, 1915, pp. 148 ff.; W. M. Sherriff, "On a Class of Graduation Formulae," *Proceedings of the Royal Society of Edinburgh*, Vol. 40, 1919, pp. 112 ff. See also on moving averages H. Wold, *A Study in the Analysis of Stationary Time Series*, Uppsala, 1938, pp. 121 ff.; F. R. Macaulay, *The Smoothing of Time Series*, New York, 1931.

TABLE 32

WEIGHTS FOR SHEPPARD'S SMOOTHING FORMULA: $g_{1,m}(j)$
$n = 1$: Straight Line, Mathematical Expectation Eliminated in the First or Second Difference
(j = Distance from Midpoint, m = Degree of Accuracy)

m	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
1	0.333 3333	0.333 3333	....	....	....	....
2	0.200 0000	0.200 0000	0.200 0000	....	....	....
3	0.142 8571	0.142 8571	0.142 8571	0.142 8571	....	....
4	0.111 1111	0.111 1111	0.111 1111	0.111 1111	0.111 1111	....
5	0.090 9091	0.090 9091	0.090 9091	0.090 9091	0.090 9091	0.090 9091
6	0.076 9231	0.076 9231	0.076 9231	0.076 9231	0.076 9231	0.076 9231
7	0.066 6667	0.066 6667	0.066 6667	0.066 6667	0.066 6667	0.066 6667
8	0.058 8235	0.058 8235	0.058 8235	0.058 8235	0.058 8235	0.058 8235
9	0.052 6316	0.052 6316	0.052 6316	0.052 6316	0.052 6316	0.052 6316
10	0.047 6190	0.047 6190	0.047 6190	0.047 6190	0.047 6190	0.047 6190

m	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$	
6	0.076 9231	....	....	....	....	....
7	0.066 6667	0.066 6667	....	....	....	....
8	0.058 8235	0.058 8235	0.058 8235	....	....	....
9	0.052 6316	0.052 6316	0.052 6316	0.052 6316	....	....
10	0.047 6190	0.047 6190	0.047 6190	0.047 6190	0.047 6190	....

first developed by Mr. W. F. Sheppard, quite independently of the variate difference method, for use in his actuarial work, is the following: We have to distinguish between the type of curve fitted and the accuracy of the fit, which are two entirely different things. The type of curve fitted is a straight line, a parabola of the second degree, of the third degree (cubic), etc.

Let us suppose that we want to fit a straight line. Then this straight line can be fitted to three points, or five points, etc. (any odd number of points) of the series, according to the degree of accuracy desired. Let us assume that we decide to fit a straight line to five successive points and that this is accurate enough for our particular purposes. Then the process of mechanical smoothing according to Sheppard is as follows: We take the first five points of our series and fit a straight line according to the method of least squares, i.e., minimizing the sum of the squares of the deviations from the fitted curve. We establish only the value of the midpoint, which is in our case item number three. This is the smoothed value of the third item.

Then we take again five consecutive points or items of our series,

TABLE 32 (continued)

WEIGHTS FOR SHEPPARD'S SMOOTHING FORMULA: $g_{2,m}(j)$

$n = 2$: Parabola of Second Degree, Mathematical Expectation Eliminated in the Third or Fourth Difference

(j = Distance from Midpoint, m = Degree of Accuracy)

m	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
2	0.485 7143	0.342 8571	–0.085 7143	....	....	....
3	0.333 3333	0.285 7143	0.142 8571	–0.095 2381	....	....
4	0.255 4113	0.233 7662	0.168 8312	0.060 6061	–0.090 9091	....
5	0.207 4592	0.195 8042	0.160 8392	0.102 5641	0.020 9790	–0.083 9161
6	0.174 8252	0.167 8322	0.146 8531	0.111 8881	0.062 9371	0.000 0000
7	0.151 1312	0.146 6063	0.133 0317	0.110 4072	0.078 7330	0.038 0090
8	0.133 1269	0.130 0310	0.120 7430	0.105 2632	0.083 5913	0.055 7276
9	0.118 9739	0.116 7625	0.110 1283	0.099 0712	0.083 5913	0.063 6886
10	0.107 5515	0.105 9170	0.101 0134	0.092 8408	0.081 3992	0.066 6885
11	0.098 1366	0.096 8944	0.093 1677	0.086 9565	0.078 2609	0.067 0807

m	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$	$j=11$
6	–0.076 9231	....	....	....	....	....
7	–0.011 7647	–0.070 5882	....	....	....	....
8	0.021 6718	–0.018 5759	–0.065 0155	....	....	....
9	0.039 3631	0.010 6148	–0.022 5564	–0.060 1504	....	....
10	0.048 7087	0.027 4600	0.002 9421	–0.024 8447	–0.055 9006	....
11	0.053 4161	0.037 2671	0.018 6335	–0.002 4845	–0.026 0870	–0.052 1739

but now items two to six. Again we fit a straight line according to the method of least squares. This gives us the smoothed value for our item number four, which is the midpoint of this series of five items. Next we take items number four to eight, and fit again a straight line according to the method of least squares. This gives us the smoothed value for item number five, etc. The whole process can, of course, be converted into smoothing by a moving average with certain weights which have been established by W. F. Sheppard and Miss Sherriff. They are represented in Table 32.[3] There is also a very interesting connection between the weights of these smoothing formulae as shown in Table 32 and the coefficients of the Gram polynomials given by Professor H. T. Davis.[4] (See also below, Appendix II, pp. 145 ff.)

The same general process would be applied if we should decide to smooth our series not with a straight line, but with a parabola of any

[3] W. M. Sherriff, *op. cit.*, p. 117; E. T. Whittaker and G. Robinson, *op. cit.*, pp. 295 ff.; O. Anderson, *op. cit.*, pp. 120 ff.

[4] H. T. Davis, *Tables of Higher Mathematical Functions*, Vol. II, Bloomington, Indiana, 1935, pp. 307 ff.

TABLE 32 (continued)

WEIGHTS FOR SHEPPARD'S SMOOTHING FORMULA: $g_{3,m}(j)$
$n = 3$: Cubic, Mathematical Expectation Eliminated in the Fifth or Sixth Difference
(j = Distance from Midpoint, m = Degree of Accuracy)

m	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
3	0.567 0995	0.324 6753	−0.129 8701	0.0216 450	....	....
4	0.417 2494	0.314 6853	0.069 9301	−0.128 2051	0.034 9650	....
5	0.333 3333	0.279 7203	0.139 8601	−0.023 3100	−0.104 8951	0.041 9580
6	0.278 4862	0.246 8120	0.160 4278	0.045 2489	−0.055 5327	−0.081 4480
7	0.239 5159	0.219 2080	0.162 3763	0.081 2964	−0.003 5723	−0.063 5866
8	0.210 2882	0.196 4754	0.157 1803	0.098 8331	0.032 1505	−0.027 8638
9	0.187 5084	0.177 6821	0.149 4145	0.106 3400	0.054 5161	0.002 4229
10	0.169 2325	0.161 9907	0.140 9919	0.108 4168	0.067 8999	0.024 5294
11	0.154 2334	0.148 7414	0.132 7231	0.107 5515	0.075 5149	0.039 8169
12	0.141 6958	0.137 4313	0.124 9375	0.105 1141	0.079 4603	0.050 0750

m	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$	$j=11$
6	0.045 2488	....	....	....	....	....
7	−0.061 9195	0.046 4397	....	....	....	....
8	−0.061 9195	−0.046 4396	0.046 4396	....	....	....
9	−0.039 0362	−0.056 5352	−0.034 3249	0.045 7666	....	....
10	−0.015 1530	−0.043 1514	−0.050 0163	−0.024 8447	0.044 7205	....
11	0.004 5767	−0.025 1716	−0.043 4783	−0.043 4783	−0.017 3913	0.043 4783
12	0.019 6568	−0.008 4958	−0.030 4848	−0.041 8124	−0.037 3813	−0.011 4943

m	$j=12$					
12	0.042 1456	....	....	....	....	....

degree. Again we have also to establish the degree of accuracy which is necessary in our case. We may want to fit a parabola, say of the second degree, to five items, or seven items, etc. Suppose we think the fit of a parabola to five items is accurate enough. Then we take the first five items of our series and fit a parabola to them, which gives us only the smoothed value for the third item, as before. We next take items number two to six, and fit to those five items again a parabola of the second degree by the method of least squares, which gives us the smoothed value for item number four, etc.

The same process holds true for parabolas of higher degree. All those smoothing processes can be replaced by applying various moving averages with different weights. Those weights are represented in Table 32. The n in Table 32 shows the type of curve which we want

TABLE 32 (continued)

WEIGHTS FOR SHEPPARD'S SMOOTHING FORMULA: $g_{4,m}(j)$

$n = 4$: Parabola of Fourth Degree, Mathematical Expectation Eliminated in the Seventh or Eighth Difference

(j = Distance from Midpoint, m = Degree of Accuracy)

m	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
4	0.619 2696	0.304 5843	–0.152 2922	0.043 5120	–0.005 4390	....
5	0.475 9358	0.322 5010	0.011 5179	–0.126 6968	0.066 2279	–0.011 5179
6	0.391 0671	0.303 1024	0.098 5083	–0.075 7756	–0.078 0489	0.073 3508
7	0.333 3333	0.277 8439	0.138 9220	–0.011 1138	–0.086 1316	–0.037 7868
8	0.291 0527	0.253 6836	0.155 8342	0.036 2405	–0.053 9984	–0.071 6113
9	0.258 5812	0.232 1712	0.160 9369	0.067 0884	–0.018 2797	–0.064 0732
10	0.232 7844	0.213 4099	0.160 0574	0.086 1848	0.011 3699	–0.042 9668

m	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$	
6	–0.016 6706	....	....	....	....	....
7	0.072 2394	–0.020 6398	....	....	....	....
8	–0.009 4225	0.067 3038	–0.023 5563	....	....	....
9	–0.051 6355	0.009 2610	0.060 8696	–0.025 6293	....	....
10	–0.059 5504	–0.032 9519	0.021 0018	0.054 1063	–0.027 0531	....

to fit. That is $n = 1$ represents a straight line, $n = 2$ a parabola of the second degree, $n = 3$ a parabola of the third degree (cubic), etc. The m represents the degree of accuracy. That is, if $m = 1$, we fit our curve to three consecutive items; if $m = 2$, to five consecutive items; if $m = 3$, to seven consecutive items, etc. The number of items in our moving average is $2m+1$.

This whole process of smoothing with moving averages[5] will be understood to be quite in the spirit of the variate difference method (cf. Chapter I, Section C). It is consistent with our assumptions about the random element. It will be remembered that our contention was that the random element is the result of the less permanent causes which affect only one or a few neighboring items of our series. Hence it is evident that their effect could be eliminated by some process of mechanical smoothing which would take out everything whose influence is not permanent, but is restricted to one or a few neighboring items.

The variate difference method accomplishes only one thing in this connection. It shows us, so to speak, how deeply rooted the influence of those nonrandom causes is in our series. It gives us an indication as to which type of curve to choose if we want to eliminate the ran-

[5] F. C. Mills, *Statistical Methods*, revised edition, New York, 1938, pp. 234 ff.

TABLE 32 (concluded)

WEIGHTS FOR SHEPPARD'S SMOOTHING FORMULA: $g_{5,m}(j)$
$n = 5$: Parabola of Fifth Degree, Mathematical Expectation Eliminated in the Ninth or Tenth Difference
(j = Distance from Midpoint, m = Degree of Accuracy)

m	$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
5	0.656 2818	0.286 4318	–0.163 6753	0.061 3783	–0.013 6396	0.001 3640
6	0.519 8857	0.321 5051	–0.032 1505	–0.110 7407	0.085 7347	–0.027 8638
7	0.436 0148	0.314 5159	0.058 2437	–0.102 3536	–0.040 0717	0.083 9641
8	0.377 3052	0.296 8098	0.110 2436	–0.053 7084	–0.084 8028	0.008 4803
9	0.333 3333	0.277 0225	0.138 5112	–0.006 5958	–0.079 1493	–0.048 0549
10	0.298 9443	0.257 9175	0.152 8400	0.030 2496	–0.055 1706	–0.067 8227

m	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$	
6	0.003 5723	....	....	....	....	....
7	–0.038 3632	0.006 0573	....	....	....	....
8	0.070 6690	–0.044 8243	0.008 4803	....	....	....
9	0.034 8634	0.054 1123	–0.048 0549	0.010 6789	....	....
10	–0.015 1410	0.045 9244	0.038 1125	–0.048 9755	0.012 5937	....

dom element to a certain degree (it determines n). The variate difference method, however, does not give us a criterion for the accuracy with which this should be done(the m is still arbitrary).[6]

It follows from Professor Anderson's exposition[7] that the type of moving average which should be chosen is determined in the following way: If the nonrandom element is already more or less eliminated in the first or second finite difference, we may take $n=1$, or a moving average which is equivalent to fitting a straight line to a number of consecutive items. If the mathematical expectation is eliminated only in the third or fourth finite difference, we have $n=2$, and we choose a moving average which is equivalent to fitting a parabola of the second degree to a selected number of consecutive items. If the nonrandom element is eliminated in the fifth or sixth finite difference, we choose a moving average which is equivalent to fitting a parabola of the third degree (a cubic) to a selected number of consecutive items ($n=3$) etc. If the nonrandom element is eliminated in the k_0th finite difference, $n = k_0/2$ for k_0 even, or $n = (k_0+1)/2$ for k_0 odd.[8]

[6] See, however, R. Zaycoff, "Ueber die Ausschaltung der zufälligen Komponente nach der 'Variate-Difference' Methode," *Publications of the Statistical Institute for Economic Research*, State University of Sofia, 1937, No. 1, p. 82.
[7] O. Anderson, *op. cit.*, p. 74 ff.
[8] R. Zaycoff, *op. cit.*, pp. 82 ff.

Even if the nonrandom element or the mathematical expectation is eliminated to a considerable degree in the k_0th difference, we cannot represent the mathematical expectation of the *whole* time series by a parabola of degree k_0—1 whose k_0th difference would vanish (Chapter IV, Section C). The variate difference method tells us only something about the *approximate* behavior of a time series *in the neighborhood of a point* and not about the whole shape of it. Hence all we can say is that in the small (*im kleinen*) the series behaves like a certain parabola and we can hence use Sheppard's smoothing formulae of a given type n as defined. But we cannot make any statement about the development of the series over the whole period considered. The variate difference method gives us no indication about the behavior of the series in the large (*im grossen*). This is one of the most important distinctions between the variate difference method and the fitting of orthogonal functions.

The number of items to which we fit those parabolas or which we include in our moving averages is still arbitrary. We shall have to balance in this case the advantage of greater accuracy against the greater labor of using longer moving averages. Tables 33 and 34 should be helpful in this connection.

We have represented in Table 33 the degree to which we reduce

TABLE 33

COEFFICIENT L_{nm} FOR THE REDUCTION OF THE RANDOM VARIANCE BY SMOOTHING WITH A MOVING AVERAGE OF TYPE n AND ACCURACY m

Type of Curve	Degree of Accuracy					
n	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$
1	0.333 3333	0.200 0000	0.142 8571	0.111 1111	0.090 9091	0.076 9231
2	....	0.485 7143	0.333 3333	0.255 4113	0.207 4592	0.174 8252
3	....	....	0.567 0995	0.417 2494	0.333 3333	0.278 4862
4	....	....	....	0.619 2696	0.475 9358	0.391 0671
5	....	....	....	....	0.656 2818	0.519 8857

Type of Curve	Degree of Accuracy					
n	$m=7$	$m=8$	$m=9$	$m=10$	$m=11$	$m=12$
1	0.066 6667	0.058 8235	0.052 6316	0.047 6190	....	....
2	0.151 1312	0.133 1269	0.118 9739	0.107 5515	0.098 1366	....
3	0.239 5159	0.210 2882	0.187 5084	0.169 2325	0.154 2334	0.141 6958
4	0.333 3333	0.291 0527	0.258 5812	0.232 7844	....	....
5	0.436 0148	0.377 3052	0.333 3333	0.298 9443	....	....

the variance of the random element if we use a moving average of a certain length. In our case of annual wheat-flour prices, for instance, we can take $n=2$, assuming that the mathematical expectation is eliminated only in the third finite difference (Table 22). It follows from Table 33[9] that we can reduce our random variance approximately to 49 per cent, simply by using the approximation $m=2$, which involves a moving average including five items. L_{nm} is the coefficient by which the original random variance V_{k_0} is multiplied if we use a moving average of type n and accuracy m. We can reduce our random variance to one-third if we make $m=3$ or use a moving average which includes seven items. We can reduce the random variance to about 26 per cent if we use a moving average which includes nine items ($m=4$). We can reduce it to approximately 21 per cent if we use a moving average which includes 11 items ($m=5$), etc. (Table 33).

We have, in our case, chosen a moving average which includes 7 items and which reduces the random variance to one-third ($n=2$, $m=3$, Table 33). We consider this value as sufficiently accurate for our purposes. The remaining random variance of our series of annual wheat-flour prices should hence be approximately 0.15, whereas the original one was about 0.44 (Table 17). This is, in our opinion, enough reduction. A greater accuracy would imply much greater labor since the moving averages become longer.

We give in Table 34 the reduction of the coefficient of random variability v. Entering it for $v=0.10$ and $n=2$, $m=3$, we see that the coefficient of the remaining random variability after the smoothing is about 0.058 or less than 6%.

We have arranged in Table 32 the values of the weights $g_{nm}(j)$ of the moving averages in the following way: j is the distance of the item from the midpoint. In our case, $n=2$, $m=3$, that is, we have decided to use a moving average which is equivalent to fitting a parabola of the second degree to seven consecutive items. The item in the middle ($j=0$) is multiplied by 0.333 3333 (Table 32). The items distant by 1 from the midpoint ($j=1$) are multiplied by 0.285 7143, the items distant by 2 from the midpoint ($j=2$) are multiplied by 0.142 8571. Finally, the items distant by 3 from the midpoint ($j=3$) are multiplied by —0.095 2381. It follows from the nature of moving averages that the sum of weights must always be equal to 1.

The first smoothed value (m') which we establish by using our moving averages is for item number four. We get it by using a moving average including the first seven items of our series of annual

[9] R. Zaycoff, *op. cit.*, p. 85, Table 4.

