LINEAR REGRESSION ANALYSIS OF ECONOMIC TIME SERIES

T. KOOPMANS

LINEAR REGRESSION ANALYSIS OF ECONOMIC TIME SERIES

LINEAR REGRESSION ANALYSIS OF ECONOMIC TIME SERIES

PROEFSCHRIFT TER VERKRIJGING VAN DEN GRAAD VAN DOCTOR IN DE WIS- EN NATUURKUNDE AAN DE RIJKSUNIVERSITEIT TE LEIDEN, OP GEZAG VAN DEN RECTOR MAGNIFICUS DR. J. VAN DER HOEVE, HOOGLEERAAR IN DE FACULTEIT DER GENEESKUNDE, VOOR DE FACULTEIT DER WIS- EN NATUURKUNDE TE VERDEDIGEN OP DONDERDAG 26 NOVEMBER 1936, DES NAMIDDAGS TE 4 UUR

DOOR

TJALLING KOOPMANS

GEBOREN TE 'S-GRAVELAND

DE ERVEN F. BOHN N.V. — HAARLEM — 1936

De voltooiïng van dit proefschrift biedt mij een welkome gelegenheid mijn dank uit te spreken aan U, Hoogleraren in Wiskunde en Natuurkunde aan de Rijksuniversiteit te Utrecht, voor de wetenschappelijke opleiding, die ik van U ontving.

Hooggeleerde K r a m e r s, Hooggeachte Promotor, het viel mij zwaar de beoefening van die wetenschap, waarin gij mij hadt opgeleid, te laten varen. Te meer ben ik getroffen door de bijzondere belangstelling en de daadwerkelijke bijstand, van U ondervonden bij de bewerking van dit proefschrift.

Hooggeleerde T i n b e r g e n, U dank ik zeer voor de hulp, door U, ondanks Uw drukke bezigheden, aan mij gegeven bij mijn studie van de econometrie, en voor de leerzame besprekingen, die ik met U had over de vraagstellingen van dit onderzoek.

I am greatly indebted to you, Professor F r i s c h, for many fruitful discussions on the subject matter of this study.

My cordial thanks are due to my friend R. S a r d for having devoted great care to the correction of the language.

Tenslotte een woord van oprechte dank aan U, Heren Provisoren, Collatoren en Bestuurders van het Sint Geertruidsleen te Abbega, voor het vertrouwen, door mijn benoeming als beneficiant dezer instelling in mij gesteld, en voor de buitengewone bereidwilligheid, waarmee gij aan al mijn wensen aangaande mijn studie zijt tegemoetgekomen.

CONTENTS

NOTATIONS

Greek characters denote known constants and unknown parameters.

Latin characters denote quantities subject to a probability distribution, frequently estimates of the parameters denoted by the corresponding Greek characters.

The *subscripts* $k, l \ldots$ can assume all values $1, 2 \ldots K$,
" " $k', l' \ldots$ " " " " $2, 3 \ldots K$,
" " $t, s \ldots$ " " " " $1, 2 \ldots T$, etc.

A *Greek subscript* occurring at least twice in the same product should be summed over all values assumed by the corresponding Latin subscript.

It need not cause confusion that the same character may be used at once as a subscript and as denoting a quantity.

$\mathbf{c}$, c_k	regression vector, set of regression coefficients
a_k	orthogonal }
b_k	elementary } regression coefficients
c_k	weighted } regression coefficients
d_k	diagonal }
$\mathbf{X}$, X_k	point, values assumed by a set of variables
$\mathbf{cX}$	$c_\varkappa X_\varkappa$
$\mathbf{m}$, $\|\|m_{kl}\|\|$	matrix with elements m_{kl}
$\|\mathbf{m}\|$, $\|m_{kl}\|$, M	the determinant value of $\mathbf{m}$
$\mathbf{m}^{-1}$, $\|\|m^{kl}\|\|$	the matrix inverse to $\mathbf{m}$
$\mathbf{cmc}$	$c_\varkappa m_{\varkappa\lambda} c_\lambda$
$\boldsymbol{\delta}$	unit matrix: $\delta_{kl} = 1$ if $k = l$, 0 if $k \neq l$
$\equiv$	is by definition equal to
$\geqslant$	is large compared with
$\leqslant$	" small " "
E	mathematical expectation of
$\exp z$	e^z
$\operatorname{sgn} r$	sign of r
est σ	estimate of σ
(5)	formula (5) of the section in which it is quoted
(3.32)	formula (32) of section 3
(16)	number 16 in the list of references

form, the object of statistical methods is the reduction of data. A quantity of data, which usually by its mere bulk is incapable of entering the mind, is to be replaced by relatively few quantities which shall adequately represent the whole, or which, in other words, shall contain as much as possible, ideally the whole, of the relevant information contained in the original data.