TABLE 34

REDUCTION OF THE COEFFICIENT OF RANDOM VARIABILITY (v) BY THE USE OF MOVING AVERAGES

($n = 1$, Straight Line; m = Degree of Accuracy)

m	$v = 0.02$	$v = 0.04$	$v = 0.06$	$v = 0.08$	$v = 0.10$
1	0.011 5470	0.023 0940	0.034 6410	0.046 1880	0.057 7350
2	0.008 9443	0.017 8885	0.026 8328	0.035 7771	0.044 7214
3	0.007 5593	0.015 1186	0.022 6779	0.030 2372	0.037 7964
4	0.006 6667	0.013 3333	0.020 0000	0.026 6667	0.033 3333
5	0.006 0302	0.012 0605	0.018 0907	0.024 1209	0.030 1511
6	0.005 5470	0.011 0940	0.016 6410	0.022 1880	0.027 7350
7	0.005 1640	0.010 3280	0.015 4919	0.020 6559	0.025 8199
8	0.004 8507	0.009 7014	0.014 5521	0.019 4028	0.024 2536
9	0.004 5883	0.009 1766	0.013 7649	0.018 3533	0.022 9416
10	0.004 3644	0.008 7287	0.013 0931	0.017 4574	0.021 8218

m	$v = 0.12$	$v = 0.14$	$v = 0.16$	$v = 0.18$	$v = 0.20$
1	0.069 2820	0.080 8290	0.092 3760	0.103 9230	0.115 4700
2	0.053 6656	0.062 6099	0.071 5542	0.080 4984	0.089 4427
3	0.045 3557	0.052 9150	0.060 4743	0.068 0336	0.075 5929
4	0.040 0000	0.046 6667	0.053 3333	0.060 0000	0.066 6667
5	0.036 1814	0.042 2116	0.048 2418	0.054 2720	0.060 3023
6	0.033 2820	0.038 8290	0.044 3760	0.049 9230	0.055 4700
7	0.030 9839	0.036 1478	0.041 3118	0.046 4758	0.051 6398
8	0.029 1043	0.033 9550	0.038 8057	0.043 6564	0.048 5071
9	0.027 5299	0.032 1182	0.036 7065	0.041 2948	0.045 8831
10	0.026 1861	0.030 5505	0.034 9148	0.039 2792	0.043 6436

m	$v = 0.25$	$v = 0.30$	$v = 0.35$	$v = 0.40$	$v = 0.45$	$v = 0.50$
1	0.144 3376	0.173 2051	0.202 0726	0.230 9401	0.259 8076	0.288 6751
2	0.111 8034	0.134 1641	0.156 5248	0.178 8854	0.201 2461	0.223 6068
3	0.094 4911	0.113 3893	0.132 2875	0.151 1858	0.170 0840	0.188 9822
4	0.083 3333	0.100 0000	0.116 6667	0.133 3333	0.150 0000	0.166 6667
5	0.075 3778	0.090 4534	0.105 5290	0.120 6045	0.135 6801	0.150 7556
6	0.069 3375	0.083 2050	0.097 0725	0.110 9400	0.124 8075	0.138 6750
7	0.064 5497	0.077 4597	0.090 3696	0.103 2796	0.116 1895	0.129 0994
8	0.060 6339	0.072 7607	0.084 8875	0.097 0142	0.109 1410	0.121 2678
9	0.057 3539	0.068 8247	0.080 2955	0.091 7663	0.103 2371	0.114 7079
10	0.054 5544	0.065 4653	0.076 3762	0.087 2871	0.098 1980	0.109 1089

TABLE 34 (continued)

REDUCTION OF THE COEFFICIENT OF RANDOM VARIABILITY (v) BY THE USE OF MOVING AVERAGES

($n = 2$, Parabola of Second Degree; m = Degree of Accuracy)

m	$v = 0.02$	$v = 0.04$	$v = 0.06$	$v = 0.08$	$v = 0.10$
2	0.013 9386	0.027 8773	0.041 8159	0.055 7546	0.069 6932
3	0.011 5470	0.023 0940	0.034 6410	0.046 1880	0.057 7350
4	0.010 1076	0.020 2153	0.030 3229	0.040 4306	0.050 5382
5	0.009 1095	0.018 2191	0.027 3286	0.036 4382	0.045 5477
6	0.008 3624	0.016 7248	0.025 0873	0.033 4497	0.041 8121
7	0.007 7751	0.015 5502	0.023 3254	0.031 1005	0.038 8756
8	0.007 2973	0.014 5946	0.021 8919	0.029 1892	0.036 4866
9	0.006 8985	0.013 7970	0.020 6956	0.027 5941	0.034 4926
10	0.006 5590	0.013 1180	0.019 6770	0.026 2360	0.032 7950
11	0.006 2654	0.012 5307	0.018 7961	0.025 0614	0.031 3268

m	$v = 0.12$	$v = 0.14$	$v = 0.16$	$v = 0.18$	$v = 0.20$
2	0.083 6318	0.097 5705	0.111 5091	0.125 4478	0.139 3864
3	0.069 2820	0.080 8290	0.092 3760	0.103 9230	0.115 4700
4	0.060 6459	0.070 7535	0.080 8612	0.090 9688	0.101 0765
5	0.054 6572	0.063 7668	0.072 8763	0.081 9858	0.091 0954
6	0.050 1745	0.058 5369	0.066 8994	0.075 2618	0.083 6242
7	0.046 6507	0.054 4258	0.062 2009	0.069 9761	0.077 7512
8	0.043 7839	0.051 0812	0.058 3785	0.065 6758	0.072 9731
9	0.041 3911	0.048 2896	0.055 1881	0.062 0867	0.068 9852
10	0.039 3540	0.045 9131	0.052 4721	0.059 0311	0.065 5901
11	0.037 5921	0.043 8575	0.050 1228	0.056 3882	0.062 6535

m	$v = 0.25$	$v = 0.30$	$v = 0.35$	$v = 0.40$	$v = 0.45$	$v = 0.50$
2	0.174 2330	0.209 0796	0.243 9262	0.278 7728	0.313 6194	0.348 4660
3	0.144 3376	0.173 2051	0.202 0726	0.230 9401	0.259 8076	0.288 6751
4	0.126 3456	0.151 6147	0.176 8838	0.202 1529	0.227 4220	0.252 6912
5	0.113 8692	0.136 6431	0.159 4169	0.182 1908	0.204 9646	0.227 7384
6	0.104 5302	0.125 4363	0.146 3424	0.167 2484	0.188 1544	0.209 0605
7	0.097 1890	0.116 6268	0.136 0646	0.155 5024	0.174 9402	0.194 3780
8	0.091 2164	0.109 4596	0.127 7029	0.145 9462	0.164 1895	0.182 4328
9	0.086 2315	0.103 4778	0.120 7241	0.137 9704	0.155 2167	0.172 4630
10	0.081 9876	0.098 3851	0.114 7826	0.131 1802	0.147 5777	0.163 9752
11	0.078 3169	0.093 9803	0.109 6437	0.125 3070	0.140 9704	0.156 6338

TABLE 34 (continued)

REDUCTION OF THE COEFFICIENT OF RANDOM VARIABILITY (v) BY THE USE OF MOVING AVERAGES

($n = 3$, Cubic; m = Degree of Accuracy)

m	$v = 0.02$	$v = 0.04$	$v = 0.06$	$v = 0.08$	$v = 0.10$
3	0.004 7628	0.009 5255	0.014 2888	0.019 0511	0.023 8138
4	0.004 0853	0.008 1707	0.012 2560	0.016 3413	0.020 4267
5	0.003 6515	0.007 3030	0.010 9544	0.014 6059	0.018 2574
6	0.003 3376	0.006 6752	0.010 0127	0.013 3503	0.016 6879
7	0.003 0953	0.006 1905	0.009 2858	0.012 3810	0.015 4763
8	0.002 9003	0.005 8005	0.008 7008	0.011 6011	0.014 5013
9	0.002 7387	0.005 4773	0.008 2160	0.010 9547	0.013 6934
10	0.002 6018	0.005 2036	0.007 8054	0.010 4072	0.013 0089
11	0.002 4838	0.004 9676	0.007 4514	0.009 9353	0.012 4191
12	0.002 3807	0.004 7614	0.007 1422	0.009 5229	0.011 9036

m	$v = 0.12$	$v = 0.14$	$v = 0.16$	$v = 0.18$	$v = 0.20$
3	0.028 5766	0.033 3394	0.038 1022	0.042 8649	0.047 6277
4	0.024 5120	0.028 5974	0.032 6827	0.036 7680	0.040 8534
5	0.021 9089	0.025 5604	0.029 2119	0.032 8634	0.036 5148
6	0.020 0255	0.023 3631	0.026 7007	0.030 0382	0.033 3758
7	0.018 5716	0.021 6668	0.024 7621	0.027 8573	0.030 9526
8	0.017 4016	0.020 3018	0.023 2021	0.026 1024	0.029 0026
9	0.016 4320	0.019 1707	0.021 9094	0.024 6481	0.027 3867
10	0.015 6107	0.018 2125	0.020 8143	0.023 4161	0.026 0179
11	0.014 9029	0.017 3867	0.019 8705	0.022 3543	0.024 8382
12	0.014 2843	0.016 6650	0.019 0458	0.021 4265	0.023 8072

m	$v = 0.25$	$v = 0.30$	$v = 0.35$	$v = 0.40$	$v = 0.45$	$v = 0.50$
3	0.059 5346	0.071 4416	0.083 3485	0.095 2554	0.107 1623	0.119 0692
4	0.051 0667	0.061 2800	0.071 4934	0.081 7067	0.091 9201	0.102 1334
5	0.045 6436	0.054 7723	0.063 9010	0.073 0297	0.082 1584	0.091 2871
6	0.041 7198	0.050 0637	0.058 4077	0.066 7516	0.075 0956	0.083 4396
7	0.038 6908	0.046 4289	0.054 1670	0.061 9052	0.069 6434	0.077 3815
8	0.036 2533	0.043 5040	0.050 7546	0.058 0053	0.065 2559	0.072 5066
9	0.034 2334	0.041 0801	0.047 9268	0.054 7735	0.061 6202	0.068 4668
10	0.032 5224	0.039 0268	0.045 5313	0.052 0358	0.058 5402	0.065 0447
11	0.031 0477	0.037 2572	0.043 4668	0.049 6763	0.055 8859	0.062 0954
12	0.029 7590	0.035 7108	0.041 6626	0.047 6144	0.053 5662	0.059 5180

TABLE 34 (continued)

REDUCTION OF THE COEFFICIENT OF RANDOM VARIABILITY (v) BY THE USE OF MOVING AVERAGES

($n = 4$, Parabola of Fourth Degree; m = Degree of Accuracy)

m	$v = 0.02$	$v = 0.04$	$v = 0.06$	$v = 0.08$	$v = 0.10$
4	0.015 7387	0.031 4775	0.047 2162	0.062 9549	0.078 6937
5	0.013 7976	0.027 5952	0.041 3929	0.055 1905	0.068 9881
6	0.012 5071	0.025 0141	0.037 5212	0.050 0283	0.062 5354
7	0.011 5470	0.023 0940	0.034 6410	0.046 1880	0.057 7350
8	0.010 7899	0.021 5797	0.032 3696	0.043 1594	0.053 9493
9	0.010 1702	0.020 3404	0.030 5105	0.040 6807	0.050 8509
10	0.009 6495	0.019 2991	0.028 9486	0.038 5982	0.048 2477

m	$v = 0.12$	$v = 0.14$	$v = 0.16$	$v = 0.18$	$v = 0.20$
4	0.094 4324	0.110 1712	0.125 9099	0.141 6486	0.157 3874
5	0.082 7857	0.096 5833	0.110 3810	0.124 1786	0.137 9762
6	0.075 0424	0.087 5495	0.100 0566	0.112 5636	0.125 0707
7	0.069 2820	0.080 8290	0.092 3760	0.103 9230	0.115 4700
8	0.064 7392	0.075 5290	0.086 3189	0.097 1087	0.107 8986
9	0.061 0211	0.071 1912	0.081 3614	0.091 5316	0.101 7018
10	0.057 8973	0.067 5468	0.077 1964	0.086 8459	0.096 4955

m	$v = 0.25$	$v = 0.30$	$v = 0.35$	$v = 0.40$	$v = 0.45$	$v = 0.50$
4	0.196 7342	0.236 0810	0.275 4279	0.314 7747	0.354 1216	0.393 4684
5	0.172 4702	0.206 9643	0.241 4584	0.275 9524	0.310 4464	0.344 9405
6	0.156 3384	0.187 6060	0.218 8737	0.250 1414	0.281 4091	0.312 6768
7	0.144 3376	0.173 2051	0.202 0726	0.230 9401	0.259 8076	0.288 6751
8	0.134 8732	0.161 8479	0.188 8226	0.215 7972	0.242 7718	0.269 7465
9	0.127 1272	0.152 5526	0.177 9781	0.203 4035	0.228 8290	0.254 2544
10	0.120 6193	0.144 7432	0.168 8671	0.192 9909	0.217 1148	0.241 2386

wheat-flour prices (w) (Table 35). The first item (5.185) is to be multiplied by —0.0952, the second item (5.305) is to be multiplied by 0.1429, the third item (4.346) is to be multiplied by 0.2857, the fourth item which is at the midpoint at the year 1893, (4.006) is to be multiplied by 0.3333. The fifth item (3.594) has to be multiplied by 0.2857, the sixth item (3.643) by 0.1429, and the seventh item (3.795) by —0.0952.

The process of calculation is exhibited in Table 35. All the mov-

TABLE 34 (concluded)

REDUCTION OF THE COEFFICIENT OF RANDOM VARIABILITY (v) BY THE USE OF MOVING AVERAGES

($n = 5$, Parabola of Fifth Degree; m = Degree of Accuracy)

m	$v = 0.02$	$v = 0.04$	$v = 0.06$	$v = 0.08$	$v = 0.10$
5	0.016 2022	0.032 4045	0.048 6067	0.064 8090	0.081 0112
6	0.014 4206	0.028 8412	0.043 2619	0.057 6825	0.072 1031
7	0.013 2063	0.026 4126	0.039 6189	0.052 8251	0.066 0314
8	0.012 2850	0.024 5701	0.036 8551	0.049 1401	0.061 4252
9	0.011 5470	0.023 0940	0.034 6410	0.046 1880	0.057 7350
10	0.010 9352	0.021 8703	0.032 8055	0.043 7406	0.054 6758

m	$v = 0.12$	$v = 0.14$	$v = 0.16$	$v = 0.18$	$v = 0.20$
5	0.097 2135	0.113 4157	0.129 6180	0.145 8202	0.162 0224
6	0.086 5237	0.100 9443	0.115 3650	0.129 7856	0.144 2062
7	0.079 2377	0.092 4440	0.105 6503	0.118 8566	0.132 0628
8	0.073 7102	0.085 9952	0.098 2803	0.110 5653	0.122 8503
9	0.069 2820	0.080 8290	0.092 3760	0.103 9230	0.115 4700
10	0.065 6110	0.076 5461	0.087 4813	0.098 4164	0.109 3516

m	$v = 0.25$	$v = 0.30$	$v = 0.35$	$v = 0.40$	$v = 0.45$	$v = 0.50$
5	0.202 5280	0.243 0337	0.283 5393	0.324 0449	0.364 5505	0.405 0561
6	0.180 2578	0.216 3093	0.252 3608	0.288 4124	0.324 4640	0.360 5155
7	0.165 0786	0.198 0943	0.231 1100	0.264 1257	0.297 1414	0.330 1571
8	0.153 5629	0.184 2755	0.214 9881	0.245 7007	0.276 4133	0.307 1258
9	0.144 3376	0.173 2051	0.202 0726	0.230 9401	0.259 8076	0.288 6751
10	0.136 6895	0.164 0274	0.191 3653	0.218 7032	0.246 0411	0.273 3790

ing averages are symmetrical around the midpoint, and hence Table 32 is sufficient. The simplest method of calculation is to multiply every item of the series by the corresponding weights, a process which is shown in Table 35. The asterisks in this table indicate the items to be summed diagonally up and down from the midpoint to give the smoothed value for the year 1893, which is 4.028. The other smoothed values in the last column are found similarly.

We can eliminate the random element in this way to the degree which we have decided on previously and get a series m' which represents the true mathematical expectation within the limits of the approximation desired. We can assume that it is not far from the course which our prices would have taken if only about one-third of the original random influences had been effective. The result of our smooth-

TABLE 35

SMOOTHING OF THE ANNUAL AMERICAN WHEAT-FLOUR PRICES BY MOVING AVERAGES

$n = 2, m = 3$

(Weights, $g_{2,3}(j)$ from Table 32)

Year	Annual value w	$j = 0$ $g = 0.3333$	$j = 8$ $g = 0.2857$	$j = 2$ $g = 0.1429$	$j = 3$ $g = —0.0952$	Smoothed value m'
1890	5.185	1.728	1.481	0.741	—0.494*	....
1891	5.305	1.768	1.516	0.758*	—0.505	....
1892	4.346	1.449	1.242*	0.621	—0.414	....
1893	4.006	1.335*	1.145	0.572	—0.382	4.028*
1894	3.594	1.198	1.027*	0.513	—0.342	3.604
1895	3.643	1.214	1.041	0.521*	—0.347	3.689
1896	3.795	1.265	1.084	0.542	—0.361*	4.066
1897	4.591	1.530	1.312	0.656	—0.437	4.317
1898	4.729	1.576	1.351	0.676	—0.450	4.348
1899	3.774	1.258	1.078	0.539	—0.359	4.183
1900	3.842	1.281	1.098	0.549	—0.366	3.819
1901	3.810	1.270	1.089	0.544	—0.363	3.651
1902	3.808	1.269	1.088	0.544	—0.363	4.036
1903	4.330	1.443	1.237	0.619	—0.412	4.614
1904	5.378	1.793	1.537	0.768	—0.512	4.907
1905	5.422	1.807	1.549	0.775	—0.516	5.005
1906	4.276	1.425	1.222	0.611	—0.407	4.949
1907	4.875	1.625	1.393	0.699	—0.464	4.956
1908	5.418	1.806	1.548	0.774	—0.516	5.239
1909	5.756	1.919	1.645	0.822	—0.548	5.552
1910	5.495	1.832	1.570	0.785	—0.523	5.557
1911	5.078	1.693	1.451	0.725	—0.484	5.239
1912	5.271	1.757	1.506	0.753	—0.502	4.836
1913	4.544	1.515	1.298	0.649	—0.433	4.939
1914	5.096	1.697	1.456	0.728	—0.485	5.120
1915	6.663	2.221	1.904	0.952	—0.635	6.556
1916	7.264	2.421	2.075	1.038	—0.692	8.191
1917	11.397	3.799	3.256	1.628	—1.086	9.762
1918	10.200	3.400	2.914	1.457	—0.971	11.505
1919	11.998	3.999	3.428	1.714	—1.143	11.965
1920	12.675	4.225	3.621	1.811	—1.207	10.836

TABLE 35 (concluded)

SMOOTHING OF THE ANNUAL AMERICAN WHEAT-FLOUR PRICES BY MOVING AVERAGES

$n = 2,\ m = 3$

(Weights, $g_{2,3}(j)$ from Table 32)

Year	Annual value w	$j=0$ $g=0.3333$	$j=1$ $g=0.2857$	$j=2$ $g=0.1429$	$j=3$ $g=-0.0952$	Smoothed value m'
1921	8.326	2.775	2.379	1.189	—0.793	9.447
1922	7.282	2.427	2.081	1.040	—0.694	7.722
1923	6.385	2.128	1.824	0.912	—0.608	6.704
1924	7.191	2.397	2.055	1.027	—0.685	7.487
1925	8.828	2.943	2.522	1.261	—0.603	7.999
1926	8.426	2.809	2.407	1.204	—0.802	8.257
1927	7.433	2.478	2.124	1.062	—0.708	7.954
1928	7.205	2.402	2.059	1.029	—0.686	7.434
1929	6.786	2.262	1.939	0.969	—0.646	6.442
1930	5.626	1.875	1.607	0.804	—0.536	5.502
1931	4.578	1.526	1.308	0.654	—0.436	4.782
1932	4.196	1.399	1.199	0.599	—0.400	4.726
1933	5.683	1.894	1.624	0.812	—0.541	5.592
1934	6.773	2.258	1.935	0.968	—0.645	6.575
1935	7.676	2.559	2.193	1.097	—0.731	....
1936	6.836	2.279	1.953	0.977	—0.651	....
1937	6.716	2.239	1.919	0.959	—0.640	....

ing process and of the approximations m' are shown in Table 35. We also show in Figure 1 the original annual wheat-flour prices, 1890 to 1937, w, and the approximation to their mathematical expectations found by an application of Sheppard's moving averages, m'.

The same smoothing process has been carried through for the following price series, 1890 to 1937: annual wool prices, annual raw-silk prices, and monthly wool prices.

It appears from our difference analysis as shown in Table 23 that we have probably eliminated the nonrandom element in the annual wool price series in the second or third difference. Assuming $n=2$ and choosing $m=4$, i.e., a moving average comprising 9 items, we smoothed the original annual wool prices w by applying the weights $g_{2,4}(j)$ (Table 32). Both the original data w and the smoothed values m', the approximations to the true mathematical expectation, are shown in Figure 2.

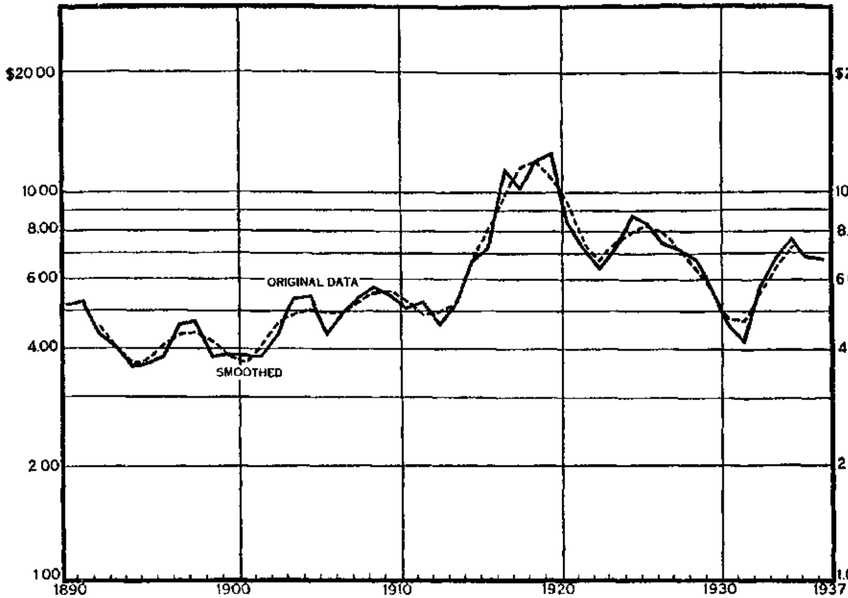

FIGURE 1.—ANNUAL WHEAT-FLOUR PRICES, 1890-1937 (LOGARITHMIC SCALE).

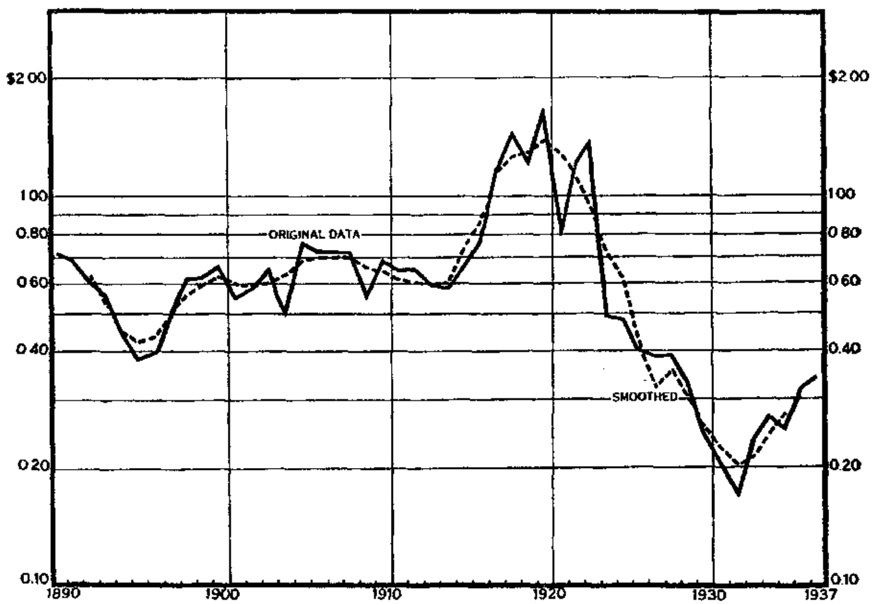

FIGURE 2.—ANNUAL WOOL PRICES, 1890-1937 (LOGARITHMIC SCALE).

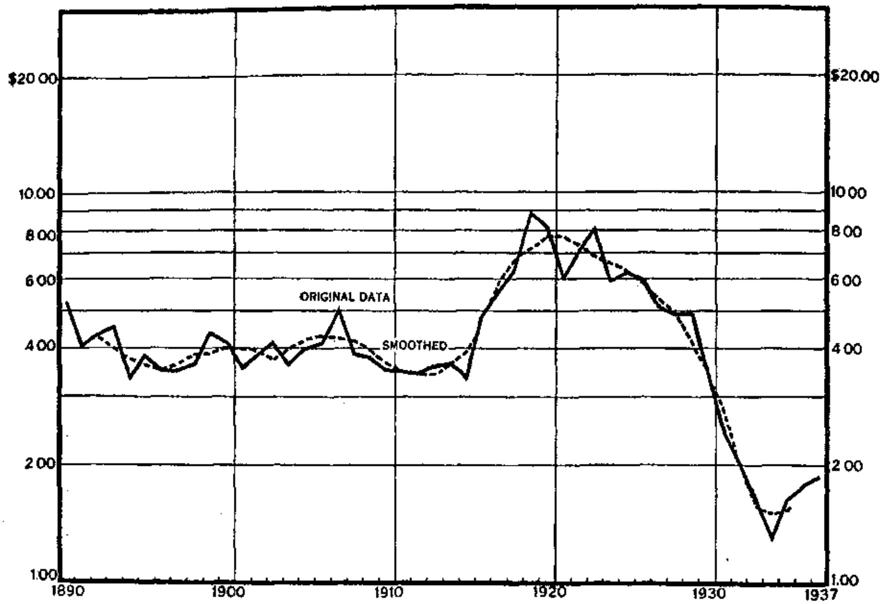

FIGURE 3.—ANNUAL RAW-SILK PRICES, 1890-1937 (LOGARITHMIC SCALE).

Table 25 shows for the annual raw-silk prices a result similar to Table 23. Assuming that we have eliminated the mathematical expectation of this series in the third finite difference, we get $n=2$. If we choose again a moving average of 9 items, we have $m=4$. This means that we take the weights $g_{2,4}(j)$ as given in Table 32. Our original random variance is in this case reduced to about one-fourth (Table 33, L_{nm} for $n=2$, $m=4$). We show in Figure 3 the original series of raw-silk prices w together with the smoothed values m' which are an empirical approximation to the mathematical expectation.

Finally, we see from the difference analysis in Table 24 that we have to go probably as far as the seventh difference in order to eliminate even approximately the nonrandom element or mathematical expectation in the monthly wool prices. Hence we get $n=4$ and choose $m=4$. This means that we reduce our original random variance to about 62% (Table 33).

CHAPTER X

CORRELATION

A. Difference Analysis

The variate difference method has been used very frequently for the purpose of calculating correlations.[1] It seems, indeed, to be very well suited for two purposes—to calculate the correlation between the random elements of two time series and also to get a measure of the linear relationship between the mathematical expectations or the non-random components of the series (see Section C). We treat in this connection only the simplest case. More involved problems, dealing with serial correlation, lag correlation, and multiple correlation, have been discussed in the literature (see Chapter II).

Following some ideas put forward by Tschuprow,[2] we can state our problem in this form: Correlation applies only to stochastic or random variables (Chapter IV, Section B). Nonrandom variables, like our mathematical expectations, can have functional relationships, which are always strictly true (at least in theory) and need not, of course, be linear. But correlation is essentially a stochastic relationship between random variables. This relationship, again, need not be linear. Two random or casual variables are stochastically independent, if the distribution of one is independent of the value of the other. If this is not the case we get certain stochastic relationships one of which may be measured by the correlation coefficient. This measures the closeness of the linear stochastic relationship between the two random variables. But if two random variables are not correlated this does not necessarily mean that there is no stochastic relationship between them. It is possible, for instance, to have a definite relationship between the values of one variable and the variance of the other.

We make an assumption similar to that in Chapter V, but we are

[1] O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, Bonn, 1929, pp. 37 ff., 81 ff., 122 ff.; *Einführung in die mathematische Statistik*, Vienna, 1935, pp. 267 ff.; "Ist die Quantitätstheorie statistisch nachweisbar?" *Zeitschrift für Nationalökonomie*, Vol. 2, 1931, pp. 523 ff.

[2] A. A. Tschuprow, *Grundbegriffe und Grundprobleme der Korrelationstheorie*, Berlin, 1925, especially pp. 12 ff.; O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, pp. 81 ff.

now considering two time series.[3] We assume that each consists of the mathematical expectation and the random element. There is no correlation between the mathematical expectations and the random element of the same series. Only corresponding elements of both series are correlated. That is, we assume that there is a correlation only between the random elements that refer to the same period of time. Our procedure can be very easily generalized for lag correlations.[4] This would assume that the correlation exists only between the random elements, which are lagged by a certain number of units. Our assumption does not involve any considerable restriction. We do not consider here the case in which there is a relationship between one item of the random element of one series and several items of the other series.[5]

We assume further that the population variance of the random element of one time series is σ_x^2 and of the other series σ_y^2. The population value of their product moment is π. The product moments of the differences of two pure random series which satisfy our conditions should be equal if properly calculated and adjusted for the multiplication by the same binomial coefficient that appears in the variances (Chapter V, see Table 9). Hence, if we have eliminated the nonrandom elements or mathematical expectations to a considerable degree in the k_0th difference, we should have the following relationship:[6]

$$p_{k_0} = p_{k_0+1} = p_{k_0+2} = \cdots ,$$

where p_k denotes the empirical approximation to the product moment of our two series calculated from the kth differences. It is an estimate of the true product moment π. This is the relationship that we have to test statistically.

The values p_k are calculated similarly to the V_k which were empirical approximations to the population variances (Chapter V). In fact, we have to divide by the same binomial coefficients (Table 9) or multiply by the corresponding factors A_{kN} (Table 10). The following are the formulae:[7]

$$p_0 = \frac{[S^0(xy) - N\bar{w}(x)\,\bar{w}(y)]}{N-1},$$

[3] See p. 32 ff. and Appendix II, "Mathematical Notes," pp. 146 ff. Appendix I, Section G, gives a summary of computations.

[4] O. Anderson, *op. cit.*, pp. 125 ff.

[5] *Ibid.*, pp. 130 ff.

[6] *Ibid.*, pp. 89 ff., pp. 127 ff.

[7] *Ibid.*, p. 127.

for the original series, and

$$p_k = \frac{S^{(k)}(xy)}{(N-k)\,{}_{2k}C_k} = S^{(k)}(xy)\,A_{kN}\,, \qquad k = 1, 2, \cdots .$$

for the first and higher differences. $\bar{w}(x)$ and $\bar{w}(y)$ in these formulae denote the arithmetic means of the two original series and $S^{(k)}(xy)$ is the sum of the products of their kth differences, the original series being counted as the 0th difference. The values of the coefficients A_{kN} have been tabulated in Table 10.

We want to find out beginning from which k_0th difference the mathematical expectations have been considerably eliminated so that only the random elements are left. Hence the product moments p_k should be approximately equal within the limits of probability. They are empirical approximations to the true population product moment of the random elements of the two series, π. The test can be performed in two ways. We may use the method of the standard errors which is permitted only with large samples. It will give us a criterion for the stability of p_k. Or we may calculate empirical independent correlation coefficients (making a transformation to z as suggested by Fisher) and test their stability statistically. Similar procedures have been used with the variances (Chapters VII and VIII). The latter procedure will be shown in Section B for the problem of correlation.

We first want to describe the method of standard errors proposed by Anderson.[8] The approximate standard error of the difference between the empirical product moments of two consecutive differences $(p_k - p_{k+1})$, $\hat{e}_k{}^0$, depends upon the square of the product moment and the product of the variances of the two series. If we take empirical approximations we get the formula:

$$\hat{e}_k{}^0 = \frac{\sqrt{L_k}}{H_{kN}}\,,$$

where

$$L_k = \frac{p_k{}^2 + V_k(x)\,V_k(y)}{2}$$

p_k has been defined before and is an empirical approximation to the product moment of the random elements of our original series. $V_k(x)$ and $V_k(y)$ are empirical approximations to the variances of the kth difference of the two series. H_{kN} has been tabulated in Table 20.

This approximation is valid only if we assume that we are not

[8] *Ibid.*, pp. 89 ff., 127 ff.

too far from normal correlation and that a term which corresponds to the kurtosis does not appear.[9] We give the correct procedure later.

We can again form an approximative criterion:

$$\hat{R}_k^0 = \frac{(p_k - p_{k+1})}{\sqrt{L_k}} \cdot H_{kN} .$$

A value of $\hat{R}_k^0$ numerically smaller than 3 will give us a fiducial limit which, from the point of view of probability, will be sufficient in most cases for testing the k_0th difference beginning from which we retain only the random elements in our series (see above, pages 33 ff.).

TABLE 36

SUMMARY OF CORRELATION, ANNUAL WOOL (x) AND RAW-SILK (y) PRICES, 1890-1937

Order of Difference k	Sum of Products $S^{(k)}(xy)$	$S^{(k)}(xy^2)$	$S^{(k)}(x^2y)$	$S^{(k)}(x^2y^2)$
0	148.236	838.585	134.868	860.730
1	4.971	....	....	9.110
2	9.575	....	....	49.820
3	26.270	....	....	425.150
4	76.568	....	....	3,713.846
5	230.206	....	....	34,410.202
6	705.936	....	....	335,864.333
7	1,912.628	....	....	3,406,367.622
8	6,869.021	....	....	35,828,491.213
9	21,872.754	....	....	389,309,546.730
10	73,040.159	....	....	4,367,448,992.124

Second-order product moment of original series: $m_{2,2} = 45.87$.

This procedure has been followed with the annual wool and raw-silk prices. They have been compiled from the wholesale price series of the United States Bureau of Labor Statistics. The data cover the period 1890 to 1937. We show in Table 36 the sums $S^{(k)}(xy)$ of the products of the originals and differences of the yearly wool (x) and raw-silk prices (y). The same table also contains other data referred to later.

In Table 37 we find the empirical approximations to the product moments of the original and differences, p_k. The product moment of the original items (0th difference) is calculated according to the formula:

[9] *Ibid.*, p. 127

$$p_0 = \frac{[S^{(0)}(xy) - N\,\overline{w}(x)\,\overline{w}(y)]}{N-1},$$

where $N=48$ is the number of items in our two series, $\overline{w}(x)$ the mean of the wool prices is 0.6218 (Table 11), $\overline{w}(y)$ the mean of the raw-silk prices is 4.2888 (Table 13). We have further from Table 36: $S^{(0)}(xy)$, the sum of the products of the items in the two original series, is 148.236 (Table 36, $k=0$). Hence we get p_0, the approximation to the product moment of the original series, equal to 0.4304 (Table 37).

The same procedure has been adopted for the higher differences, where we use the approximate formula $p_k = S^{(k)}(xy)\,A_{kN}$. The values of the first factor, the sums of the products of the kth differences of the two series, are given in Table 36. The coefficients A_{kN} have been tabulated in Table 10. We want, for instance, to find p_3, the empirical approximation to the product moment of the third differences of the two series. We have from Table 36 under $k=3$: $S^{(k)}(xy) = 26.270$. We get from Table 10 by interpolation for $k=3$, $N=48$: $A_{kN} = 0.001121$. Hence we have for p_3 the value 0.02945, which has again been tabulated in Table 37. Table 37 also gives all the other values of the p_k up to $k=10$.

We have now to calculate the values of $L_k = [p_k^2 + V_k(x)\,V_k(y)]/2$. They are again tabulated in Table 37. Suppose, to give an example,

TABLE 37

DIFFERENCE ANALYSIS OF CORRELATION,
ANNUAL WOOL AND RAW-SILK PRICES, 1890-1937

Order of Difference	Product Moment				Standard Error Ratio	Approximate Standard Error Ratio	Correlation Coefficient
k	p_k	L_k	$\hat{Q}_k$	M_k	$\hat{R}_k$	$\hat{R}^0_k$	r_k
0	0.4304	0.247 167	6.74	0.2665	5.1144	5.1421	0.7743
1	0.05332	0.006 752	12.38	0.05640	2.7601	3.0194	0.5167
2	0.03500	0.004 127	15.62	0.03541	1.3493	1.5644	0.4175
3	0.02945	0.003 474	17.82	0.03219	1.3124	1.5797	0.3777
4	0.02511	0.002 997	20.09	0.02229	1.3396	1.5955	0.3429
5	0.02146	0.002 625	22.31	0.01412	1.3543	1.5626	0.3101
6	0.01835	0.002 328	24.52	0.00771	2.3378	2.5782	0.2792
7	0.01375	0.002 070	26.69	0.00293	0.1291	0.1352	0.2187
8	0.01353	0.001 906	29.17	—0.00114	1.2361	1.2098	0.2246
9	0.01168	0.001 736	31.50	—0.00384	0.8769	0.8046	0.2023
10	0.01052	....	......	....	....	....	0.1888

we want to find L_7. We have from Table 37: $p_7 = 0.01375$; and from Table 23: the variance of the seventh differences of the annual wool prices, $V_7(x) = 0.02577$. Table 25 gives the variance of the seventh differences of the annual raw-silk prices, $V_7(y) = 0.1533$. Hence we establish by a short calculation the value of $L_7 = 0.002\,070$. This value and the values of all the L_k for $k=0, 1, \cdots, 9$ are given in Table 37.

The approximate criterion $\hat{R}_k^0$ which assumes a normal or nearly normal correlation is given by: $\hat{R}_k^0 = (p_k - p_{k+1})H_{kN}/\sqrt{L_k}$. We get the values for p_k and L_k from Table 37. H_{kN} has been tabulated in Table 20. If we want, for instance, to find the approximate criterion $\hat{R}_5^0$ for the difference between the product moments of the fifth and sixth differences, $p_5 - p_6$, we have the following data: From Table 37, $p_5 = 0.02146$, $p_6 = 0.01835$, and $L_5 = 0.002\,625$. Interpolating in Table 20 for $k=5$ and $N=48$, we get the value 25.741 for H_{kN}. A short calculation gives $\hat{R}_5^0 = 1.5626$ as an approximate criterion for the difference between the product moments of the fifth and sixth differences of our yearly series.

The hypothesis that we eliminate the nonrandom element in our product moments of the annual wool and raw-silk prices in the second differences seems to be somewhat justified by the values of the approximate criterion $\hat{R}_k^0$ given in Table 37. Its absolute values become smaller than 3 and stay so beginning with the second difference. We may hence be justified from the point of view of large-sample theory in making this statement. The correlation between the random elements of the two yearly series can hence be estimated by the correlation coefficient of the second differences of both series. We have, in Table 37, $r_2 = 0.4175$. In general, the correlation coefficient of the kth differences is $r_k = p_k/\sqrt{V_k(x)V_k(y)}$. The coefficient of determination[9a] (square of the correlation coefficient) is about 0.17. This means that about 17% of the variance of the random element of one variable can be explained linearly by the variation of the random element of the other.

For the annual series of wool and raw-silk prices and for their differences we have also applied the more rigorous method which involves the knowledge of a term corresponding to the kurtosis in the case of the difference analysis of one single series. It necessitates a knowledge of the sums of the products of the squares of our two series and their differences as well as of the sums of the products of the squares of the one variate times the values of the other variate. The parameters M_k which are necessary for our calculations are shown in

[9a] M. Ezekiel, *Methods of Correlation Analysis*, New York, 1930, p. 120.

Table 37. We construct a better criterion $\hat{R}_k$ according to the formulae:[10]

$$\hat{R}_k = \frac{(p_k - p_{k+1})\hat{Q}_k}{\sqrt{L_k}}$$

and

$$\hat{Q}_k = \frac{H_{kN}}{\sqrt{1 + \frac{J_{kN} M_k}{L_k}}}.$$

H_{kN} and J_{kN} have been tabulated in Tables 20 and 21. The values M_k are given for the differences, $k = 1, 2, \cdots$ by the formula:

$$M_k = \left[\frac{S^{(k)}(x^2y^2)}{N-k} - B_k L_k\right] C_k, \qquad k = 1, 2, \cdots.$$

The coefficients B_k and C_k are given in Tables 14 and 16, N is the number of items in the original series (in our case $N=48$), and L_k has been calculated before and is given in Table 37 for our example. Table 36 shows the values of the $S^{(k)}(x^2y^2)$, i.e., the sums of the products of the squares of the two original series and the difference series.

The calculation of M_0 is somewhat more complicated. We have:

$$M_0 = \left[\frac{m_{2,2}}{N} - E_N L_0\right] F_N.$$

E_N and F_N are coefficients which have been tabulated in Tables 18 and 19. $m_{2,2}$ is the second-order product moment and the best estimate of it is given by:

$$m_{2,2} = S^{(0)}(x^2y^2) - 2\,\bar{w}(y)S^{(0)}(x^2y) - 2\,\bar{w}(x)S^{(0)}(xy^2)$$
$$+ \bar{w}(y)^2 S_2^{(0)}(x) + \bar{w}(x)^2 S_2^{(0)}(y) + 4\bar{w}(x)\bar{w}(y)S^{(0)}(xy)$$
$$- 3N\,\bar{w}(x)^2\bar{w}(y)^2.$$

The sums $S^{(0)}(x^2y^2)$, $S^{(0)}(x^2y)$, $S^{(0)}(xy^2)$, and $S^{(0)}(xy)$ are given in the summary (Table 36, $k=0$). We get for the annual wool prices the sum $S_2^{(0)}(x)$ and the mean $\bar{w}(x)$ from Table 11. Table 13 gives the values for raw-silk prices of $S_2^{(0)}(y)$ and $\bar{w}(y)$. The second-order product moment $m_{2,2}$ is also given in Table 36. For our example, $m_{2,2} = 45.87$, $M_0 = 0.2665$ (Table 37).

The results which appear in Table 37 are really not very different

[10] *Ibid.*, p. 127. See also Appendix II, pp. 147 ff., and Appendix I, Section H for a summary of computations.

from those reached before with the approximate criterion $\hat{R}_k^0$. In fact, we see that probably the mathematical expectations have been eliminated and the random elements common to the annual raw-silk and wool prices have been separated out in the first or second difference. None of the approximate criteria $\hat{R}_k^0$ or the new better criteria $\hat{R}_k$ are greater than 1.6 for a difference higher than $k=1$. Hence we should conclude, at least in this particular case, that it was not really necessary to make the exact test and the approximate test was more or less sufficient. There is some additional labor involved in the more exact tests but it cannot be considered prohibitive.

B. Selected Comparisons[11]

We can also try to find the k_0th difference, beginning from which the differences of our two series contain approximately only the random elements, in another way. We start with the same assumptions as before but have also to assume that we are dealing with normal or nearly normal correlation between the random elements. We calculate the correlation coefficient r_k' for the correlation between selected items for the kth differences of our two series and another independent correlation coefficient r'_{k+1} for the correlation between selected items of the $(k+1)$th differences of our two series. If we have eliminated to a considerable degree the nonrandom elements in our k_0th difference, then the correlation coefficients r'_{k_0} and r'_{k_0+1} and coefficients of selected independent higher differences should be equal within the limits of probability.

The selections are again made by use of Table 26, which gives all the possible arrangements. The selections are arranged in such a way that the differences of order k and $k+1$ which they contain are independent and hence also the correlation coefficients (Table 38).

We use Fisher's transformation[12] for z_k' which is now also given in the tables by Fisher and Yates.[13] The difference $z'_k - z'_{k+1}$ should be distributed normally with a standard error of

$$\sqrt{\frac{1}{n-3}+\frac{1}{n-1}}$$

for $k=0$, that is, a comparison of the correlation between the original

[11] G. Tintner, "On Tests of Significance in Time Series," *Annals of Mathematical Statistics*, Vol. 10, 1939, pp. 139 ff. See also Appendix II, pp. 148 ff., and Appendix I, Section J, for a summary of computations.