This object is accomplished by constructing a hypothetical infinite population, of which the actual data are regarded as constituting a random sample. The law of distribution of this hypothetical population is specified by relatively few parameters, which are sufficient to describe it exhaustively in respect of all qualities under discussion. Any information given by the sample, which is of use in estimating the values of these parameters, is relevant information. Since the number of independent facts supplied in the data is usually far greater than the number of facts sought, much of the information supplied by any actual sample is irrelevant. It is the object of the statistical processes employed in the reduction of data to exclude this irrelevant information, and to isolate the whole of the relevant information contained in the data."

Thus there is constructed a *"hypothetical infinite population"* or *"parent distribution"* or *"universe"*, that is, a probability distribution of the observational variable(s) of given mathematical form, however containing one or more unknown *parameters*. The data are considered as a random sample drawn from this distribution. The problems which arise in reduction of data are divided by Fisher (**5**, p. 313, see also **10**, p. 8) into the following three types:

"(1) Problems of *specification*. These arise in the choice of the mathematical form of the population.

(2) Problems of *estimation*. These involve the choice of methods of calculating from a sample statistical derivatives, or as we shall call them *statistics*, which are designed to *estimate* the values of the parameters of the hypothetical population.

(3) Problems of *distribution*. These include discussions of the distribution of statistics derived from samples, or in general any functions of quantities whose distribution is known.

> It will be clear that when we know (1) what parameters are required to specify the population from which the sample is drawn, (2) how best to calculate from the sample estimates of these parameters, and (3) the exact form of the distribution, in different samples, of our derived statistics, then the theoretical aspect of the treatment of any particular body of data has been completely elucidated."

These concepts are adopted as the theoretical basis of the developments in this paper. The reliability of the statistics as estimates of the parameters of the parent distribution will be judged from a theoretical study of the distribution of these statistics "in repeated samples". Nevertheless, the special requirements of the present field of application will necessitate a treatment deviating in important features from existing regression theory based on these same concepts.

This theory has been widely applied to data obtained from agricultural experiment or from measurements in biological populations. There are some essential differences between data of this kind and those usually encountered in economic problems. In agricultural experiment some of the determining variables can be completely controlled by the experimenter (for instance manurial treatment). In the design of his experiment he can secure as much independent variation of each of these variables as is needed for a reliable estimate of the corresponding regression coefficients. Other determining variables less under his control (rainfall, temperature, etc.) are usually by their nature subject to adequate independent variation. In that respect they bear a resemblance to the variables representing measurable characteristics of individuals of a biological population, which are usually conceived as random drawings from a stable probability distribution, eventually the multivariate normal distribution. In investigations of this latter type regression coefficients are sometimes not used as quantitative measures of a causal relationship between one "dependent" variable and its "determining" variables, but rather as quantities descriptive of a joint distribution of a set of variables without specifying one-way causal connections between them. In both cases the data are considered as manifestations of some underlying stable natural law. If they are not sufficiently numerous to yield the desired information, more observations can be obtained.

In economic analysis variables at the control of an experimenting institution are exceptional. Further only a few types of variables, mainly those directly connected with crop yields, are so erratic in nature, that they could reasonably be regarded as drawings from any stable distribution. In a great deal of the problems variables are developing in time in cyclical oscillations, apparently to a large extent governed by some internal causal mechanism, and only besides that influenced, more or less, according to the nature of the variable, by erratic shocks due to technical inventions, variations in crop yields, etc. [1]). At any rate, they are far from being random drawings from any distribution whatever. It is for that reason that a great deal of the extensive work now available on sampling from a multivariate normal distribution, as reviewed by Rider (**24**), may find only very limited application in the analysis of economic time series.

Further the relations between the variables studied in this type of analysis are themselves subject to gradual or abrupt changes, according to institutional or technical changes in society (and the same holds for the causal mechanism referred to above). Therefore the number T of observations from which the regression coefficients have to be estimated is limited by the very nature of the problem.

These considerations have led some authors to doubt the possibility of fruitfully applying "sampling theory" to the interpretation of economic time series. In the sections 4 and 6 statements made by Frisch, E. J. Working and Bartlett, which more or less explicitly express such doubts will be quoted and discussed. In my opinion, there can be assigned a well defined task to sampling theory in economic regression analysis. The point is that, mathematically, we can advance even if we assume a probability distribution only for the accidental errors in the variables, which prevent the regression equation from being exactly satisfied by the observations. Following Frisch (**16**, p. 51), each of the variables may be conceived as the sum of two components, a *"systematic component"* or *"true value"* and an *"erratic component"* or *"disturbance"* or *"accidental error"*. The systematic components are assumed to satisfy the regression equation exactly. In the determining

[1]) For this concept of economic variables see for instance Frisch (**15**) and Tinbergen (**32**).