[12] R. A. Fisher, *Statistical Methods for Research Workers*, 7th ed., London, 1938, pp. 202 ff., Chapter VI, Section 35.

[13] R. A. Fisher and F. Yates, *Statistical Tables*, London, 1938, p. 36.

TABLE 38

SELECTED COMPARISONS OF THE CORRELATION OF ANNUAL WOOL AND RAW-SILK PRICES, 1890-1937

Selection	Number of Items in Selection	Correlation Coefficients		Transformations		Ratio of the Difference $z'_k - z'_{k+1}$ to its Standard Error
		Lower Difference	Higher Difference	Lower Difference	Higher Difference	
	n	r'_k	r'_{k+1}	z'_k	z'_{k+1}	
0-A	16	0.8270	0.0973	1.179	0.097	2.85
0-B	15	0.6187	0.1609	0.723	0.172	1.40
0-C	15	0.8136	0.8545	1.138	1.272	—0.34
1-A	9	0.7848	0.4769	1.058	0.517	1.08
1-B	9	0.6798	0.0526	0.829	0.053	1.56
1-C	9	0.5536	0.5706	0.624	0.648	—0.05
1-D	9	0.4371	0.8480	0.469	1.250	—1.57
1-E	8	—0.0921	0.0484	—0.092	0.048	—0.26
2-A	6	—0.8361	0.2173	—1.208	0.221	—2.26
2-B	6	0.7624	0.8607	1.002	1.295	—0.46
2-C	6	0.8075	0.6789	1.120	0.827	0.46
2-D	6	0.1290	—0.7285	0.130	—0.925	1.67
2-E	6	0.7581	—0.5689	0.992	—0.645	2.58
2-F	6	0.9219	0.8503	1.601	1.258	0.54
2-G	6	—0.3482	0.5554	—0.363	0.626	—1.57
3-A	5	—0.8225	0.9150	—1.165	1.557	—3.84
3-B	5	—0.5143	0.8105	—0.569	1.129	—0.79
3-C	5	0.8392	0.0224	1.218	0.022	1.69
3-D	5	0.8344	—0.9778	1.202	—2.240	—4.86
3-E	4	0.8322	—0.8907	1.196	—1.425	3.21
3-F	4	0.9192	—0.8803	1.583	—1.377	3.62
3-G	4	0.5112	0.6486	0.565	0.773	—0.25
3-H	4	—0.6424	0.9743	—0.762	2.170	—3.59
3-I	4	—0.9248	0.8815	—1.620	1.382	—3.68
4-A	4	0.7991	—0.5583	1.097	—0.630	2.12
4-B	4	0.2717	0.8346	0.279	1.203	—1.13
4-C	4	0.0140	0.8951	0.014	1.446	—1.88
4-D	4	0.4869	0.9119	0.532	1.537	—1.22
4-E	4	—0.7064	0.9716	—0.880	2.120	—3.66
4-F	3	—0.9114	0.9734	—1.535	2.151	—3.69
4-G	3	0.6426	0.4774	0.762	0.520	0.24
4-H	3	0.9166	0.0903	1.564	0.090	1.47
4-I	3	0.6036	0.5925	0.699	0.681	0.02
4-J	3	0.9858	—0.4537	2.461	—0.489	2.95
4-K	3	0.9962	—0.8123	3.130	—1.133	4.26

TABLE 38 (concluded)

SELECTED COMPARISONS OF THE CORRELATION OF ANNUAL WOOL AND RAW-SILK PRICES, 1890-1937

Selection	Number of Items in Selection	Correlation Coefficients		Transformations		Ratio of the Difference $z'_k - z'_{k+1}$ to its Standard Error
		Lower Difference	Higher Difference	Lower Difference	Higher Difference	
	n	r'_k	r'_{k+1}	z'_k	z'_{k+1}	
5-A	3	—0.7238	0.8101	—0.917	1.127	—2.04
5-B	3	—0.1727	0.8068	—0.174	1.119	—0.95
5-C	3	0.6087	0.0604	0.707	0.060	0.65
5-D	3	0.5396	—0.8283	0.604	—1.182	—1.79
5-E	3	0.4908	0.0932	0.537	0.093	0.44
5-F	3	0.9475	0.7151	1.807	0.898	0.91
5-G	3	0.8807	—0.5391	1.379	—0.603	1.98
5 H	3	0.5275	—0.6360	0.587	—0.751	1.34
5-I	3	0.5776	0.3803	0.659	0.400	1.06
5-J	3	—0.7098	0.6470	—0.887	0.770	1.66
5-K	2	—0.4296	—0.6968	—0.460	—0.861	—0.28
5-L	2	0.6775	0.8794	0.825	1.373	—0.39
5-M	2	—0.0286	0.9735	—0.029	2.156	—1.54
6-A	3	0.6256	—0.4726	0.733	—0.514	1.25
6-B	3	0.7295	—0.4085	0.928	—0.434	1.36
6-C	3	0.5935	—0.6824	0.683	—0.834	1.52
6-D	3	—0.0714	—0.8190	—0.071	—1.154	—1.08
6-E	2	—0.6302	—0.0667	—0.742	—0.067	—0.48
6-F	2	—0.8822	0.7592	—1.383	0.994	—1.68
6-G	2	—0.7514	0.5810	—0.975	0.664	—1.16
6-H	2	—0.5679	0.0277	—0.643	0.028	0.47
6-I	2	—0.5249	0.7362	—0.583	0.942	—1.08
6-J	2	—0.7760	0.7930	—1.035	1.080	—1.49
6-K	2	—0.6714	0.6198	—0.814	0.725	1.09
6-L	2	—0.6264	—0.8921	—0.736	—1.432	—0.49
6-M	2	0.8174	—0.2216	1.149	—0.225	0.97
6-N	2	0.8785	—0.9853	1.369	—2.453	2.69
6-O	2	0.1404	—0.7115	0.141	—0.890	0.73

series and the first difference (selections 0-A, 0-B, and 0-C). The standard error is equal to $\sqrt{2/(n-1)}$ for all other k's, i.e., for selections of order one or higher. The n is the number of items in the selections in question.[14] This follows from the fact that we need not calculate the arithmetic means of the differences whose true means are zero and hence lose only one degree of freedom here, whereas we have

[14] R. A. Fisher, *op. cit.*, pp. 208 ff.

to calculate the arithmetic mean of the original series and lose there three degrees of freedom.

We have applied this method to the coefficients of correlation of selected comparisons of the annual wool and raw-silk prices. The results are more or less encouraging and show that the mathematical expectation has probably been eliminated in the first or second differences—the same result which we reached before. They are presented in Table 38.

To give an example: For selections 0-A we take items number 1, 4, 7, 10 etc. of the original series and items number 2, 5, 8, 11, etc. of the first differences (Table 26). This gives us a correlation coefficient of $r_k' = 0.8270$ for the original series and the corresponding z_k' from Fisher's table is equal to 1.179 (Table 38). The correlation coefficient between the selected items of the first differences is $r'_{k+1} = 0.0973$ and the corresponding z'_{k+1} from Fisher's table is 0.097 (Table 38). The difference is $z_k - z_{k+1} = 1.082$ and its standard error is $\sqrt{1/13 + 1/15} = 0.379$ since we have $n=16$ and the order of selection is zero. The difference is about 2.85 times the standard error.

Take for instance selection 1-A, that is, items 1, 6, 11, 16, etc. of the first differences and items 3, 8, 13, 18, etc. of the second differences (Table 26). We get a correlation coefficient of $r_k' = 0.7848$ for the first differences and $r'_{k+1} = 0.4769$ for the second differences. The corresponding z'_k and z'_{k+1} are 1.058 and 0.517 (Table 38). Their difference is 0.541 and the standard error is $\sqrt{1/4} = 0.5$ since $n=9$. Hence the difference is equal to only about once its standard error and hence probably not significant.

We show in Table 38 the ratios of the differences of the z's, $z_k' - z'_{k+1}$, to their standard errors. Here again as in Chapter VIII we should not hope for a too close agreement with the theory, since we have probably not eliminated *all* the nonrandom elements in the second differences of our series. Our samples also become very small with higher selections. Table 38 seems however to support the hypothesis, to a certain degree, that we eliminate the nonrandom parts in the second differences. The exceptions are not too numerous. The difference of the z's exceeds three times its own standard error only in rare cases.

C. The Linear Relationship between the Mathematical Expectations

We can now try to get an estimate of the linear relationship between the nonrandom elements in the wool and raw-silk prices.

Tschuprow,[15] Anderson,[16] and recently Wold[17] have shown that the empirical correlation coefficient between two stochastic or random variables consists really of elements of three different correlations: the correlation between the random elements, the correlation between the arithmetic means of the random elements, and the correlation (i.e., a measure of linear relationship) between the mathematical expectations. The second terms disappear under our assumptions (and also with Wold) and we are left with only elements pertaining to the correlation between the random elements and to the linear relationship between the mathematical expectations.[18]

Denoting the empirical approximation to the product moment of the mathematical expectation by $p(m)$ and the empirical approximation to the product moment of the random parts by p_{k_0}, we have the following relationship:

$$p_0 = \frac{N-1}{N} p_{k_0} + p(m).$$

We get also a similar relationship for the variances of both price series.

If we carry the calculations through we get the following results for the linear relationship between the mathematical expectations of the annual wool prices and raw-silk prices: We have from Table 37 the values for the product moments: $p_0 = 0.4304$ and (since $k_0 = 2$, the random elements being more or less isolated in the second differences) $p_{k_0} = p_2 = 0.03500$. Hence we get in our case for an estimate of the product moment of the mathematical expectation from $p(m) = p_0 - [(N-1)p_{k_0}/N]$ with $N = 48$, $p(m) = 0.3961$. We take the second differences as the ones in which we have eliminated the nonrandom element in so far as it enters into the correlation of the yearly wool and raw-silk prices. By reasoning similar to that before we have from Table 23 for the wool prices (x): $V_0(x) = 0.1069$ and the approximation to the random variance: $V_2(x) = 0.02590$. We have the formula $V[m(x)] = V_0(x) - [(N-1)V_{k_0}(x)/N]$. The approximation to the variance of the mathematical expectations of the wool prices is 0.08154. Table 25 gives the data for the approximation

[15] A. A. Tschuprow, *op. cit.*

[16] O. Anderson, *Die Korrelationsrechnung in der Konjunkturforschung*, p. 124, formula 79.

[17] H. Wold, *A Study in the Analysis of Stationary Time Series*, Uppsala, 1938, p. 200.

[18] See also Appendix II, pp. 148 ff., and Appendix I, Section J, for a summary of computations.

to the variance of the mathematical expectations of the raw-silk price $V[m(y)]$ which is 2.6257. Hence the correlation which measures the linear relationship between the mathematical expectations of the annual wool and raw-silk prices is

$$r(m) = p(m)/\sqrt{V[m(x)][V\,m(y)]} = 0.8560.$$

We want to stress again the fact that the correlation coefficient between mathematical expectations has not the same meaning as a correlation coefficient for random variables. It cannot be interpreted stochastically, i.e., from the point of view of probability, and measures only the average linear relationship of the two mathematical expectations over the period considered.

APPENDIXES

APPENDIX I

SUMMARY OF COMPUTATIONS

A. Calculation of the Variances of Differences

Calculate the differences of the series w_i: $\Delta^{(k)}w_i$. Find the sum of the original series $S_1{}^{(0)} = \sum_{i=1}^{N} w_i$. Square the original items and the differences and find the sum of the squares of the original series and differences $S_2{}^{(k)} = \sum_{i=1}^{N-k} (\Delta^{(k)}w_i)^2$, where the 0th difference $\Delta^{(0)}w_i = w_i$ is the original series. Form the mean of the series $w = S_1{}^{(0)}/N$ where N is the number of items in the original series. The best estimate of the variance of the original series is $V_0 = (S_2{}^{(0)} - N\,\bar{w}^2)/(N-1)$. The estimates of the variances of the kth differences are: $V_k = S_2{}^{(k)}A_{kN}$. The coefficient A_{kN} is given in Table 10.

B. Calculation of the Variances Corrected for Seasonal

Find the means of all the Januaries, Februaries, etc.: w_i^*. Differencing this series and proceeding as before we have $S_2{}^{*(k)} = \sum_{i=1}^{12} (\Delta^{(k)}w_i^*)^2$. The corrected variance of the original items is: $V_0^* = [S_2{}^{(0)} - N\bar{w}^2 - (N/12)(S_2{}^{*(0)} - 12\,\bar{w}^2)]/(N-1)$. The corrected values for the variances of the differences are: $V_k^* = [S_2{}^{(k)} - (N/12)\ S_2{}^{*(k)}]\ A_{kN}$. The coefficient A_{kN} is given in Table 10.

C. Difference Analysis: Approximate Criterion

Form the variances V_k or the corrected variances V_k^* as described in Sections A and B. The approximate criterion of the stability of the uncorrected variances is: $R_k{}^0 = (V_k - V_{k+1})H_{kN}/V_k$; and for the variances corrected for seasonal: $R_k{}^{*0} = (V_k^* - V_{k+1}^*)H_{kN}/V_k^*$. The coefficient H_{kN} is given in Table 20. An absolute value of $R_k{}^0$ or $R_k{}^{*0}$ smaller than 3 indicates stability of the variances.

D. Difference Analysis: Exact Criterion

Find the values of $S_1{}^{(0)}$, $S_2{}^{(k)}$ as in Section A. Calculate the sum of the cubes of the original items $S_3{}^{(0)} = \sum_{i=1}^{N} w_i^3$ and the sums of the fourth powers of the original series and the differences: $S_4{}^{(k)} = \sum_{i=1}^{N-k} (\Delta^{(k)}w_i)^4$ (here again, the original series is the 0th difference). Find the fourth moment about the mean of the original series: $m_4 = (S_4{}^{(0)}/N) - (4S_3{}^{(0)}\bar{w}/N) + (6S_2{}^{(0)}\bar{w}^2/N) - 3\bar{w}^4$. The best estimate of the kurtosis of the original series is $D_0 = (m_4 - E_N V_0{}^2)F_N$. E_N and F_N are given in Tables 18 and 19. The best estimates of the kurtosis of the differences are: $D_k = \{[S_4{}^{(k)}/(N-k)] - B_k V_k{}^2\}C_k$; B_k and C_k are tabulated in Tables 14 and 16. Form the values $G_k = D_k/V_k{}^2$ for the original series and differences, and the values $Q_k = H_{kN}/\sqrt{(1 + J_{kN}\ G_k)}$, again for the original series and differences. The coefficients H_{kN} and J_{kN} are given in Tables 20 and

21. The exact criterion for the stability of the variances of the difference series is then: $R_k = (V_k - V_{k+1}) Q_k / V_k$. A numerical value of R_k smaller than 3 indicates stability of the variances.

E. Difference Analysis: Tests of Significance

Make selected comparisons of the kth and $(k+1)$th differences according to Table 26. Find the sum of squares of the selected differences: $\bar{S}_2^{(k)} = \Sigma' (\Delta^{(k)} w_i)^2$ and $\bar{S}_2^{(k+1)} = \Sigma'' (\Delta^{(k+1)} w_i)^2$ where Σ' and Σ'' denote summation over the selections given in Table 26. Form the ratio $\bar{S}_2^{(k)} / \bar{S}_2^{(k+1)}$. If the variances of the kth and $(k+1)$th differences are equal, then this ratio should fall within the limits given in Tables 27, 28, and 29. These limits are calculated from the point of view of a 5%, 1%, and 0.1% level of significance.

F. Smoothing by Sheppard's Formulae

Let k_0 be the difference at which the variances become stable. Put $n = k_0/2$ for k_0 even and $n = (k_0+1)/2$ for k_0 odd. V_{k_0} is the best estimate of the true random variance. Table 32 gives the weights of the smoothing formulae $g_{nm}(j)$ where j is the distance from the midpoint. The degree of accuracy m can be determined by taking into consideration that the variance of the smoothed series m_i' is $V(m') = L_{nm} V_{k_0}$ (Table 33). Table 34 shows the reduction of the coefficient of random variability $v = \sqrt{V_{k_0}} / \bar{w}$, if the weights given in Table 32 are applied to the original series w_i.

G. Correlation: Approximate Criterion

Given two series x and y, find the sums of the original items $S_1^{(0)}(x)$ and $S_1^{(0)}(y)$ and their means $\bar{w}(x) = S_1^{(0)}(x)/N$ and $\bar{w}(y) = S_1^{(0)}(y)/N$. Find the sums of the squares of the original and differences, $S_2^{(k)}(x)$ and $S_2^{(k)}(y)$. Calculate the estimates of the variances of the originals and differences of both series, $V_k(x)$ and $V_k(y)$, as in Section A. Find also the sums of the products of the originals and the differences of both series, $S^{(k)}(xy)$. The best estimate of the product moment of the original two series is: $p_0 = [S^{(0)}(xy) - N\bar{w}(x)\,\bar{w}(y)]/(N-1)$, and the best estimate for the product moments of the differences is: $p_k = S^{(k)}(xy)\,A_{kN}$. The coefficient A_{kN} is given in Table 10. Form also $L_k = [p_k^2 + V_k(x)\,V_k(y)]/2$ for the original and differences. The approximate criterion for the stability of the product moments p_k is then given by: $\hat{R}_k^0 = (p_k - p_{k+1}) H_{kN} / \sqrt{L_k}$ (H_{kN} is tabulated in Table 20). A numerical value smaller than 3 indicates stability.

H. Correlation: Exact Criterion

Find the values of $S_1^{(0)}(x)$, $S_1^{(0)}(y)$, $S_2^{(0)}(x)$, $S_2^{(0)}(y)$, $S^{(k)}(xy)$ as before (Section G) and also $S^{(0)}(x^2y)$ and $S^{(0)}(xy^2)$, the sums of the products of the values of one original series times the squares of the other. Calculate $S^{(k)}(x^2y^2)$, the sums of the products of the squares of the original series or difference series times the squares of the other. Find the second-order product moment of the original series $m_{2,2} = S^{(0)}(x^2y^2) - 2\bar{w}(y)S^{(0)}(x^2y) - 2\bar{w}(x)S^{(0)}(xy^2) + \bar{w}(y)^2 S_2^{(0)}(x) + \bar{w}(x)^2 S_2^{(0)}(y) + 4\bar{w}(x)\bar{w}(y)S^{(0)}(xy) - 3N\bar{w}(x)^2\bar{w}(y)^2$. Calculate the parameters $M_0 = [(m_{2,2}/N) - E_N L_0]F_N$, where L_0 is calculated as above (Section G) and E_N and F_N are given in Tables 18 and 19. For the differences: $M_k = [S^{(k)}(x^2y^2)/(N-k) - B_kL_k]C_k$, where L_k is calculated as before and B_k and C_k are given in Tables 14 and 16. Finally, form for the original series and the differences the values: $\hat{Q}_k = H_{kN}/\sqrt{1 + (J_{kN}M_k/L_k)}$, where H_{kN} and J_{kN} are given in Tables 20 and 21. The exact criterion is then $\hat{R}_k = (p_k - p_{k+1})\hat{Q}_k/\sqrt{L_k}$. Again, a numerical value smaller than 3 indicates stability.

I. Correlation: Tests of Significance

Use again the selected comparisons as given in Table 26. Find the variances and product moments for the kth and $(k+1)$th differences, using selected items; p'_k, p'_{k+1}, $V'_k(x)$, $V'_k(y)$, $V'_{k+1}(x)$, $V'_{k+1}(y)$. Form the correlation coefficients for the kth and $(k+1)$th differences based upon the selected items: $r'_k = p'_k/\sqrt{V'_k(x)V'_k(y)}$ and $r'_{k+1} = p'_{k+1}/\sqrt{V'_{k+1}(x)\ V'_{k+1}(y)}$. Find the transformed values z'_k and z'_{k+1} from Fisher's table (R. A. Fisher, *Statistical Methods for Research Workers*, 7th ed., London, 1938, Table V-B; R. A. Fisher and F. Yates, *Statistical Tables*, London, 1938, Table VII). Let n be the number of items in the selection. Then the standard error of the difference $z'_k - z'_{k+1}$ is $\sqrt{[1/(n-3)] + [1/(n-1)]}$ for $k=0$, i.e., selected comparisons of order zero. The standard error is $\sqrt{2/(n-1)}$ for higher selections. A difference smaller than three times its standard error indicates stability.

J. Correlation of the Random Elements and Linear Relationship of the Mathematical Expectations

Let k_0 be the difference at which the product moments p_k or the correlation coefficients of the selected comparisons r'_k become stable. Compute the variance of the mathematical expectation of x: $V[m(x)] = V_0(x) - [(N-1)\ V_{k_0}(x)/N]$; the variance of the mathematical expectation of y: $V[m(y)] = V_0(y) - [(N-1)\ V_{k_0}(y)/N]$; and the product moment of the mathematical expectations: $p(m) = p_0 - [(N-1)p_{k_0}/N]$. The correlation between the random elements in x and y gives the correlation coefficient: $r_{k_0} = p_{k_0}/\sqrt{V_{k_0}(x)\ V_{k_0}(y)}$. The linear relationship between the mathematical expectations of x and y is measured by: $r(m) = p(m)/\sqrt{V[m(x)]\ V[m(y)]}$.