variables the erratic component is taken as an error in the literal sense of the word. It can be due to inadequate weighting of an index number, or to inaccuracy in the statistical recording procedure; it can arise if, for lack of statistical data, a determining variable is represented by the values of some other variable that is known to be highly correlated with the former (for instance an index of industrial or total production as a representative for total real income). In the dependent variable, however, the erratic component will, besides this "technical" error, contain an additional component representing the influence exerted on X_1 by determining variables of minor importance, not included in the set $X_2 \ldots . X_K$. These distinctions enable us to give a more clear-cut sense to the notion of a complete set of determining variables introduced above. The set $X_2 \ldots . X_K$ will be called complete if the combined influence on X_1 of all other variables, not included in the set, may be represented by a relatively "small" summand in X_1 of accidental nature. If, in any concrete situation, a standard is established for what should be understood by a "small summand of accidental nature", the notion of a complete set of determining variables is fixed by that standard.

Above the distinction has been drawn between a regression coefficient conceived as a quantitative measure of a causal relationship and a regression coefficient conceived as a quantity descriptive for a multivariate distribution. It will be clear from the last paragraph that here the former sense is adopted (see also Frisch **13**, p. 95). It is assumed that there is a "true regression equation" which would be exactly satisfied by the "true values" of the variables. This should not be taken as a matter of principle. In fact, in numerous applications in which at least some of the variables are index numbers it would be rather difficult, if not impossible, to maintain this simplified picture with all its consequences. Nevertheless, this simplification may prove useful, as in many cases the "errors" in the empirical regression coefficients which are studied in this investigation using this simplification will be of dominant quantitative importance compared with eventual corrections that may be obtained by a refinement of the conceptions taking into account the fact that in reality index numbers are only some kind of mean values of distributions.

This conceptual scheme of the structure of the variables being

accepted, it is clear that the expression *"repeated sampling"* requires an interpretation somewhat different from that prevailing in applications of sampling theory in the agricultural and biological field. In the latter domains in many cases repeated samples may in principle be obtained in any number. The distribution of a statistic "in repeated samples" is therefore something which, if only sufficiently extensive efforts are applied, may be found or controlled by experience. But in the conditions under which sampling theory is used here such a distribution is much more hypothetical in nature. The observations (1) constituting one sample, a repeated sample consists of a set of values which the variables would have assumed if in these years the systematic components had been the same and the erratic components had been other independent random drawings from the distribution they are supposed to have.

The meaning of the sampling distribution of a statistic is closely bound up with the plausibility of the hypotheses specifying the parent distribution from which samples are imagined to be drawn. In this paper these hypotheses are embodied in the decomposition of the variables into systematic and erratic components, in the supposed existence of a linear relation between the systematic components, and in the distribution assumed for the erratic components. These hypotheses are in concrete economic applications much less liable to empirical verification than are the hypotheses underlying the use of sampling theory in those domains of science in which this theory has by now found widespread recognition and application. It is only the urgent need — brought about by the recent development of econometrics — for a basis for appreciation of the reliability of empirical regression equations fitted to economic time series, which might supply the justification for a procedure based to such an extent on hypothetical foundations. Therefore, any measure of the reliability of regression coefficients, reached by this procedure, will hardly have the precision attachable to the results of sampling theory in many other situations. In view of the need for such a measure, however, it may be better to have some point of support, obtained by the use of a set of simplifying assumptions, than none at all.

Designing a set of assumptions for this purpose requires thorough consideration. Part II deals with some lines of approach

followed by three different authors on regression theory. The sets of assumptions underlying these lines of thought will be discussed with regard to their adequacy to the special requirements of the analysis of economic time series. Some general remarks on the choice of basic assumptions may be made here.

If the methods of the sampling theory are applied, this choice takes the form of a specification of the parent distribution of which the observations are considered as a sample. Though parameters of various nature may be introduced in this parent distribution, attention is often concentrated upon the estimation of parameters of a special kind — in our case the regression coefficients. For brevity such parameters will be referred to as the *"required parameters"*, though, properly speaking, reliable *estimates* of these parameters are required.

I am aware of the following rules by which the choice of a specification of the parent distribution which is adequate to the special nature of the problem dealt with should be guided. They are of a different nature, and even in part conflicting.

I. *In the specification should be implied all a priori knowledge on the nature of the data which is relevant to the estimation of the required parameters.*

For instance, in sufficiently high dilutions of a fluid containing micro-organisms the number of these organisms in samples of a given volume may on a priori grounds be assumed to be distributed according to Poisson's law. Further in many cases the process by which observations are obtained warrants independence between successive values of variables.

II. *In specifying the parent distribution such a posteriori information as to its form should be worked up as may be extracted from the sample itself with a reasonable degree of reliability and is relevant to the estimation of the required parameters.*

Thus, in estimating the mean of a univariate parent distribution, evidence as to the normality may be obtained from a sufficiently large sample. And in general, as Fisher remarks (**5**, p. 314), K. Pearson's χ^2-test of goodness of fit (**21**) may supply a powerful tool to check from a sufficiently large sample the adequacy of any assumed specification.

The restriction that only information relevant to the required estimation procedure should be implied deserves particular attention in connection with the next rule:

III. *Extension of the specification by the introduction of additional assumptions not imposed by the rules* I *and* II *should be avoided as much as possible.*

For any extension of this kind necessarily implies a limitation of the range of applicability of the results deduced by means of the specification. An example is Bartlett's remark (**2**, p. 542) that the sampling distribution of the coefficient of correlation between two normal variables, independent mutually as well as in successive drawings, which serves as the basis for a test of significance of correlation, may also be derived if only one of the variables is assumed to be so distributed, the other having any distribution, possibly showing serial correlation, but independent of that of the first variable.

The qualification "as much as possible" has been added to rule III because it is often difficult to bring this rule into accordance with the exigencies of the following rule which is of very practical nature.

IV. *The specified form of the parent distribution should be neither so general nor so complicated as to make mathematical treatment of the problems of estimation and distribution too cumbersome and intricate or even impossible.*

In cases of apparent conflict between the two last rules there may be several ways out. One of them is that the development of mathematical tools may open the way for a treatment of problems which had hitherto seemed insoluble. In estimating the mean of a normal univariate distribution from a large sample, the variance of this distribution may according to rule II be taken as estimated from the sample. In small samples, where the reliability of this procedure is very low, the way out was opened by the introduction of Student's ratio (**28**) and the "*t*-test" (**8**) based on the sampling distribution of this ratio.

If no such improvement of mathematical tools seems possible, a line of compromise may be followed:

V. *In so far as the specification extends beyond the elements supplied by the rules* I, II *and* III, *the accuracy of estimation of the required parameters should not be very sensitive to an inexact fulfilment of the additional assumptions.*

Thus, a series of sampling experiments by E. S. Pearson and others (**19**) has shown that the distribution of Student's ratio in samples of moderate size is not much affected by con-

siderable departures from normality in the parent distribution. By similar experiments (**20**) the above mentioned test of absence of correlation was found to be approximately valid for samples from a wider range of non-normal distributions.

These rules will be referred to as the *specification rules*. Perhaps, in some applications of sampling theory, more or other requirements have to be imposed on a specification of the parent distribution, adequate to the problem under consideration. The set of rules given here will serve as a basis for discussing the specification problem with regard to the application of sampling theory to regression analysis of economic time series.

2. *Mathematical tools.*

Frequent use will be made in this investigation of the elementary theory of *determinants* and of their relations to *systems of linear equations*. Further we shall need the first elements of the theory of *matrices, linear and quadratic forms*, and their behaviour under *orthogonal* and other *linear transformations*.

No more is supposed to be known than may be found in nearly every elementary textbook on these subjects. A useful résumé of the elements of matrix theory in so far as it is related with statistical applications can be found in an extremely instructive paper of Frisch (**13**).

An appeal to geometric imagination is made by the extensive use of the presentation of the variables (1.1) in K-dimensional space, one rectangular coordinate corresponding to each of the variables. The word *vector* and the notation $\mathbf{c}$ are sometimes used for the set of coefficients c_k in a linear regression equation [1])

$$c_\varkappa x_\varkappa - c = 0 \tag{1}$$

between these variables. In that case the set of coefficients c_k is thought to be represented by a vector connecting the origin O with the point with coordinates $c_1, c_2 \ldots . c_K$, thus having a direction perpendicular to the plane (1). This perpendicularity is by definition not affected by a change

$$X_k^* = \frac{X_k}{\theta_k}$$

[1]) Summation should everywhere be made over any Greek index occurring at least twice in the same product.

in the units of measurement of the variables because in such a change the c_k must transform according to

$$c_k^* = \theta_k c_k$$

if (1), written in the asterisk quantities, is again to represent the same relation. This almost trivial remark is only made in order to prevent confusion with another kind of perpendicularity, to be considered in section 4, which is in general not preserved under a change in the units.