APPENDIX II

Mathematical Notes

The following notes will give the mathematical statistician some of the necessary background for the ideas involved in the procedures of the variate difference method. The reader is referred, however, to the original work of Anderson[1] and Zaycoff[2] for a more detailed treatment.

Notes to Chapter IV, Section A

Let N be the total number of cases observed and N_1 the number of them which show a certain characteristic. The frequency definition of the probability[3] p is then:

$$p = \lim_{N \to \infty} \frac{N_1}{N}. \tag{1}$$

We take a sample from our original series in a way which is independent of the characteristic, but arbitrary otherwise. We have N' cases in our sample and N_1' of them show the characteristic. We should have in a true collective again:

$$p' = \lim_{N' \to \infty} \frac{N_1'}{N'} = p\,, \tag{2}$$

and similarly for any further samples N'', N''', etc., we should always get $p = p' = p'' = p''' = \cdots$.

Notes to Chapter IV, Section B

A random or casual or stochastic variable[4] x can assume a number (say M) of values $x_1, \cdots, x_M$ with certain definite probabilities $p_1, \cdots, p_M$, where $\sum_{i=1}^{M} p_i = 1$. The values of x and their probabilities are the distribution of x. The distribution law is continuous if x varies continuously.

The mathematical expectation[5] of x, $E(x)$, is defined as:

$$E(x) = \sum_{i=1}^{M} x_i\, p_i\,. \tag{3}$$

[1] O. Anderson, **Die Korrelationsrechnung in der Konjunkturforschung,** Bonn, 1929, especially pp. 101 ff. See also G. Tintner, **Prices in the Trade Cycle,** Vienna, 1935, pp. 81 ff.

[2] R. Zaycoff, "Ueber die Ausschaltung der zufälligen Komponente nach der 'Variate-Difference' Methode," **Publications of the Statistical Institute for Economic Research,** State University of Sofia, 1937, No. 1, pp. 75 ff.

[3] R. von Mises, **Probability, Statistics and Truth,** London, 1939; **Wahrscheinlichkeitsrechnung,** Vienna, 1931, pp. 7 ff. See also, R. von Mises and H. Pollaczek-Geiringer, article "Probability" in **Encyclopaedia of the Social Sciences,** New York, 1937, Vol. 12, pp. 426 ff.; A. Wald, "Die Widerspruchsfreiheit des Kollektivbegriffs der Wahrscheinlichkeitsrechnung," **Ergebnisse eines mathematischen Kolloquiums,** No. 8, Vienna, 1937, pp. 38 ff.; S. S. Wilks, **Statistical Inference,** Princeton, 1937, pp. 2, ff.

[4] J. V. Uspensky, **Introduction to Mathematical Probability,** New York, 1937, pp. 161 ff.

[5] **Ibid.,** pp. 163 ff.

We have $E(ax + by) = aE(x) + bE(y)$, for any constants a, b, and any dependent or independent random variables x, y. The relation $E(xy) = E(x)E(y)$ holds true only for uncorrelated random variables.

The mathematical expectations of the powers of x are the moments[6] about zero, μ'_a:

$$(4) \qquad E(x^a) = \sum_{i=1}^{M} x_i^a \, p_i = \mu'_a ,$$

and the mathematical expectations of the powers of the deviations from the mathematical expectation of x, $E(x)$, are the moments about the mathematical expectation, μ_a:

$$(5) \qquad E[x - E(x)]^a = \sum_{i=1}^{M} [x - E(x)]^a \, p_i = \mu_a .$$

The second moment about the mathematical expectation, $E[x - E(x)]^2 = \sigma^2$, is the population variance, its square root the population standard deviation σ.

We denote throughout population values by Greek symbols and assume a hypothetical infinite population, of which our observations are a sample. These assumptions are justified by the economic considerations given in Chapter I, Section C.

The following formulae are useful:

$$(6) \qquad \mu'_0 = \mu_0 = 1 ,$$

$$(7) \qquad \mu'_1 = E(x) ,$$

$$(8) \qquad \mu_1 = 0 ,$$

$$(9) \qquad \mu_2 = \sigma^2 = \mu_2' - \mu_1'^2 ,$$

$$(10) \qquad \mu_3 = \mu_3' - 3\,\mu_2'\,\mu_1' + 2\,\mu_1'^3 ,$$

$$(11) \qquad \mu_4 = \mu_4' - 4\,\mu_3'\,\mu_1' + 6\,\mu_2'\,\mu_1'^2 - 3\,\mu_1'^4 .$$

Notes to Chapter IV, Section C

Given a time series w_i $(i = 1, 2, \cdots, N)$ we define the first difference of w_i as:

$$(12) \qquad \Delta^{(1)} w_i = w_{i+1} - w_i .$$

The second difference is the difference of the first difference:

$$(13) \qquad \Delta^{(2)} w_i = \Delta^{(1)} w_{i+1} - \Delta^{(1)} w_i = w_{i+2} - 2w_{i+1} + w_i .$$

and generally the kth difference is defined as:

$$(14) \qquad \Delta^{(k)} w_i = {}_kC_0 \cdot w_{i+k} - {}_kC_1 \cdot w_{i+k-1} + {}_kC_2 \cdot w_{i+k-2} - \cdots + (-1)^k \cdot {}_kC_k \cdot w_i ;$$

where ${}_kC_\alpha$ is the number of combination of k things taken α at a time, a binomial coefficient:

$$(15) \qquad {}_kC_\alpha = \frac{k(k-1)(k-2)\cdots(k-\alpha+1)}{1 \cdot 2 \cdot 3 \cdots \alpha} .$$

6 S. S. Wilks, op. cit., pp. 7 ff.

Anderson[7] gives the special form of the finite difference of polynomials, exponential functions, hyperbolas, and trigonometric functions (see also above, Chapter IV, Section C).

Notes to Chapters V, VI, and VII

Given an economic time series w_i, $i = 1, \cdots, N$, we assume that it is composed of two parts:[8] a random element x_i and the mathematical expectation or smooth part $m_i = E(w_i)$, which can be eliminated by differencing. We assume that the connection is linear:

$$w_i = m_i + x_i \,. \tag{16}$$

The two parts are not correlated:

$$E(m_i x_j) = 0\,, \qquad i, j = 1, 2, \ldots, N\,. \tag{17}$$

The elements of the random part are not correlated with each other:

$$E(x_i x_j) = 0\,, \qquad i, j = 1, 2, \ldots, N\,, i \neq j\,. \tag{18}$$

The mathematical expectation of x is zero and its population variance σ^2:

$$E(x_i) = 0\,, \qquad i = 1, 2, \cdots, N\,, \tag{19}$$

$$E(x_i^2) = \sigma^2 = \mu_2^{(0)}\,, \qquad i = 1, 2, \cdots, N\,. \tag{20}$$

We get for the kth difference of x:

$$E(\Delta^{(k)} x_i) = 0\,, \tag{21}$$

because of (19) and since the differences are linear forms in the x_i's in (14).

The expectation of the squares of the differences is:

$$\begin{aligned} E(\Delta^{(k)} x_i)^2 = \mu_2^{(k)} &= E({}_kC_0 \cdot x_{i+k} - {}_kC_1 \cdot x_{i+k-1} + \cdots)^2 \\ &= E({}_kC_0^2 \cdot x^2_{i+k} + {}_kC_1^2 \cdot x^2_{i+k-1} + \cdots - 2 \cdot {}_kC_0 \cdot {}_kC_1 \cdot x_{i+k}\, x_{i+k-1} + \cdots) \\ &= {}_{2k}C_k \,.\, E(x_i^2) = {}_{2k}C_k \,.\, \mu_2^{(0)} = {}_{2k}C_k \cdot \sigma_2\,, \end{aligned} \tag{22}$$

because of (18) and since

$${}_kC_k^2 + {}_kC_{k-1}^2 + \cdots + {}_kC_1^2 + {}_kC_0^2 = {}_{2k}C_k\,. \tag{23}$$

If we define the population variance of the kth difference of the random element as $\sigma_k^2 = \mu_2^{(k)}/{}_{2k}C_k$, we have the following relations (0th difference is the original series x_i):

$$\sigma_0^2 = \sigma_1^2 = \sigma_2^2 = \cdots = \sigma^2\,. \tag{24}$$

If we have eliminated the smooth nonrandom element or the mathematical expectation m in the k_0th difference of w, we should have approximately:

$$V_{k_0} = V_{k_0+1} = V_{k_0+2} = \cdots\,, \tag{25}$$

where V_k is an empirical approximation to σ_k^2.

[7] O. Anderson, op. cit., pp. 108 ff.
[8] Ibid., pp. 110 ff.

In the k_0th difference and higher differences of w_i only the random element x_i remains and hence relations (25) hold true. This is the relationship which we have to test statistically.

For large samples, if the standard errors can be calculated with sufficient accuracy, we have two important tests of significance. (An exact test for small samples will be given later.) First is Tchebycheff's inequality[9] which holds true for any distribution. If t is a positive constant and if x is a random variable, with variance σ^2, then the probability P of the inequality $|x—E(x)| \leq t\sigma$, is:

$$P \geq 1 - \frac{1}{t^2}. \tag{26}$$

If x is a random variable, which is normally distributed, such that its probability law is:

$$p(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{[x-E(x)]^2}{2\sigma^2}\right)dx, \tag{27}$$

then the probability of a deviation $|t| = |[x—E(x)]/\sigma| \leq t_0$ is given by the integral:

$$P = \int_{-t_0}^{t_0} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt, \tag{28}$$

which has been frequently tabulated. (It should be noted that these tests necessitate a knowledge of the true population variance σ^2, which can be estimated with some accuracy only in large samples.)

In order to test differences between the variances of consecutive series of differences, we have to find the standard error of the difference $V_k - V_{k+1}$, whose square is

$$e_k^2 = E(V_k - V_{k+1})^2, \tag{29}$$

where V_k is the variance of the kth sample difference, an empirical approximation to the true value σ_k^2.

We have:

$$\bar{w} = \frac{\sum_{i=1}^{N} w_i}{N} \tag{30}$$

as the arithmetic mean of the original series w, and

$$S_a^{(k)} = \sum_{i=1}^{N-k} (\Delta^{(k)} w_i)^a, \qquad k = 0, 1, 2, \cdots, \tag{31}$$

the sums of the kth differences raised to the αth power. The original series is the 0th difference.

It follows from formulae given by Tschuprow, Anderson, Fisher, and Zaycoff[10] that an unbiased estimate of the population variance σ_0^2 of the original

[9] S. S. Wilks, op. cit., p. 8; J. V. Uspensky, op. cit., p. 204.

[10] A. A. Tschuproff, "On the Mathematical Expectation of the Moments of Frequency Distributions," Biometrika, Vol. 12, 1919, pp. 140 ff.; O. Anderson, op. cit., pp. 103 ff.; Mathematische Statistik, Vienna, 1935, pp. 213 ff.; R. A. Fisher, Statistical Methods for Research Workers, 7th ed., London, 1938, pp. 77 ff., Chapter III, appendix; R. Zaycoff, loc. cit., p. 78.

series is:

(32) $$V_0 = \frac{\sum_{i=1}^{N} (w_i - \bar{w})^2}{N-1} = \frac{S_2^{(0)} - N\bar{w}^2}{N-1}.$$

The unbiased estimate of the variances of the higher differences, σ^2_k, follows from (22) if we take into account that the true mean or mathematical expectation of x is zero (Formula 19):

(33) $$V_k = \frac{\sum_{i=1}^{N-k} (\Delta^{(k)} w_i)^2}{(N-k)\,{}_{2k}C_k} = S_2^{(k)} A_{kN}.$$

Similarly we get the estimates for the kurtosis of the original series, $\mu_4^{(0)} - 3(\sigma_0^2)^2$:

(34) $$D_0 = \frac{1}{1 - \frac{4}{N} + \frac{6}{N^2} - \frac{3}{N^3}} \left\{ \frac{1}{N} \sum_{i=1}^{N} (w_i - \bar{w})^4 - 3\left[\frac{(N-1)}{N}\right]^2 V_0^2 \right\}$$

$$= (m_4 - E_N V_0^2) F_N;$$

$$m_4 = (S_4^{(0)}/N) - (4S_3^{(0)}\bar{w}/N) + (6S_2^{(0)}\bar{w}^2/N) - 3w^4.$$

For the series of differences, $\mu_4^{(k)} - 3(\sigma_k^2)^2$ is estimated by:

(35) $$D_k = \frac{1}{\sum_{i=0}^{k} ({}_kC_i)^4} \left\{ \frac{1}{(N-k)} \sum_{i=1}^{N-k} (\Delta^{(k)} w_i)^4 - 3({}_{2k}C_k)^2 V_k^2 \right\}$$

$$= \left[\frac{S_4^{(k)}}{N-k} - B_k V_k^2\right] C_k.$$

The constants A_{kN}, E_N, F_N, B_k, C_k have been tabulated in Tables 10, 18, 19, 14, and 16.

Anderson[11] gives the square of the standard error of the difference between two variances of consecutive differences in the following form:

(36) $$e_k^2 = E(V_{k+1} - \sigma^2)^2 + E(V_k - \sigma^2)^2 - 2E(V_k - \sigma^2)(V_{k+1} - \sigma^2),$$

(37) $$E(V_k - \sigma^2)^2 = \frac{\mu_4 - 3(\sigma^2)^2}{N-k}\left[1 - \frac{2S_{(k)}}{({}_{2k}C_k)^2 (N-k)}\right] + \frac{2(\sigma^2)^2}{N-k}\left[\frac{{}_{4k}C_{2k}}{({}_{2k}C_k)^2} - \frac{k}{2(N-k)}\right];$$

$$S_{(k)} = \sum_{i=0}^{k-1} ({}_kC_i \cdot {}_kC_{i+1})^2 + 2\sum_{i=0}^{k-2} ({}_kC_i \cdot {}_kC_{i+2})^2 + \ldots + k({}_kC_0 \cdot {}_kC_k)^2$$

and

11 O. Anderson, Die Korrelationsrechnung in der Konjunkturforschung, pp. 112 ff.

$$E(V_k-\sigma^2)(V_{k+1}-\sigma^2)=\frac{\mu_4-3(\sigma^2)^2}{N-k}\left[1-\frac{2S'_{(k)}}{{}_{2k}C_k\cdot{}_{2k+2}C_{k+1}\,(N-k-1)}\right]$$

(38)

$$+\frac{2(\sigma^2)^2}{N-k}\cdot\left[\frac{{}_{4k+1}C_{2k}}{{}_{2k}C_k\cdot{}_{2k+2}C_{k+1}}\cdot\frac{2N-2k-1}{N-k-1}-\frac{k+1}{2(N-k-1)}\right];$$

$$S'_{(k)}=\sum_{i=0}^{k-1}({}_kC_i\cdot{}_{k+1}C_{i+2})^2+2\sum_{i=0}^{k-2}({}_kC_i\cdot{}_{k+1}C_{i+3})^2+\cdots+k({}_kC_0\cdot{}_{k+1}C_{k+1})^2.$$

Anderson and Zaycoff[12] have calculated the values of these complicated formulae for $k=0, 1, 2, \cdots, 10$. We use the values V_k as the best empirical approximations to the true population variance σ^2, and the values D_k as the best empirical approximations to the true kurtosis of the population $\mu_4-3(\sigma^2)^2$. We get for the empirical approximation of the standard error:

$$e_k=\frac{V_k}{Q_k}, \tag{39}$$

$$Q_k=\frac{H_{kN}}{\sqrt{1+J_{kN}G_k}}, \tag{40}$$

$$H_{kN}=\sqrt{\frac{(N-k)(N-k-1)}{b'_kN+b''_k-\dfrac{N}{(N-k)(N-k-1)}}}, \tag{41}$$

$$G_k=D_k/V_k^2, \tag{42}$$

$$J_{kN}=\frac{b_k-\left(c_k+\dfrac{c'_k}{N}\right)\dfrac{N}{(N-k)(N-k-1)}}{b'_kN+b''_k-\dfrac{N}{(N-k)(N-k-1)}}. \tag{43}$$

The values of the constants, b, b', b'', c, and c' necessary for the calculation of the other magnitudes have been established by Zaycoff[13] and are reproduced in Table 39. We also give tables for the constants H_{kN} and J_{kN} for $k=1, 2, \cdots, 10$ and selected values of N (Tables 20, 21).

If we have a normal or nearly normal distribution, the kurtosis can be neglected compared with the square of the variance, i.e., if G_k is small. We may use as an approximation:

$$e_k^0=\frac{V_k}{H_{kN}}. \tag{44}$$

Anderson gives an asymptotic formula for large N and $k\geqq 6$:

$$(e_k^0)^2=\frac{(3k+1)\,V_k^2\,\sqrt{2\pi k}}{2(2k+1)^3\,(N-k-1)} \tag{45}$$

12 Ibid., p. 57; R. Zaycoff, op. cit., pp. 78 ff.
13 R. Zaycoff, op. cit., p. 79.

TABLE 39

COEFFICIENTS FOR THE CALCULATION OF H_{kN} AND J_{kN}

Order of Difference k	b_k	b'_k	b''_k	e_k	c'_k
0	0.500 000	1.000 000	1.000 000	0.500 000	0.000 000
1	0.277 778	0.222 222	1.111 111	0.055 556	0.444 444
2	0.254 444	0.108 889	1.093 333	0.134 444	0.286 667
3	0.209 592	0.067 347	1.080 817	0.101 837	0.384 490
4	0.187 314	0.046 838	1.072 058	0.094 092	0.415 470
5	0.169 365	0.034 973	1.065 577	0.084 752	0.462 166
6	0.156 063	0.027 391	1.060 548	0.078 301	0.500 876
7	0.145 367	0.022 202	1.056 511	0.072 936	0.538 428
8	0.136 617	0.018 466	1.053 180	0.068 562	0.573 423
9	0.129 268	0.015 672	1.050 373	0.064 872	0.606 639
10	0.122 988	0.013 518	1.047 967	0.061 715	0.638 200

We form two criteria R_k and $R_k{}^0$, based upon the accurate value (39) or approximation (44), which give the ratio of the difference between variances of successive differences and its standard error:

(46)
$$R_k = \frac{V_k - V_{k+1}}{e_k} = \frac{(V_k - V_{k+1})\,Q_k}{V_k},$$

(47)
$$R_k{}^0 = \frac{V_k - V_{k+1}}{e_k{}^0} = \frac{(V_k - V_{k+1})\,H_{kN}}{V_k}.$$

These values should be used for a judgment about the probability that the variances of the higher differences are more or less equal, and that hence the nonrandom component or mathematical expectation has been eliminated. This problem has to be decided from the point of view of fiducial limits. Generally, a value of $|R_k|$ or $|R_k{}^0| < 3$ will be sufficient.

Notes to Chapter VIII[14]

We make here the same assumptions as before but we also have to assume that the random element x_i is normally distributed with mean zero and variance σ^2. We can then use Snedecor's F table[15] or Fisher's z distribution[16] for an exact test of significance which does not involve a knowledge of the true variance of the random element and can hence be applied in even small samples. The test involves, however, some loss of information.

We propose the method of selection.[17] Suppose we want to compare the variance of the kth series and the $(k+1)$th series of differences. We will then select the items j, $j+(2k+3)$, $j+2(2k+3)$, $j+3(2k+3)$, $\cdots$ from the series of kth dif-

[14] I am greatly indebted to Professors H. Hotelling, (Columbia), S. S. Wilks (Princeton), G. W. Snedecor (Iowa State College), Dr. W. G. Madow (Columbia), and Mr. G. C. Cochran (Rothamsted) for having had the opportunity to discuss the following ideas with them.

[15] G. W. Snedecor, **Statistical Methods**, Ames, Iowa, 1938, pp. 184 ff.

[16] R. A. Fisher, **op. cit.**, pp. 250 ff., Table VI.

[17] G. Tintner, "On Tests of Significance in Time Series," **Annals of Mathematical Statistics**, Vol. 10, 1939, pp. 139 ff.

ferences and the items number $j+k+1$, $j+k+1+(2k+3)$, $j+k+1+2(2k+3)$, $j+k+1+3(2k+3)$, $\cdots$ from the series of $(k+1)$th differences, where $j=1, 2, 3, \cdots, 2k+3$. There are altogether $2k+3$ possible selections, each of which gives an independent estimate of σ^2, but which of course are not independent of each other (Table 26).

Let us now calculate the variances V'_k and V'_{k+1} from the selected items:

$$V'_k = \frac{\Sigma'(\Delta^{(k)}w_i)^2}{\frac{N-k}{2k+3}\cdot {}_{2k}C_k}, \tag{48}$$

$$V'_{k+1} = \frac{\Sigma''(\Delta^{(k+1)}w_i)^2}{\frac{N-k-1}{2k+3}\cdot {}_{2k+2}C_{k+1}}, \tag{49}$$

where Σ' and Σ'' denote summation of the *selected differences*. If $(N-k)/(2k+3)$ or $(N-k-1)/(2k+3)$ are not whole numbers, they should be replaced by the nearest integers. The correction will be small with somewhat large N. We form $z = \frac{1}{2}\log_e (V_k/V_{k+1})$. It is distributed with $n_1 = (N-k)/(2k+3)$ and $n_2 = (N-k-1)/(2k+3)$ degrees of freedom. We could also enter Snedecor's F table for these numbers of degrees of freedom. The z tends to be normally distributed for moderately large samples with a standard error of $\sqrt{(2k+3)/(N-k-1)}$ since with somewhat large N the number of degrees of freedom will be nearly equal.

We proceed as follows: Suppose we take always exactly the same number of selected items of the kth and $(k+1)$th differences in order to assure approximately the normal distribution of z, which is then based upon $(N-k-1)/(2k+3)$ degrees of freedom. According to Fisher[18] it is nearly normally distributed with mean zero and standard deviation $s(z)_{Nk} = \sqrt{(2k+3)/(N-k-1)}$. We get the values $z_{Nk}^{(5\%)} = 1.96\, s(z)_{Nk}$, $z_{Nk}^{(1\%)} = 2.576\, s(z)_{Nk}$, and $z_{Nk}^{(0.1\%)} = 3.29\, s(z)_{Nk}$ for the levels of significance, 5%, 1%, 0.1%. We write generally $z_{Nk}^{(s)}$ where $s =$ 5%, 1%, 0.1% are the levels of significance.

Let us now define the new variances of the differences:

$$\bar{V}_k = \frac{\Sigma'(\Delta^{(k)}w_i)^2}{\frac{N-k-1}{2k+3}\cdot {}_{2k}C_k} = \frac{\bar{S}_2^{(k)}}{\frac{N-k-1}{2k+3}\cdot {}_{2k}C_k}, \tag{50}$$

$$\bar{V}_{k+1} = \frac{\Sigma''(\Delta^{(k+1)}w_i)^2}{\frac{N-k-1}{2k+3}\cdot {}_{2k+2}C_{k+1}} = \frac{\bar{S}_2^{(k+1)}}{\frac{N-k-1}{2k+3}\cdot {}_{2k+2}C_{k+1}}, \tag{51}$$

where again Σ' is summation over the selected values of the kth differences and Σ'' summation over the selected values of the $(k+1)$th differences. The symbols $\bar{S}_2^{(k)}$ and $\bar{S}_2^{(k+1)}$ are the sums of squares of the selected differences. Then, according to Fisher's analysis of variance, half the natural logarithm of the ratio of the variances must be smaller than $z_{Nk}^{(s)}$ if the variances are not to be considered unequal from the point of view of a level of significance s.

[18] R. A. Fisher, op. cit., p. 233, Chapter VII, Section 41.

The ratio should always be greater than 1, since z is not negative. Hence we have to distinguish two cases: If $\bar{V}_k > \bar{V}_{k+1}$,

$$\tfrac{1}{2} \log_e \frac{\bar{V}_k}{\bar{V}_{k+1}} < z_{Nk}{}^{(s)}, \tag{52}$$

or

$$\frac{\bar{V}_k}{\bar{V}_{k+1}} = \frac{\bar{S}_2{}^{(k)} \cdot {}_{2k+2}C_{k+1}}{\bar{S}_2{}^{(k+1)} \cdot {}_{2k}C_k} < \exp(2z_{Nk}{}^{(s)}) = F_{Nk}{}^{(s)}, \tag{53}$$

where $F_{Nk}{}^{(s)}$ is Snedecor's variance ratio for the level of significance s and the number of degrees of freedom $n_1 = n_2 = (N-k-1)/(2k+3)$.

Introducing the symbol $\alpha_k = {}_{2k+2}C_{k+1}/{}_{2k}C_k$ we get

$$\frac{\bar{S}_2{}^{(k)}}{\bar{S}_2{}^{(k+1)}} < \frac{F_{Nk}{}^{(s)}}{\alpha_k}. \tag{54}$$

If, on the other hand, $\bar{V}_k < \bar{V}_{k+1}$, we get by similar reasoning:

$$\frac{\bar{S}_2{}^{(k)}}{\bar{S}_2{}^{(k+1)}} > \frac{1}{F_{Nk}{}^{(s)}\, \alpha_k}. \tag{55}$$

We finally have the following limits for the ratio of the sum of squares of selected kth differences to the sum of squares of selected $(k+1)$th differences:

$$\frac{1}{F_{Nk}{}^{(s)}\, \alpha_k} < \frac{\bar{S}_2{}^{(k)}}{\bar{S}_2{}^{(k+1)}} < \frac{F_{Nk}{}^{(s)}}{\alpha_k}. \tag{56}$$

These limits are of course different for different levels of significance s and have been tabulated for selected N and for $k = 0, 1, \cdots, 9$, $s = 5\%, 1\%, 0.1\%$, in Tables 27, 28, 29.

Notes to Chapter IX

The following ingenious method, proposed by Anderson,[19] connects the variate difference method with Sheppard's smoothing formulae,[20] and, incidentally, also with the Gram polynomials.[21]

Suppose that we have established to our satisfaction that the nonrandom element or mathematical expectation of the time series is eliminated to a considerable degree by taking k_0 differences. We will, in order to simplify our notation, put $k_0/2 = n$ for k_0 even and $(k_0+1)/2 = n$ for k_0 odd.

We know that only the random element remains in the $2n$th and higher differences. But we cannot reconstruct it from the differences. We will try to find a function f_i, such that the variance:

$$D = E(x_i + f_i)^2 = \text{minimum}. \tag{57}$$

19 O. Anderson, op. cit., pp. 117 ff.

20 E. T. Whittaker and G. Robinson, The Calculus of Observations, London, 1924, pp. 291 ff.

21 H. T. Davis, Tables of Higher Mathematical Functions, Vol. 2, Bloomington, Indiana, 1935, pp. 307 ff.

Assume that f_i is a linear function of the $2n$th and higher central differences of x_i or w_i:

$$f_i = b_1 \Delta^{(2n)} x_{i-n} + b_2 \Delta^{(2n+2)} x_{i-n-1} + \cdots, \tag{58}$$

where the b's are constants, which are to be determined by the method of least squares. But remembering that:

$$\Delta^{(2n)} x_{j-n} = (-1)^n [{}_{2n}C_0 (x_{j+n} + x_{j-n}) - {}_{2n}C_1 (x_{j+n-1} + x_{j-n+1}) + \cdots + (-1)^n {}_{2n}C_n x_j], \tag{59}$$

we may also write f_i as a weighted average:

$$f_i = \sum_{j=1}^{m} W_j (x_{i+j} + x_{i-j}) + W_0 x_i, \tag{60}$$

where the W_j are certain weights, if we decide to take approximation $m-n+1$, i.e., to include only the terms, $b_1, \cdots, b_{m-n+1}$ in formula (58). This gives $(2m+1)$ terms in (60). The values of the b's which give the least-square solution have been established by Sheppard and Anderson. They give the optimum weights, say $g_{nm}(j)$, for W_j and have been tabulated in Table 32. Replacing the $2n$th and higher differences of x_i in (58) by the same differences of w_i (they are by assumption equal since m_i has been eliminated) we get the smoothing formula:

$$m_i' = \sum_{j=1}^{m} g_{nm}(j) (w_{i+j} + w_{i-j}) + g_{nm}(0) w_i, \tag{61}$$

where m_i' is the approximation to the true mathematical expectation m_i and w_i an item of our original time series. The variance of m' is found by:

$$\sigma^2(m') = \sigma^2 \left[2 \sum_{j=1}^{m} g_{nm}{}^2(j) + g_{nm}{}^2(0) \right]; \tag{62}$$

and as an approximation we get:

$$V(m') = V_{k_0} L_{nm}, \tag{63}$$

where L_{nm} is the term in brackets in (62) (Table 33). If $v = \sqrt{V_{k_0}} / \bar{w}$ is the coefficient of random variability, we get for the coefficient of variability of the approximation to the mathematical expectation:

$$v(m') = v\sqrt{L_{nm}} \quad \text{(Table 34)}.$$

It is interesting to note the connection between our weights $g_{nm}(j)$ of the moving averages of Sheppard's formulae and the Gram polynomials as tabulated by Professor H. T. Davis.[22] We have for instance:

$$g_{1m}(j) = A, \tag{64}$$

where A is the value for $p = m$ of his coefficients for the fitting of a straight line;[23]

22 Ibid. I am greatly indebted to Professor H. T. Davis (Northwestern University) for having brought this connection to my attention.

23 Ibid., pp. 326 ff.

$$g_{2m}(j) = A + j^2B\,, \tag{65}$$

where A and B are the values for $p = m$ in his tables for fitting a parabola;[24]

$$g_{3m}(j) = A + j^2B + j^4C\,, \tag{66}$$

where A, B, and C are the values for $p = m$ for fitting a quartic;[25] and

$$g_{4m}(j) = A + j^2B + j^4C + j^6D\,, \tag{67}$$

where A, B, C, and D are the values for $p = m$ for his tables for fitting a sextic;[26] and similarly for higher values of the g's.

Notes to Chapter X

Let $w(x)_i$ and $w(y)_i$ be two time series of the same type as discussed; $m(x)_i$ and $m(y)_i$ are the mathematical expectations of the two series, which again are "smooth" and can be eliminated by differencing. The random elements x_i and y_i have the mathematical expectation zero and the population variances $\sigma^2(x)$ and $\sigma^2(y)$. They are not correlated with the mathematical expectations and only corresponding items x_i and y_i are correlated. The product moment of the population is π.

$$w(x)_i = m(x)_i + x_i\,, \tag{68}$$

$$w(y)_i = m(y)_i + y_i\,, \tag{69}$$

$$E(x_i) = E(y_i) = 0\,, \tag{70}$$

$$E(x_i^2) = \sigma^2(x), \qquad E(y_i^2) = \sigma^2(y)\,, \tag{71}$$

$$E(x_iy_i) = \pi\,, \qquad i = 1, 2, \cdots, N\,, \tag{72}$$

$$E(x_iy_j) = E(x_ix_j) = E(y_iy_j) = 0\,, \qquad i \neq j\,, \tag{73}$$

$$E[m(x)_ix_j] = E[m(x)_iy_j] = E[m(y)_ix_j] = E[m(y)_iy_j] = 0\,, \qquad i, j = 1, 2, \cdots, N\,. \tag{74}$$

We want first to test by Anderson's method[27] the stability of the product moment by its standard error, which is possible only in large samples. We can prove by an argument similar to that for the variances of the differences that the properly adjusted product moments of the differences of random series are equal. The standard error has the same form as (36), except that we get $E(x^2y^2) - E(x^2)E(y^2) - 2[E(xy)]^2$ instead of $\mu_4 - 3(\sigma^2)^2$ and $[E(xy)]^2 + E(x^2)E(y^2)$ instead of $2(\sigma^2)^2$. Hence the formulae calculated by Anderson and Zaycoff can be used with little adaptation.

Let $\bar{w}(x)$ and $\bar{w}(y)$ be the arithmetic means of the two original series and $S^{(k)}(xy)$ the sum of the products of their kth differences:

$$S^{(k)}(xy) = \sum_{i=1}^{N-k} \Delta^{(k)}w(x)_i\,\Delta^{(k)}w(y)_i\,, \qquad k = 0, 1, 2, \dots. \tag{75}$$

[24] Ibid., pp. 332 ff.
[25] Ibid., pp. 342 ff.
[26] Ibid., pp. 350 ff.
[27] O. Anderson, op. cit., pp. 123 ff.

Define the empirical approximations to the product moments as:

(76) $$p_0 = \left[\sum_{i=1}^{N} [w(x)_i - \overline{w}(x)][w(y)_i - \overline{w}(y)] \right] / (N-1)$$

$$= [S^{(0)}(xy) - N\overline{w}(x)\overline{w}(y)] / (N-1),$$

(77) $$p_k = \left[\sum_{i=1}^{N-k} \Delta^{(k)} w(x)_i \, \Delta^{(k)} w(y)_i \right] / (N-k) \; {}_{2k}C_k$$

$$= S^{(k)}(xy) A_{kN}, \qquad k = 1, 2, \cdots.$$

If we decide to neglect the first term mentioned above, i.e., the term involving the product moment of the squares of x and y (which would be justified in the case of normal or nearly normal correlation), we get as an approximation to the standard error of the difference between the empirical product moments of the kth and the $(k+1)$th differences, $(p_k - p_{k+1})$:

(78) $$\hat{e}_k{}^0 = \frac{\sqrt{L_k}}{H_{kN}},$$

where L_k is an empirical approximation to $[\pi^2 + \sigma^2(x)\sigma^2(y)]/2$,

(79) $$L_k = \frac{p_k{}^2 + V_k(x)\, V_k(y)}{2},$$

and $V_k(x)$ and $V_k(y)$ are the sample variances of the kth differences of the series $w(x)_i$ and $w(y)_i$ as defined in (32) and (33).

It is also possible to form again an approximate criterion $\hat{R}_k{}^0$:

(80) $$\hat{R}_k{}^0 = \frac{(p_k - p_{k+1}) H_{kN}}{\sqrt{L_k}}.$$

A fiducial limit $|\hat{R}_k{}^0| < 3$ will in most cases be accurate enough for testing the k_0th difference beginning from which the product moments contain only the random elements x_i and y_i.

We get a more exact criterion $\hat{R}_k$ if we do not neglect the term $E(x^2y^2) - E(x^2)E(y^2) - 2[E(xy)]^2$. In order to find the best approximation to its population value we have to calculate:

(81) $$S^{(0)}(x^\alpha y^\beta) = \sum_{i=1}^{N} [w(x)_i]^\alpha [w(y)_i]^\beta$$

for the original two series; and for the original series and differences:

(82) $$S^{(k)}(x^2y^2) = \sum_{i=1}^{N-k} [\Delta^{(k)} w(x)_i]^2 [\Delta^{(k)} w(y)_i]^2, \qquad k = 1, 2, \cdots.$$

Further we must obtain the second-order product moment of the two original series:

(83) $$m_{2,2} = S^{(0)}(x^2y^2) - 2\overline{w}(y)\, S^{(0)}(x^2y) - 2\overline{w}(x)\, S^{(0)}(xy^2)$$

$$+ \overline{w}(y)^2 S_2{}^{(0)}(x) + \overline{w}(x)^2 S_2{}^{(0)}(y) + 4\overline{w}(x)\, \overline{w}(y)\, S^{(0)}(xy) - 3N\overline{w}(x)^2 \overline{w}(y)^2.$$

The best empirical approximation to the term $E(x^2y^2) - E(x^2)E(y^2) - 2[E(xy)]^2$ is, then, in terms of the original two series:

(84)
$$M_0 = \frac{\dfrac{m_{2,2}}{N} - 3\left(\dfrac{N-1}{N}\right)^2 L_0}{1 - \dfrac{4}{N} + \dfrac{6}{N^2} - \dfrac{3}{N^3}} = \left[\frac{m_{2,2}}{N} - E_N L_0\right] F_N ;$$

and in terms of the difference series of both time series:

(85)
$$M_k = \frac{\dfrac{S^{(k)}(x^2y^2)}{N-k} - 3({}_{2k}C_k)^2 L_k}{\sum_{i=0}^{k} ({}_kC_i)^4}$$
$$= \left[\frac{S^{(k)}(x^2y^2)}{N-k} - B_k L_k\right] C_k , \qquad k = 1, 2, \ldots .$$

We form the quantities:

(86)
$$\hat{Q}_k = H_{kN} / \sqrt{1 + (J_{kN} M_k / L_k)} ;$$

and the true criterion appears as:

(87)
$$\hat{R}_k = (p_k - p_{k+1}) \hat{Q}_k / \sqrt{L_k} ,$$

the ratio of the difference $p_k - p_{k+1}$ to its own standard error.

In order to apply Fisher's test[28] we have to calculate our empirical product moments, say p'_k and p'_{k+1}, from selected and hence independent items of the kth and $(k+1)$th differences of the two series. Then we calculate the empirical correlation coefficients r'_k and r'_{k+1} for two selections of kth and $(k+1)$th differences (Table 26):

(88)
$$r'_k = \frac{p'_k}{\sqrt{V'_k(x)\, V'_k(y)}} ,$$

and similarly for r'_{k+1}. We then make the transformation: $z'_k = \tfrac{1}{2}[\log_e(1 + r'_k) - \log_e(1 - r'_k)]$, and similarly for z'_{k+1}.[29] The difference $z'_k - z'_{k+1}$ will be normally distributed with a standard error of $\sqrt{1/(n-3) + 1/(n-1)}$ if $k = 0$ and $\sqrt{2/(n-1)}$ if $k = 1, 2, \ldots$. The n is the number of items contained in the selection. The difference in the number of degrees of freedom arises from the fact that we have not calculated the means of the differences.

This test of Fisher will again give us the k_0th difference beginning from which we can assume that the correlations contain only the random elements since they are approximately equal.

We can use our knowledge of the k_0th difference beginning from which we can be reasonably sure that only the random element is contained in our series,

[28] R. A. Fisher, op. cit., pp. 208 ff., Chapter VI, Section 35; G. Tintner, loc. cit., pp. 139 ff.
[29] R. A. Fisher and F. Yates, Statistical Tables, London, 1938, p. 36; R. A. Fisher, op. cit., p. 215, Table V-B.

for getting an approximation to the linear relationship between the mathematical expectations $m(x)_i$ and $m(y)_i$.[30] Denoting their product moments and variances by m, we have:

$$p_0 = \frac{N-1}{N} p_{k_0} + p(m), \tag{89}$$

$$V_0(x) = \frac{N-1}{N} V_{k_0}(x) + V[m(x)], \tag{90}$$

$$V_0(y) = \frac{N-1}{N} V_{k_0}(y) + V[m(y)], \tag{91}$$

where the terms with the index k_0 are the best empirical approximations to the product moment, and the variances of the population of the random elements. From these relationships we can calculate the parameters referring to the mathematical expectations and we get finally as a measure of the linear relationship:

$$r(m) = \frac{p(m)}{\sqrt{V[m(x)]\, V[m(y)]}}. \tag{92}$$

The correlation between the random elements is given by:

$$r_{k_0} = \frac{p_{k_0}}{\sqrt{V_{k_0}(x)\, V_{k_0}(y)}}. \tag{93}$$

30 O. Anderson, op. cit., p. 124; H. Wold, Analysis of Stationary Time Series, Uppsala, 1938, pp. 200 ff.

APPENDIX III

SEASONAL VARIATION

It has been suggested by Dr. A. Wald in his monograph on seasonal variations[1] that sometimes these fluctuations may prevent the successful application of the variate difference method. This would be the case if they belong to the category of extreme zigzag oscillations with strong negative correlation between subsequent items. See also above, pp. 16 ff. We have, therefore, made some experiments on the monthly wool prices as compiled by the United States Department of Labor. The period covered is 1890 to 1937.

TABLE 40

SUMMARY, SEASONAL MONTHLY WOOL PRICES, 1890-1937

Month	Monthly Means	Month	Monthly Means
January	0.6335	July	0.6212
February	0.6344	August	0.6260
March	0.6356	September	0.6242
April	0.6298	October	0.6210
May	0.6235	November	0.6235
June	0.6127	December	0.6221

SUMS OF SQUARES OF THE SEASONAL

Order of Difference k	$S^{*(k)}_2$	Order of Difference k	$S^{*(k)}_2$
0	4.697 36729	6	0.104 73618
1	0.000 43916	7	0.397 82858
2	0.000 82308	8	1.530 33384
3	0.002 32342	9	5.940 61810
4	0.007 79800	10	23.218 43194
5	0.028 10450	....	....

We have computed the seasonal in the following way. We took the means of the Januaries, Februaries, etc. (Table 40) and using these means as our data we carried through a difference analysis. The result of this difference analysis is shown in Table 41. We have corrected our variances of the original series and of the higher differences V_k of the wool prices for the seasonal variation by deducting from the original variances the part which is more or less explained by

[1] A. Wald, Berechnung und Ausschaltung von Saisonschwankungen, Vienna, 1936. See also, O. Anderson in Zeitschrift für Nationalökonomie, Vol. 8, 1937, pp. 251 ff.; R. Zaycoff, "Ueber die Ausschaltung der zufälligen Komponente nach der 'Variate-Difference' Methode," Publications of the Statistical Institute for Economic Research, State University of Sofia, 1937, No. 1, pp. 100 ff.

the seasonals. The assumption is that there is no correlation between the seasonal and the random element. Then we get a result which is not very different from the one shown previously.[2]

TABLE 41

DIFFERENCE ANALYSIS, MONTHLY WOOL PRICES, 1890-1937
(*corrected for seasonal*)

Order of Difference	Variances Corrected for Seasonal	Percentage Seasonal	Approximate Standard-Error Ratio Corrected for Seasonal
k	V^*_k		R_k^{*0}
0	0.112 0	0.09	23.5178
1	0.002 035	0.93	20.1179
2	0.001 225	0.97	8.0857
3	0.001 087	0.82	5.5869
4	0.001 020	0.87	5.3627
5	0.000 9693	0.96	4.3444
6	0.000 9353	1.02	3.4117
7	0.000 9123	1.06	2.9135
8	0.000 8949	1.11	2.4103
9	0.000 8819	1.17	1.6997
10	0.000 8735	1.20	....