A central place is taken by the *moment matrix*

(2) $$m_{kl} = m_{lk} \equiv (X_k^{(\tau)} - \overline{X}_k)\,(X_l^{(\tau)} - \overline{X}_l) = (X_k^{(\tau)} - \overline{X}_k)\,X_l^{(\tau)}$$

of the variables (1.1) (where

(3) $$\overline{X}_k \equiv \frac{1}{T}\,\Sigma_\tau X_k^{(\tau)}$$

is the mean of the k-th variable). It is *positive definite*, that is, it satisfies

(4) $$c_\varkappa m_{\varkappa\lambda} c_\lambda \geq 0$$

whatever the direction of the non vanishing vector $\mathbf{c}$, since this quadratic form can, in consequence of (2), be written as the sum of T squares

(5) $$[c_\varkappa\,(X_\varkappa^{(t)} - \overline{X}_\varkappa)]^2, \quad t = 1, 2 \ldots . T.$$

If T exceeds K, (as will be supposed in what follows), for every non vanishing vector $\mathbf{c}$ at least one of these squares will be positive, and the inequality sign in (4) will hold throughout, whence $\mathbf{m}$ will have the rank K — unless the points (1.1) happen to lie in some hyperplane of $K - 1$ or less dimensions.

Quadratic forms as that in (4) will sometimes be written in the abbreviate notation $\mathbf{cmc}$. Similarly $\mathbf{cx}$ will stand for $c_\varkappa x_\varkappa$. A more complete suppression of subscripts would have the disadvantage of suggesting a homogeneity between the several variables which does not exist. In particular, the abbreviate notation will never be used for expressions which are not invariant for the above mentioned change in the units of measurement.

For a positive definite matrix $\mathbf{m}$ there exists a number of inequalities between the determinant

(6) $$M \equiv |\,\mathbf{m}\,| \equiv |\,m_{kl}\,|$$

and its *principal minors* of any order. Of these we only mention the inequalities

$$M \leq m_{kk} M_{kk} \tag{7}$$

(where M_{kk} represents the cofactor of m_{kk} in $|\mathbf{m}|$) and the corresponding inequalities for the principal minors. As a consequence of these inequalities

$$M \leq m_{11} m_{22} \ldots . m_{KK} \tag{8}$$

where the equality sign can be shown to hold only if $\mathbf{m}$ has the *diagonal form*, that is, if

$$m_{kl} = 0 \text{ for } k \neq l.$$

On several occasions the argument will deal with the relations between the different degrees of approximate linear dependence in the swarm of points (1.1) and the relative magnitudes of M and its principal minors. For a full discussion of this point we may refer to Frisch (**13**, **14**, **16**). Here we shall only indicate a geometric picture which may help the reader in grasping the essentials of the situation and which gives a geometric meaning to the inequalities (7) and (8).

In the space representation used throughout in this investigation, the variables (1.1) are represented by a swarm of T points in a space of K dimensions. They can, however, also be taken as representing K points or vectors in a space of T dimensions, one coordinate corresponding to every year, one vector to every variable. Taking for simplicity

$$\overline{X}_k = 0, \tag{9}$$

the square moment m_{kk} of the k-th variable then equals the squared length of the k-th vector, the cross moment m_{kl} equals the inner product of the k-th and the l-th vector, and the determinant (6) measures the square of the volume (in units of K-dimensional space) of the generalized parallelepiped constructed on these vectors as edges, while similar relations hold for the principal minors.

This picture suggests that, if

$$M \ll m_{11} m_{22} \ldots . m_{KK}, \tag{10}$$

that is, if the volume of the above parallelepiped is small compared with that of a completely rectangular parallelepiped with edges of the same lengths, the K vectors will be nearly linearly

dependent, or the variables (1.1) will approximately satisfy at least one linear relation. An indication as to whether only one or even two or more linear relations (with considerably diverging coefficients!) are approximately satisfied by the variables (1.1) is given by the principal minors M_{kk} of order $K-1$. If at least one of these minors decidedly does not satisfy a relation similar to (10), the approximate linear dependence disappears if the corresponding variable is omitted from the set of variables, whence only one linear relation can be approximately satisfied by the complete set of variables. In the opposite case, the swarm of scatter points in the K-dimensional representation is flattened in at least two directions, and so on.

A considerable part of the subsequent sections is based on the classical solution of the problem of the extrema of the quadratic form (with symmetric matrix)

$$\mathbf{\alpha m \alpha} \equiv \alpha_\varkappa m_{\varkappa\lambda} \alpha_\lambda \tag{11}$$

for vectors $\boldsymbol{\alpha}$ restricted by

$$\alpha_\varkappa \alpha_\varkappa = 1. \tag{12}$$

In general there are K such extrema $m_1, m_2 \ldots . m_K$, the *characteristic values* of $\mathbf{m}$, which are obtained as roots of the algebraic equation of degree K

$$|\mathbf{m} - m\boldsymbol{\delta}| = \begin{vmatrix} m_{11}-m & m_{12} & \ldots\ldots & m_{1K} \\ m_{21} & m_{22}-m & \ldots\ldots & m_{2K} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ m_{K1} & m_{K2} & \ldots\ldots & m_{KK}-m \end{vmatrix} = 0, \tag{13}$$

$\boldsymbol{\delta}$ representing the unit matrix

$$\begin{cases} \delta_{kl} = 1 \text{ if } k = l, \\ \quad = 0 \text{ if } k \neq l. \end{cases} \tag{14}$$

To every single root m_n corresponds one *characteristic vector* $\mathbf{a}^{(n)}$ satisfying (12), uniquely defined by the K equations

$$(m_{k\lambda} - m_n \delta_{k\lambda})\, a_\lambda^{(n)} = 0, \tag{15}$$

and for which (11) assumes its extremum m_n. Any two such characteristic vectors are mutually perpendicular.

To a p-fold root of (13) corresponds a p-dimensional subspace of characteristic vectors, in which an arbitrary set of p mutually perpendicular vectors $\mathbf{a}^{(n)}$ can be chosen. Thus, whether mul-

tiple roots are present or not, there is at least one set of characteristic vectors satisfying

$$a_\varkappa^{(m)} a_\varkappa^{(n)} = \delta^{(mn)}. \tag{16}$$

By the orthogonal transformation which makes the new coordinate axes coincide with these vectors, $\mathbf{m}$ assumes the diagonal form:

$$\left\| \begin{matrix} m_1 & 0 & \dots & 0 \\ 0 & m_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & m_K \end{matrix} \right\|. \tag{17}$$

As a consequence of this, m_{kl} can be written

$$m_{kl} = a_k^{(\nu)} m_\nu a_l^{(\nu)}. \tag{18}$$

If the characteristic values are numbered in order of increasing size, m_1 is the absolute minimum of (11) under the restriction (12), m_K the absolute maximum. The remaining extremes have a saddle point character. If $m_1 \geq 0$, the form (11) is positive definite.

If $\mathbf{m}^{-1} \equiv ||m^{kl}||$ is the *inverse matrix* of $\mathbf{m}$, its characteristic vectors coincide or may be chosen coincident with those of $\mathbf{m}$, while its characteristic values are inverse to those of $\mathbf{m}$:

$$m^{kl} = a_k^{(\nu)} m_\nu^{-1} a_l^{(\nu)}. \tag{19}$$

The form of the swarm of scatter points (1.1) is roughly indicated by its *ellipsoid of inertia*

$$\mathbf{x}\mathbf{m}^{-1}\mathbf{x} \equiv x_\varkappa m^{\varkappa\lambda} x_\lambda = 1 \tag{20}$$

the axes of which fall alongside the characteristic vectors of $\mathbf{m}$ and have lengths $m_k^{\frac{1}{2}}$. This picture, though, may be misleading because it is not independent of the units of measurement of the variables.

The above theorems are easily extended to the case in which (12) is replaced by the condition

$$\boldsymbol{\alpha}\boldsymbol{\epsilon}\boldsymbol{\alpha} = 1, \tag{21}$$

where $\boldsymbol{\epsilon}$ is a non singular positive definite matrix. Instead of (13) and (15), then, the equations

$$|\, \mathbf{m} - l\boldsymbol{\epsilon} \,| = 0, \tag{22}$$

$$(m_{k\lambda} - l_n \varepsilon_{k\lambda})\, c_\lambda^{(n)} = 0, \tag{23}$$

PART I

Introduction

1. *The problem.*

This publication is an attempt to deal with some of the difficulties encountered in the statistical procedure of fitting a linear regression equation to a set of time series representing the values assumed by some related variables.

In general and, therefore, vague terms the problem may be formulated like this. For each of K variables

$$X_1,\ X_2 \ldots . X_K,$$

there is given a series of T observations

$$(1) \qquad X_k^{(t)},\ k = 1, 2 \ldots . K,\ t = 1, 2 \ldots . T,$$

relating to T successive time periods of equal length. For brevity these periods will be called "years", though, of course, they may as well be months, weeks, or any other period. On "a priori" grounds it is expected that in the period of years considered the values of one of these variables, say X_1, are determined, but for small discrepancies of accidental nature, by the values of the remaining variables. For that reason, X_1 may be referred to as the *dependent variable*, $X_2 \ldots . X_K$ as a *complete set of determining variables*[1]) of X_1. Moreover, the *"regression equation"* expressing X_1 by means of its determining variables is assumed to be linear. This restriction is only introduced since it may be convenient first to treat the simpler problems before the more complicated ones are raised. The "a priori" grounds mentioned above may be supplied by general experience or deductive reasoning from such experience or by

[1]) The expression "independent variables" has been avoided because the discussion of the problem in the subsequent pages is just to a large extent concerned with the difficulties that arise in cases where the variables $X_2 \ldots . X_K$ fail to be independent.

any other source of information in the domain of science to which the variables belong; the expectation that a significant relation exists between the variables $X_1 \ldots . X_K$ should, however, not be exclusively derived from the observations (1). The reasons for, and the importance of, this reserve will become clear in the discussions in sections 3 and 6. Now our problem is to find from the observations (1) estimates of the coefficients of the expected relation and to obtain an idea of the reliability of these estimates.