Our corrected variance V_k^* can be calculated by the following formulae: $V_0^* = [S_2^{(0)} - N\,\bar{w}^2 - (N/12)\ (S_2^{*(0)} - 12\bar{w}^2)]/(N-1)$ for the original series; and for the differences: $V_k^* = [S_2^{(k)} - (N/12)\ S_2^{*(k)}]\ A_{kN}$, $k = 1, 2, \cdots$, where $\bar{w}$ is the mean of the original series, $S_2^{(k)}$ the sums of the squares of the kth differences of the uncorrected series, and $S_2^{(k)*}$ the sum of the squares of the kth differences of the seasonal as defined above (Table 40). The coefficient A_{kN} is given in Table 10.

We show in Table 41 the corrected variances V_k^* for the wool prices as well as the percentage of the seasonal. It appears that these percentages do not vary greatly. The percentage of the seasonal is throughout about 1%. (Table 41). Hence it is not probable that the seasonal in this price is of the nature of an extreme zigzag element which would make the variate difference method inapplicable. In this case the percentage of the seasonal in the variances would increase the higher the difference. Since it stays more or less stable it follows that we may be able to make the assumption that the seasonal in this commodity is not of a nature which would prevent the application of the ordinary methods of separating the mathematical expectation and the random element by taking higher differences.

This point of view is also confirmed by Table 41. Here we show difference analyses of the variances which have been corrected for the seasonal V_k^*. We use only an approximate criterion R_k^{*0} which does not make any use of the kurtosis. This criterion is calculated according to the formula:

[2] See Appendix I, Sections B and C, for a summary of computations.

$$R_k^{*0} = (V_k^* - V_{k+1}^*) H_{kN} / V_k^*$$

and is the ratio of the difference between the variances of the kth and the $(k+1)$th differences of the series corrected for the seasonal $(V_k^* - V_{k+1}^*)$ to its standard error.

The criteria lead more or less to the same results and the differences in the old and new R's (Tables 24 and 41) are really very minute. In the case of the monthly wool prices we should assume that we probably have eliminated the mathematical expectation in the eighth difference because our criteria then become smaller than 2 and stay so. The results would probably have been even better had we calculated the more accurate criterion which involves the kurtosis, but it would have meant some additional labor.

We suppose that this experiment has been successful in so far as our results are not much affected by the elimination of the seasonal. Hence we should be led to believe that, at least in price data, we shall rarely if ever find a seasonal element which is of this extreme zigzag nature. Should this, however, be the case we have given this method for correcting for the seasonal variation and avoiding any possible error.

APPENDIX IV

THE STANDARD ERRORS OF SOME DERIVED STATISTICAL SERIES

We have seen how the variate difference method permits us to estimate the random variance σ^2. We have also given some formulae in Chapter IX which indicate the reduction of the original random variance by the use of appropriate Sheppard smoothing formulae (Table 33).

The problem which we want to treat here is the following: Suppose we have a time series and the estimate of its random variance V_{k_0}. What is the random variance or the standard error of some statistical series derived from the original series by applying moving averages? This is a continuation and amplification of some formulae given in a previous publication.[1]

We want first to state again two propositions which are important for the method of moving averages. This procedure seems to us preferable to any other method when dealing with economic time series. Suppose we have given a time series with the general term y_i. We smooth this with certain weights W', where $\sum_{i=-n}^{n} W'_i = 1$. This moving average hence has the length $2n+1$. We have the smoothed values: $y'_0 = \sum_{i=-n}^{n} W'_i\, y_i$. We smooth the series y' again with another moving average with the weights W'', $\sum_{i=-m}^{m} W''_i = 1$, which has the length $2m+1$. We get $y''_0 = \sum_{i=-m}^{m} W''_i\, y'_i$. This process of double smoothing can be replaced by a simple smoothing,[2] $y''_0 = \sum_{i=-m-n}^{m+n} W'''_i\, y_i$, of the original series with a moving average with weights $W'''_i = \sum_{j=0}^{m+n-i} W'_{i-m+j}\, W''_{m-j}$.

If we derive from an original series y with random variance σ^2 by mechanical smoothing a new series $y'_i = \sum_{j=-n}^{n} W'_j\, y_{i+j}$, the variance of the new series y' is equal to $\sigma^2 \sum_{i=-n}^{n} W'^2_i$, provided we make the assumption that the errors in the original y_i are mutually independent and have the same variance. This formula covers a wide range of parameters. If we take for instance the arithmetic mean of the series y, or make the weights $W_j = 1/N$ where N is the number of items in the series, we get the well-known formula for the variance of the arithmetic mean,[3] σ^2/N.

It is also of some interest to investigate the serial correlations which are introduced by applying moving averages. Suppose again that we have a random element y_i whose items are independently distributed about zero with a variance σ^2. We apply a moving average W' to the series y_i, which contains $2n+1$ items. We have: $y'_0 = \sum_{i=-n}^{n} W'_i\, y_i$. The serial correlation between the items y_j and y'_{j+k} is then equal to: $r_{(K)} = (\sum_{i=0}^{2n-K} W'_{-n+i}\, W'_{K-n+i}) / (\sum_{i=-n}^{n} W'^2_i)$.

To give a simple example:[4] Suppose that a time series with a random vari-

[1] G. Tintner, **Prices in the Trade Cycle**, Vienna, 1935, pp. 84 ff.
[2] Ibid., p. 84, formula (52).
[3] See, e.g., H. T. Davis and W. F. C. Nelson, **Elements of Statistics**, 2nd ed., Bloomington, Indiana, 1937, p. 195.
[4] G. Tintner, op. cit., p. 85.

ance V_{k_0} is smoothed with a thirteen-month moving average in order to eliminate seasonal variation. The weights W' are in this case: 1/24, 1/12, ··· , 1/12, 1/24. The random variance of the smoothed series is $[(2/24^2) + (11/12^2)] V_{k_0} = 0.0799 V_{k_0}$. We get for the serial correlations of items of the new series which is free of seasonal variations: for $K=1$, i.e., items distant by one time unit, $r_{(1)} = 0.96$; for $K = 2$, i.e., items distant by two time units, $r_{(2)} = 0.87$. The values of the other serial correlations are $r_{(3)} = 0.78$, $r_{(4)} = 0.70$, $r_{(5)} = 0.61$, $r_{(6)} = 0.52$, $r_{(7)} = 0.44$, $r_{(8)} = 0.35$, $r_{(9)}$ 0.26, $r_{(10)} = 0.17$, $r_{(11)} = 0.09$, $r_{(12)} = 0.002$. All other serial correlations are zero. There is no serial correlation between items in the new series distant from each other by 13 or more units.

APPENDIX V

Alternative Methods

A. Sequences and Reversals

An interesting alternative method for the treatment of time series has been developed by Mr. Herbert E. Jones of the Cowles Commission. He studies random series by the method of sequences and reversals.[1]

A sequence is defined as the case in which two or more consecutive first differences of the series have like signs. A reversal occurs at a change of signs of the first differences.

The expected number of reversals in a random series with normal distribution is given by the following asymptotic formula:

$$E(R) = (2/3)(n-2),$$

where $E(R)$ is the expected number of reversals, n the number of items in the series. The standard error is given by:

$$\sigma_R = \sqrt{2n/3}\,.$$

We treat our example of the annual prices of wheat flour according to this method. The results are shown in Table 42.

Its meaning is this: In our original series for instance (zeroth difference) we

TABLE 42

Reversals, Annual American Wheat-Flour Prices, 1890-1937

Order of Difference	Number of Items	Number of Reversals	Expected Number of Reversals	Standard Error
k	n	R	$E(R)$	σ_R
0	48	19	30.7	3.3
1	47	25	30.0	3.2
2	46	30	29.4	3.2
3	45	33	28.6	3.2
4	44	32	28.0	3.1
5	43	32	27.4	3.1
6	42	32	26.7	3.1
7	41	31	26.0	3.1
8	40	30	25.4	3.0
9	39	29	24.7	3.0

[1] H. E. Jones, "Theory of Runs as Applied to Time Series," Cowles Commission for Research in Economics, **Report of Third Annual Research Conference on Economics and Statistics**, Colorado Springs, 1937, pp. 33 ff.; A. Cowles and H. E. Jones, "Some a Posteriori Probabilities in Stock Market Action," **Econometrica**, Vol. 5, 1937, pp. 280 ff.; see also L. von Bortkiewicz, **Die Iterationen**, Berlin, 1917.

have 48 items and 19 reversals. This means its own difference series, i.e., our series of first differences, changes 19 times from plus to minus or from minus to plus. Next, the series of first differences includes 47 items and has 25 reversals. Its own difference series, that is our series of second differences, shows 25 times a change from plus to minus or from minus to plus, etc.

It is interesting to note that this analysis enforces the results already established in Chapters VII and VIII with the help of the variate difference method. We saw there that the mathematical expectation was probably already eliminated in the second or third difference and that all higher differences hence contain only the random element. We see from Table 42 above that the difference between the actual and the expected number of reversals exceeds two times its standard error only in the original series (0th difference). In all other differences the expected and the actual values of the number of reversals come so close that they could be considered as true random series.

B. Serial Correlations

Professor G. U. Yule[2] was the first to consider extensively the problem of serial or autocorrelations. Prof. H. T. Davis[3] also established some significant results in this field. A recent book of Dr. H. Wold[4] is an attempt to study stationary time series mainly from the point of view of serial correlations.

The close connection of serial correlations with the fundamental ideas of the variate difference method has been established by Anderson.[5] He showed the relationship between the serial correlation coefficients ($r_{(L)}$) and the variances of the differences of a random series (V_k).

His result appears as follows: We form the serial or autocorrelations of a time series by correlating the series with itself, lagged by a lag L. Hence every item w_i of the original series is correlated with the item w_{i+L}. It appears that we have $r_{(0)} = 1$ since the series correlated with itself has of course perfect correlation (lag zero). Further $r_{(L)} = r_{(-L)}$. A serial correlation with negative lag is the same as the one with positive lag. All serial correlations (except for lag zero) are zero in the case of independence.

We form a series $\cdots r_{(-4)}, r_{(-3)}, r_{(-2)}, r_{(-1)}, r_{(0)}, r_{(1)}, r_{(2)}, r_{(3)}, r_{(4)}, \cdots$ and the central differences of $r_{(0)}$. The $2k$th central difference is denoted by $\Delta^{(2k)} r_{(-k)}$. It can be shown[6] that in a true random series the central difference $\Delta^{(2k)} r_{(-k)}$ is proportional to $V_k \cdot {}_{2k}C_k$, the variance of the kth difference. We have also for a long series:

$${}_{2k}C_k \cdot V_k = V_0[{}_{2k}C_k - 2 \cdot {}_{2k}C_{k+1} \cdot r_{(1)} + 2 \cdot {}_{2k}C_{k+2} \cdot r_{(2)} - \cdots + (-1)^k 2 r_{(k)}] .$$

We show in Table 43 the serial correlations of the annual wheat-flour price

[2] G. U. Yule, "Why Do We sometimes Get Nonsense-Correlations Between Time Series?" Journal of the Royal Statistical Society, Vol. 89, 1926, pp. 103 ff. See also H. Von Schelling, Die wirtschaftlichen Zeitreihen als Problem der Korrelationsrechnung, Bonn, 1931.

[3] H. T. Davis, "The Econometric Problem," Cowles Commission for Research in Economics, Report of Third Annual Research Conference on Economics and Statistics, Colorado Springs, 1937, pp. 11 ff.; H. T. Davis, "Mathematical Adventures in Social Science," American Mathematical Monthly, Vol. 45, 1938, pp. 101 ff.

[4] H. Wold, A Study in the Analysis of Stationary Time Series, Uppsala, 1938. See also M. M. Flood, "Recursive Methods and the Analysis of Time Series," Cowles Commission for Research in Economics, Report of Fourth Annual Research Conference on Economics and Statistics, Colorado Springs, 1938, pp. 90 ff.

[5] O. Anderson, Die Korrelationsrechnung in der Konjunkturforschung, Bonn, 1929, pp. 11 ff., 114 ff.

[6] Ibid., p. 114.

TABLE 43

SERIAL CORRELATION COEFFICIENTS
ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Lag in Years	Serial Correlation Coefficient	Lag in Years	Serial Correlation Coefficient	Lag in Years	Serial Correlation Coefficient
L	$r_{(L)}$	L	$r_{(L)}$	L	$r_{(L)}$
1	0.8546	17	—0.0905	33	—0.0925
2	0.6854	18	—0.1025	34	0.0876
3	0.5114	19	—0.1292	35	0.3604
4	0.3489	20	—0.1522	36	0.5592
5	0.3091	21	—0.2766	37	0.7511
6	0.2724	22	—0.3957	38	0.5865
7	0.2820	23	—0.5983	39	0.0263
8	0.2950	24	—0.6487	40	—0.6390
9	0.2427	25	—0.5830	41	—0.9765
10	0.1471	26	—0.3727	42	—0.7744
11	0.0096	27	—0.1414	43	—0.4128
12	—0.0999	28	0.0881	44	0.6417
13	—0.1564	29	0.3473	45	0.4991
14	—0.1580	30	0.4061	46	—0.9333
15	—0.1098	31	0.2629		
16	—0.0830	32	0.0084		

data 1890 to 1937. We have calculated all possible serial correlation coefficients from a lag $L = 1$ to a lag $L = 46$. Their meaning is as follows: The annual wheat-flour prices (as given in Table 7) have been correlated with themselves shifted by one year for a lag $L = 1$. That is, we correlate the price for 1890 with the price for 1891, the price for 1891 with the one for 1892, etc. The result of this correlation is a serial correlation coefficient for $L = 1$, $r_{(1)} = 0.8546$. In the same way we found the serial correlation for a lag $L = 2$. Here we correlated the price for 1890 with the price for 1892, the price for 1891 with the one for 1893, etc. This gives us a serial correlation coefficient for $L = 2$, $r_{(2)} = 0.6854$. The other serial correlation coefficients have been formed in the same way. The correlogram[7] is shown in Figure 4.

We complete this series by adding to it the serial correlation coefficients for negative lags, remembering that $r_{(-L)} = r_{(L)}$. Then we difference the resulting series and form the central differences for $r_{(0)}$, $\Delta^{(2k)} r_{(-k)}$. They are shown in Table 44 for $k = 0, 1, 2, \cdots, 10$. We see that beginning from about $k = 2$ the ratios $\Delta^{(2k)} r_{(-k)} / {}_{2k}C_k \cdot V_k$ become rather stable and are all about 1/5. We should hence conclude that the random element alone or almost alone is present beginning from the second or third difference. Or, that we have then eliminated to a considerable degree the nonrandom element or the mathematical expectation of the annual wheat-flour prices. This is the same result as the one established previously by the difference analysis (Chapter VII) or by the test of significance (Chapter VIII). The analysis of the serial correlations of the annual wheat-flour prices substantiates our previous results.

[7] H. Wold, op. cit., p. 147 ff.

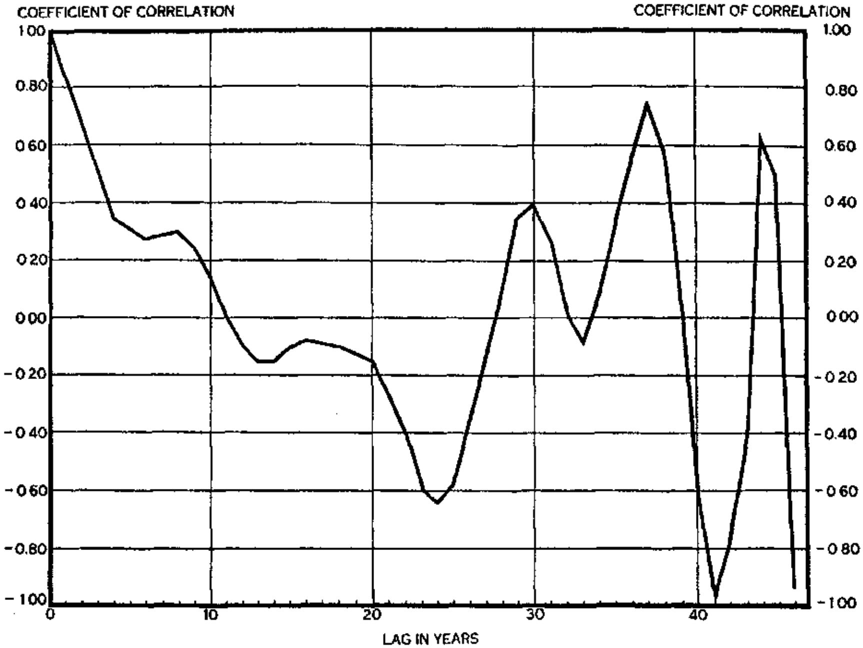

FIGURE 4.—SERIAL CORRELATION COEFFICIENTS, ANNUAL WHEAT-FLOUR PRICES, 1890-1937

TABLE 44

ANALYSIS OF SERIAL CORRELATION
ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Order of Difference	Central Differences of Serial Correlation Coefficients	Ratio
k	$\Delta^{(2k)} r_{(-k)}$	$\Delta^{(2k)} r_{(-k)} / {}_{2k}C_k \cdot V_k$
0	1.0000	0.2085
1	— 0.2908	—0.2071
2	0.5340	0.2022
3	— 1.5640	—0.1989
4	5.1826	0.1965
5	—17.8978	—0.1939
6	63.0248	0.1923
7	—224.5974	—0.1869
8	807.2740	0.1831
9	—2,921.3642	—0.1802
10	10,632.4254	0.1775

TABLE 45

SERIAL CORRELATION COEFFICIENTS, SELECTED COMPARISONS OF ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Lag in Years L	Selection	Number of Items in the Selection n	Serial Correlation Coefficient $r'_{(L)}$
1	1-α	24	0.7648***
	1-β	23	0.9485***
2	2-α	16	0.6852**
	2-β	15	0.7786***
	2-γ	15	0.6991**
3	3-α	12	0.4470
	3-β	11	0.4826
	3-γ	11	0.3257
	3-δ	11	0.7966**
4	4-α	9	0.3932
	4-β	9	0.5189
	4-γ	9	0.6121
	4-δ	9	0.1047
	4-ε	8	0.1966
5	5-α	8	0.2988
	5-β	7	0.4539
	5-γ	7	0.7833*
	5-δ	7	0.6495
	5-ε	7	0.2185
	5-ζ	7	0.2064
6	6-α	6	0.0013
	6-β	6	0.2779
	6-γ	6	0.2214
	6-δ	6	0.3176
	6-ε	6	0.9080*
	6-ζ	6	0.8987*
	6-η	6	0.3087
7	7-α	6	0.4242
	7-β	5	0.8540*
	7-γ	5	0.3949
	7-δ	5	0.3867
	7-ε	5	0.3579
	7-ζ	5	0.4582
	7-η	5	0.1403
	7-θ	5	-0.2283
8	8-α	5	0.8480
	8-β	5	0.3762
	8-γ	5	0.3005
8	8-δ	5	0.1357
	8-ε	4	0.2644
	8-ζ	4	0.0214
	8-η	4	-0.3210
	8-θ	4	-0.1136
	8-ι	4	0.2804
9	9-α	4	0.0265
	9-β	4	-0.1361
	9-γ	4	0.0336
	9-δ	4	-0.2103
	9-ε	4	0.4900
	9-ζ	4	0.7294
	9-η	4	0.8184
	9-θ	4	0.7424
	9-ι	4	-0.0264
	9-κ	3	0.0881
10	10-α	4	0.2100
	10-β	4	0.3609
	10-γ	4	0.7655
	10-δ	4	0.6785
	10-ε	4	0.8033
	10-ζ	3	0.6233
	10-η	3	0.0560
	10-θ	3	-0.0595
	10-ι	3	-0.3678
	10-κ	3	-0.2715
	10-λ	3	-0.4562
11	11-α	4	0.3077
	11-β	3	0.7631
	11-γ	3	0.9003
	11-δ	3	0.6267
	11-ε	3	-0.2084
	11-ζ	3	-0.1667
	11-η	3	-0.3769
	11-θ	3	-0.3406
	11-ι	3	-0.2485
	11-κ	3	0.6938
	11-λ	3	0.9961
	11-μ	3	0.4662

* Falls outside 5% limits ** Outside 1% and 5% limits. *** Outside 0.1%, 1%, and 5% limits.