Only applications to economic data are discussed. If the results of this paper would prove to have some value in other fields I should consider that as an unexpected gain. For even within the field of economic applications of regression analysis nearly every situation imposes its own requirements on the method of treatment. The idea of a technique which, "like a stone of the wise, solves all the problems of testing "significance" with which the economic statistician is confronted" is rejected by F r i s c h (**16**, p. 192) in the significant words:

> "No statistical technique, however refined, will ever be able to do such a thing. The ultimate test of significance must consist in a network of conclusions and cross checks where theoretical economic considerations, intimate and realistic knowledge of the data and a refined statistical technique concur."

In view of this statement any discussion which focusses attention on the statistical procedure as such is bound to suffer from an inevitable degree of abstraction and schematism. As an attempt to compensate for this remoteness from concrete situations, illustrations of the results in some practical examples will be given.

The method applied in this work may be characterized as an application of the theoretical concepts of the English School of mathematical statistics to the special situation prevailing in economic analysis. R. A. F i s h e r, who contributed most to the clarification of the theoretical basis and tools on which the impressive and comprehensive results of this school of statistical thought rely, formulated the purpose of statistical methods in the following words (**5**, p. 311):

> "In order to arrive at a distinct formulation of statistical problems, it is necessary to define the task which the statistician sets himself: briefly, and in its most concrete

determine the solution, while (16) is replaced by

$$\mathbf{c}^{(m)} \boldsymbol{\epsilon}\, \mathbf{c}^{(n)} = \delta^{(mn)}. \tag{24}$$

From the theory of probability we shall use the proposition that two linear combinations

$$y_1 = \eta_1^{(\tau)} z^{(\tau)}, \quad y_2 = \eta_2^{(\tau)} z^{(\tau)}, \tag{25}$$

of a set of normally distributed variables $z^{(t)}$ with means 0 and variances and covariances (square and cross moments)

$$Ez^{(t)} z^{(s)} = \pi^{(ts)} \tag{26}$$

are again jointly normally distributed with variances and covariance given by

$$Ey_k y_l = \eta_k^{(\tau)}\, \pi^{(\tau\sigma)}\, \eta_l^{(\sigma)}, \quad k, l = 1, 2. \tag{27}$$

Here the symbol E followed by a quantity subject to a probability distribution denotes the *mathematical expectation* or *mean value* of that quantity. If Z_1 and Z_2 are such quantities, the simple rules of the *calculus of mathematical expectations* are

$$\begin{cases} E\,(\zeta_1 Z_1 + \zeta_2 Z_2) = \zeta_1 EZ_1 + \zeta_2 EZ_2, \\ EZ_1 Z_2 = EZ_1 . EZ_2 \text{ if } Z_1 \text{ and } Z_2 \text{ are mutually independent.} \end{cases} \tag{28}$$

By the first of these rules (27) follows from (25) and (26).

If in (26) $\| \pi^{(ts)} \|$ is a multiple of the unit matrix,

$$\pi^{(ts)} = \sigma^2\, \delta^{(ts)}, \tag{29}$$

the variables $z^{(t)}$ are independent and of equal variance σ^2. Their distribution will be called the *spherical normal distribution* with variance σ^2, the $z^{(t)}$ themselves *spherically normal variables* of that variance.

It follows from (27) that spherically normal variables are by orthogonal transformation transformed into new spherically normal variables of the same variance.

On several occasions the following proposition will be used: *A quadratic form*

$$S \equiv z^{(\tau)} \zeta^{(\tau\sigma)} z^{(\sigma)} \tag{30}$$

in T spherically normal variables with a symmetric matrix $\| \zeta^{(ts)} \|$, can, by orthogonal transformation, be written as a sum of squares of such variables if and only if

$$\zeta^{(t\tau)} \zeta^{(\tau s)} = \zeta^{(ts)}, \tag{31}$$

while the number of such squares in S *is given by the trace*

$$T' \equiv \zeta^{(\tau\tau)} \tag{32}$$

of the matrix $||\zeta^{(ts)}||$. The proof runs as follows. If S is to be such a sum, the characteristic values $\zeta^{(t)}$ of $||\zeta^{(ts)}||$ must consist of zero's and ones only. Exactly this condition is expressed by the relations (31), which, by an orthogonal transformation which brings $||\zeta^{(ts)}||$ to the diagonal form $||\zeta^{(t)}\delta^{(ts)}||$, assume the form

$$\zeta^{(t)2} = \zeta^{(t)}. \tag{33}$$

The number of ones among the $\zeta^{(t)}$ is given by (32), which transforms into