It is interesting to test the significance of the serial correlation coefficients. Fisher's exact test[8] of significance, however, cannot be directly applied, since of course the elements entering into a serial correlation coefficient are not independent. But we can again make them independent by selection (Table 45).[9]

Take for instance the serial correlation coefficient $r'_{(1)}$ which refers to a lag $L = 1$. We make a selection 1-α by taking only the first, the third, the fifth, etc. pairs. This gives us a serial correlation coefficient for the selected items $r'_{(1)} = 0.7648$. If we make the other possible selections 1-β, (i.e., include only the pairs number 2, 4, 6, etc.) we get in our case of the annual wheat-flour prices a serial correlation coefficient for the selected items $r'_{(1)} = 0.9485$. Similarly we make selection 2-α by taking only the pairs 1, 4, 7, etc.; 2-β by taking the pairs 2, 5, 8, etc.; 2-γ by taking the pairs 3, 6, 9, etc. This gives the serial correlation coefficients for selected items $r'_{(2)}$ which are 0.6852, 0.7786, and 0.6991. All the serial correlation coefficients for selected and hence independent items are tabulated in Table 45.

We can now use Fisher's exact test, since our items are independent. The number of degrees of freedom is $n-2$ if n is the number of pairs in the particular selection (Table 45). The problem can be dealt with from different levels of significance. We have again in Table 45 designated by * a serial correlation coefficient, which is significant from the point of view of a level of significance of 5% but not from a point of view of levels of significance 1% or 0.1%; ** designates a serial correlation coefficient which is significant from the point of view of levels of significance 5% and 1% but not from the level of 0.1%. Finally, *** indicates that the serial correlation coefficient is significant from all three levels of significance.

To give an example, selection 2-α gives a serial correlation coefficient of 0.6852. We have in this particular selection $n = 16$ and hence the number of degrees of freedom equal to 14. Entering Table VI in Fisher and Yates tables[10] for 14 degrees of freedom we get a value of 0.4973 for 5%, 0.6226 for 1%, and 0.7420 for the 0.1% level of significance. Our empirical serial coefficient, 0.6852, is greater than the first two values but smaller than the third. We hence conclude that the serial correlation is significant for levels of 5% and 1% but not for 0.1%. The serial coefficient hence has ** in Table 45.

Glancing over Table 45 we see that hardly any of the serial correlation coefficients with $L = 3$ or higher are significant. Hence, we should conclude that the third and higher serial correlation coefficients are probably already zero or near to zero. This again agrees well with our contention that we have eliminated the nonrandom element in the third difference, as stated above (Chapter VII).

[8] R. A. Fisher, **Statistical Methods for Research Workers**, 7th ed., London, 1938, pp. 198 ff., Chapter VI, Section 34.

[9] G. Tintner, "On Tests of Significance in Time Series," **Annals of Mathematical Statistics**, Vol. 10, 1939, pp. 141 ff.

[10] R. A. Fisher and F. Yates, **Statistical Tables**, London, 1938, Table VI; see also R. A. Fisher, op. cit., Table V A.

APPENDIX VI

The Variability of the Random Variance Through Time

Dr. Zaycoff pointed out in an interesting article,[1] that we frequently find the random variance not constant in time. Hence the assumption that all elements of the random part of the series have the same variance is not fulfilled for long series. We have performed an experiment, which substantiates his contention.

Table 46 exhibits the standard deviations of the monthly wool prices and their differences for every year from 1890 to 1936. A glance at this table confirms the impression that the standard deviations are not constant in time. The years in which especially high values occur are the war years, 1917 to 1920, and also the years of great crises, 1903, 1923, 1932.

We show in Table 47 the correlation coefficients for correlations between the average annual prices, also given in Table 46, and the standard deviations of the monthly wool prices and their differences. All these correlation coefficients are too high not to be significant. We have 47 items in each series. Entering Fisher's table[2] for 45 degrees of freedom we see that the value of the correlation coefficient is 0.4648 for a level of significance 0.1%. All our correlation coefficients are higher than this.

If we leave out the years 1903, 1917-20, 1923, and 1932 from our calculations we have only 40 items left. The correlation coefficients between the annual wool prices and the standard deviations of the monthly wool prices and their differences, disregarding these exceptional years, are also given in Table 47. The levels of significance are now, from Fisher's table for 38 degrees of freedom: 0.4032 for a level of significance 1% and 0.5013 for a level of significance 0.1%. None of our correlation coefficients are significant from the point of view of the latter level of significance. But two of them are also not significant from the point of view of the level 1%. Hence we may conclude that there is probably not much correlation between the monthly variances and the annual prices if we disregard exceptional years.

This result fits in with our economic considerations as shown in Chapter I, Section C. We should expect that the random errors will increase in years of war or exceptional economic change. People will be more likely to make errors of the first kind since they have to adapt themselves to changing conditions. The effect of the nonpermanent causes in economic life can be expected to be greater in these years.

[1] R. Zaycoff, "Ueber die Ausschaltung der zufälligen Komponente nach der 'Variate-Difference' Methode," **Publications of the Statistical Institute for Economic Research**, State University of Sofia, 1937, No. 1, pp. 86 ff.

[2] R. A. Fisher, **Statistical Methods for Research Workers**, 7th ed., London, 1938, Table V-A; R. A. Fisher and F. Yates, **Statistical Tables**, London, 1938, Table VI, p. 36. See also G. W. Snedecor, **Statistical Methods**, Ames, Iowa, 1938, Table 7.2, p. 133.

TABLE 46

ANNUAL PRICES AND ANNUAL STANDARD DEVIATIONS, MONTHLY WOOL PRICES, 1890-1936
(k = Order of Difference)

Year	Annual Prices	Annual Standard Deviations										
		$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
1890	0.716	0.0582	0.0058	0.0049	0.0046	0.0045	0.0045	0.0044	0.0040	0.0036	0.0036	0.0034
1891	0.686	0.0308	0.0102	0.0079	0.0072	0.0067	0.0063	0.0061	0.0061	0.0062	0.0061	0.0060
1892	0.612	0.0271	0.0054	0.0044	0.0038	0.0031	0.0029	0.0032	0.0034	0.0034	0.0033	0.0035
1893	0.564	0.0563	0.0209	0.0119	0.0092	0.0082	0.0077	0.0070	0.0065	0.0063	0.0061	0.0058
1894	0.445	0.0220	0.0126	0.0093	0.0081	0.0069	0.0060	0.0055	0.0051	0.0051	0.0051	0.0055
1895	0.377	0.0141	0.0098	0.0084	0.0082	0.0080	0.0079	0.0079	0.0080	0.0080	0.0078	0.0074
1896	0.394	0.0171	0.0087	0.0055	0.0047	0.0042	0.0043	0.0044	0.0042	0.0041	0.0040	0.0041
1897	0.496	0.0683	0.0182	0.0109	0.0098	0.0094	0.0091	0.0088	0.0088	0.0088	0.0088	0.0088
1898	0.615	0.0334	0.0084	0.0065	0.0061	0.0062	0.0070	0.0080	0.0089	0.0095	0.0098	0.0100
1899	0.623	0.0721	0.0196	0.0128	0.0119	0.0115	0.0109	0.0100	0.0089	0.0079	0.0070	0.0065
1900	0.659	0.0763	0.0154	0.0083	0.0079	0.0078	0.0077	0.0075	0.0075	0.0076	0.0077	0.0077
1901	0.545	0.0091	0.0058	0.0037	0.0034	0.0033	0.0034	0.0037	0.0042	0.0048	0.0054	0.0060
1902	0.577	0.0330	0.0110	0.0082	0.0083	0.0085	0.0086	0.0087	0.0087	0.0085	0.0082	0.0078
1903	0.655	0.0257	0.0485	0.0402	0.0382	0.0374	0.0369	0.0365	0.0361	0.0358	0.0356	0.0354
1904	0.487	0.0178	0.0415	0.0318	0.0301	0.0295	0.0293	0.0292	0.0292	0.0292	0.0292	0.0292
1905	0.759	0.0192	0.0091	0.0058	0.0052	0.0049	0.0046	0.0044	0.0042	0.0041	0.0041	0.0042
1906	0.718	0.0100	0.0071	0.0058	0.0055	0.0054	0.0054	0.0055	0.0055	0.0055	0.0056	0.0057
1907	0.718	0.0153	0.0082	0.0067	0.0063	0.0061	0.0059	0.0057	0.0055	0.0053	0.0050	0.0046
1908	0.716	0.0163	0.0071	0.0058	0.0052	0.0049	0.0050	0.0051	0.0053	0.0054	0.0055	0.0054
1909	0.543	0.0240	0.0058	0.0053	0.0056	0.0058	0.0057	0.0056	0.0055	0.0054	0.0054	0.0054
1910	0.686	0.0208	0.0082	0.0075	0.0059	0.0056	0.0055	0.0057	0.0061	0.0066	0.0073	0.0082
1911	0.647	0.0499	0.0173	0.0131	0.0120	0.0117	0.0116	0.0116	0.0116	0.0115	0.0111	0.0105
1912	0.647	0.0412	0.0100	0.0088	0.0086	0.0085	0.0083	0.0082	0.0081	0.0080	0.0080	0.0080
1913	0.589	0.0359	0.0096	0.0061	0.0057	0.0054	0.0052	0.0050	0.0050	0.0049	0.0050	0.0052
1914	0.579	0.0161	0.0106	0.0089	0.0088	0.0097	0.0107	0.0112	0.0113	0.0112	0.0110	0.0108
1915	0.665	0.0210	0.0151	0.0115	0.0101	0.0082	0.0058	0.0039	0.0026	0.0020	0.0019	0.0019

TABLE 46 (concluded)

ANNUAL PRICES AND ANNUAL STANDARD DEVIATIONS, MONTHLY WOOL PRICES, 1890-1936
(k = Order of Difference)

Year	Annual Prices	Annual Standard Deviations										
		$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
1916	0.775	0.0602	0.0149	0.0101	0.0094	0.0092	0.0090	0.0087	0.0085	0.0084	0.0082	0.0079
1917	1.164	0.1926	0.0470	0.0245	0.0216	0.0200	0.0190	0.0183	0.0178	0.0173	0.0168	0.0163
1918	1.440	0.0990	0.0498	0.0330	0.0298	0.0312	0.0105	0.0336	0.0322	0.0302	0.0283	0.0266
1919	1.189	0.0936	0.1797	0.1439	0.1350	0.1292	0.1249	0.1221	0.1208	0.1221	0.1252	0.1296
1920	1.673	0.4733	0.1400	0.0986	0.0906	0.0850	0.0804	0.0769	0.0734	0.0680	0.0592	0.0482
1921	0.791	0.0620	0.0479	0.0316	0.0298	0.0305	0.0318	0.0329	0.0339	0.0352	0.0367	0.0386
1922	1.219	0.1086	0.0385	0.0309	0.0313	0.0316	0.0319	0.0322	0.0324	0.0322	0.0316	0.0305
1923	1.376	0.0672	0.1670	0.1393	0.1338	0.1308	0.1288	0.1273	0.1261	0.1249	0.1237	0.1229
1924	0.489	0.0445	0.0197	0.0144	0.0134	0.0136	0.0140	0.0146	0.0154	0.0161	0.0519	0.0163
1925	0.482	0.0528	0.0210	0.0145	0.0140	0.0139	0.0135	0.0129	0.0119	0.0106	0.0093	0.0082
1926	0.402	0.0108	0.0094	0.0062	0.0051	0.0045	0.0044	0.0044	0.0044	0.0044	0.0042	0.0041
1927	0.386	0.0112	0.0050	0.0039	0.0035	0.0032	0.0031	0.0030	0.0030	0.0030	0.0029	0.0028
1928	0.389	0.0153	0.0046	0.0037	0.0034	0.0035	0.0036	0.0036	0.0036	0.0036	0.0035	0.0035
1929	0.327	0.0240	0.0087	0.0042	0.0036	0.0030	0.0026	0.0022	0.0019	0.0017	0.0016	0.0015
1930	0.246	0.0104	0.0050	0.0037	0.0034	0.0032	0.0031	0.0031	0.0030	0.0030	0.0030	0.0031
1931	0.205	0.0104	0.0065	0.0047	0.0039	0.0034	0.0030	0.0028	0.0029	0.0032	0.0036	0.0041
1932	0.166	0.0231	0.0156	0.0160	0.0164	0.0164	0.0166	0.0169	0.0172	0.0175	0.0177	0.0179
1933	0.233	0.0729	0.0186	0.0096	0.0068	0.0063	0.0061	0.0060	0.0059	0.0058	0.0058	0.0018
1934	0.268	0.0268	0.0076	0.0056	0.0057	0.0058	0.0059	0.0059	0.0060	0.0060	0.0062	0.0064
1935	0.250	0.0266	0.0108	0.0065	0.0054	0.0049	0.0048	0.0047	0.0046	0.0044	0.0042	0.0041
1936	0.318	0.0278	0.0140	0.0090	0.0064	0.0056	0.0055	0.0057	0.0060	0.0061	0.0062	0.0061

TABLE 47

CORRELATION COEFFICIENTS OF ANNUAL STANDARD DEVIATIONS OF MONTHLY WOOL PRICES AND DIFFERENCES, AND ORIGINAL ANNUAL WOOL PRICES

Order of Difference k	All Items 1890-1936	All Items except 1903, 1917-1920, 1923, 1932	Order of Difference k	All Items 1890-1936	All Items except 1903, 1917-1920, 1923, 1932
0	0.6875	0.4634	6	0.6844	0.4772
1	0.7159	0.3744	7	0.6775	0.4724
2	0.6850	0.4362	8	0.6640	0.4649
3	0.6825	0.4739	9	0.6168	0.3161
4	0.6849	0.4836	10	0.6150	0.4636
5	0.6434	0.4816			

APPENDIX VII

THE NORMALITY OF THE RANDOM ELEMENT

It is interesting to investigate the normality of the distribution of the random element in the price series we have considered.[1] We should from a priori economic reasons (Chapter I, Section C) expect a more or less normal or at least symmetrical distribution.

We show in Table 48 the *k*-statistics, the skewness (g_1), and kurtosis[2] (g_2) of the original data and differences of the annual wheat-flour prices, 1890 to 1937. They are calculated from the sums given in Table 8. We also exhibit the standard error of the skewness (e_1') and of the kurtosis (e_2'). It appears that the skewness is rather insignificant in the higher differences. The kurtosis on the other hand is significant and stays so even for high differences. Since it is positive we have a distribution that is more peaked than a normal one.

The same impression is also gained from Figure 5 which shows the distributions of the original monthly wool prices, 1890-1937, as well as the distributions of the fifth and tenth differences of the same price series. We have also exhibited in these graphs the corresponding normal distributions. All the distributions show kurtosis, the peaks being much higher than in a normal distribution with the same variance. The original prices also have considerable skewness, probably mainly due to the secular trend, which is already more or less eliminated in the higher differences.

1 R. Zaycoff, "Ueber die Ausschaltung der zufälligen Komponente nach der 'Variate-Difference' Methode," **Publications of the Statistical Institute for Economic Research**, State University of Sofia, 1937, No. 1, pp. 109 ff.

2 R. A. Fisher, **Statistical Methods for Research Workers**, 7th ed., 1938, p. 78 f., Chapter III, Appendix; G. W. Snedecor, **Statistical Methods**, Ames, Iowa, 1938, pp. 147 ff. The *k*-statistics seem to be related to the semi-invariants of Thiele. See A. Fisher, **The Mathematical Theory of Probability**, 2nd ed., New York, 1930, pp. 191 ff.

TABLE 48

SKEWNESS AND KURTOSIS OF ANNUAL AMERICAN WHEAT-FLOUR PRICES, 1890-1937

Order of Difference	k-statistics				Skewness	Kurtosis	Standard Error of Skewness	Standard Error of Kurtosis
k	k_1	k_2	k_3	k_4	g_1	g_2	e_1'	e_2'
0	6.0722	4.8034	14.1358	42.6961	1.3428	1.8501	0.3432	0.6744
1	0.0326	1.4217	—0.1622	10.7509	—0.09568	5.3190	0.3466	0.6809
2	—0.0052	2.6759	—4.0131	24.4853	—0.9168	3.4195	0.3501	0.6876
3	0.0400	7.9692	8.5540	240.6643	0.3802	3.7895	0.3537	0.6946
4	0.0174	26.7160	34.0001	2,702.9546	0.2462	3.7870	0.3575	0.7016
5	0.1491	93.5263	—298.2719	33,780.3522	—0.3298	3.8619	0.3614	0.7090
6	0.0517	332.7998	—776.2496	432,348.5131	—0.1279	3.9036	0.3654	0.7167
7	0.3843	1,217.0490	5,132.7523	5,501,485.7272	0.1209	3.7142	0.3695	0.7245
8	0.2645	4,459.4503	17,467.4201	90,564,246.7731	0.0587	4.5540	0.3738	0.7326
9	1.5008	16,429.3353	—228,274.0717	957,632,794.5676	—0.1084	3.5478	0.3783	0.7410
10	1.0357	60,865.8319	—697,449.7349	12,779,691,619.9162	—0.0464	3.4496	0.3828	0.7497

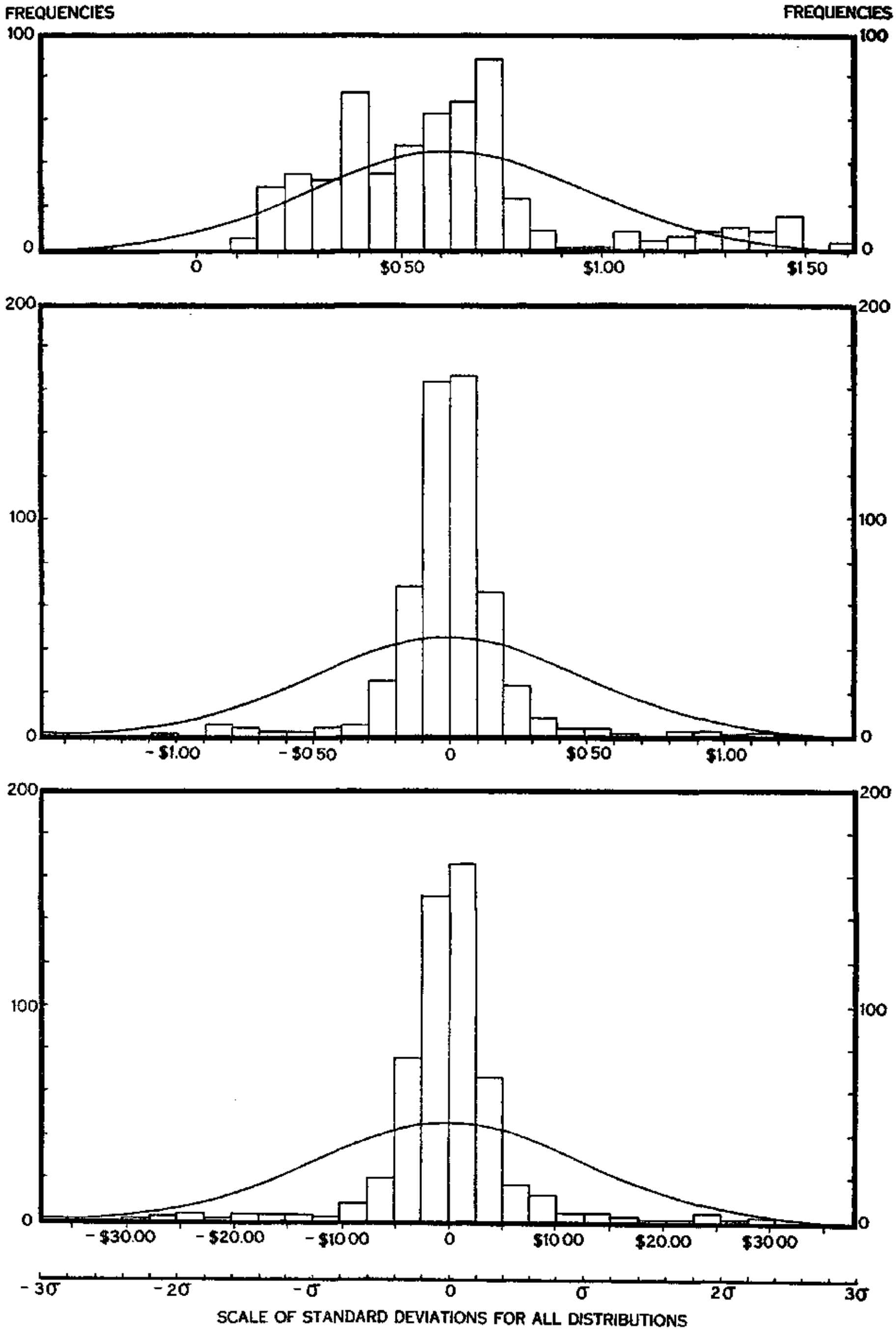

FIGURE 5.—FREQUENCY DISTRIBUTIONS OF MONTHLY WOOL PRICES, 1890-1937: ORIGINAL, FIFTH DIFFERENCES, AND TENTH DIFFERENCES

INDEXES

INDEX OF AUTHORS

INDEX OF SUBJECTS