$$T' = \Sigma_\tau \zeta^{(\tau)}. \tag{34}$$

Further, *two symmetric forms*

$$S_p \equiv z^{(\tau)}\zeta_p^{(\tau\sigma)}z^{(\sigma)}, \quad p = 1, 2,$$

which both satisfy (31) *are mutually independent sums of squares of spherically normal variables, if and only if*

$$\zeta_1^{(t\tau)}\zeta_2^{(\tau s)} = 0. \tag{35}$$

For, if $||\zeta_1^{(ts)}||$ is by orthogonal transformation brought to the diagonal form with ones only in the T_1 first places in the diagonal, the new values for $||\zeta_2^{(ts)}||$ must satisfy

$$\zeta_2^{(ts)} = 0, \quad t = 1, 2 \ldots . T_1, \quad s = 1, 2 \ldots . T,$$

that is, S_2 has become a form in the last $T - T_1$ variables only, and, by another orthogonal transformation only of these variables, it can be made a sum of squares of variables which are independent of the first T_1 new variables appearing in S_1.

In section 3 we shall deal with the distribution (3.48) of the sum (30), the so-called χ^2-distribution, and with the distribution (3.53) of the ratio of one of a set of spherically normal variables to the square root of the mean of the squares of the remaining ones — the well-known t-distribution. A very useful collection of proofs for the mathematical forms of these and other distributions can be found in a recent paper of Fisher (**12**), and also in Derksen (**4**).

PART II

Three lines of approach in existing work

3. *R. A. Fisher — the elementary regression.*

In this section, we shall consider the theory of linear regression given by R. A. Fisher (**6, 8**). In the specification used by him *the determining variables* $X_2 \ldots . X_K$ *are supposed to be measured without errors*. The dependent variable X_1 is taken as differing from a linear combination of the determining variables by an "error" which is normally distributed, independently in different years, and with the same mean 0 and variance σ^2 for every year. The observations (1.1) are considered as one sample. In repeated samples — which are, for the class of applications studied in this paper, imaginary in nature — the values

$$(1) \qquad X_{k'}^{(t)}, \quad k' = 2, 3 \ldots . K, \quad t = 1, 2 \ldots . T,$$

remain the same. In order to express that explicitly in the notations used, these assumptions may be written somewhat pleonastically in the following form:

$$(2) \qquad \begin{cases} X_1^{(t)} = \xi_1^{(t)} + z^{(t)}, \\ X_{k'}^{(t)} = \xi_{k'}^{(t)}, \end{cases}$$

$$(3) \qquad \beta_\varkappa \xi_\varkappa^{(t)} - \beta = 0, \quad \beta_1 \neq 0,$$

where the $z^{(1)} \ldots . z^{(T)}$, having the distribution function

$$(4) \qquad (2\pi\sigma^2)^{-\frac{1}{2}T} \exp\left[-\frac{1}{2\sigma^2} \Sigma_\tau z^{(\tau)2}\right],$$

are spherically normal variables with variance σ^2.

The duplicity in the notation of the $X_{k'}$ is adopted — and maintained later on for means and moments of these variables — firstly in order to hold to the convention that observations and functions of observations are denoted by latin characters, whereas unknown parameters as well as known constants of the

parent distribution are expressed by greek symbols, and secondly to facilitate comparison with the formulae to be derived in Part III.

The regression equation (3) is given in homogeneous form. Explicitly, it is [1])

$$\xi_1 = -\frac{\beta_{x'}\,\xi_{x'} - \beta}{\beta_1}. \tag{5}$$

Therefore, the β_k are related to the usual regression coefficients in Yule's notation (**39**) by

$$\beta_{1k'.23\ldots k'-1\,k'+1\ldots K} = -\frac{\beta_{k'}}{\beta_1}.$$

For the present, it is convenient to adopt as the rule which normalizes the "vector" $\boldsymbol{\beta}$ the equation

$$\beta_1 = 1. \tag{6}$$

In that case σ, $\beta_{k'}$ and β constitute the unknown parameters of the parent distribution. The "true" values $\xi_1^{(t)}$ of the first variable are by means of the unknowns $\beta_{k'}$ and β expressed in terms of the known constants $\xi_{k'}^{(t)}$.

The method of estimation, followed by Fisher, is the *method of maximum likelihood*. If in the joint distribution function of all observational values, which depends on the unknown parameters, the actual observations are inserted, there remains a function of the parameters only, which is called by Fisher (**5**) the *likelihood function*. Those values of the parameters for which this function is a maximum are taken as the *maximum likelihood estimates* of the unknown real values of these parameters.

In consequence of (4) and (2), the $X_1^{(t)}$ have the distribution function

$$(2\pi\sigma^2)^{-\frac{1}{2}T} \exp\left[-\frac{1}{2\sigma^2}\,\Sigma_\tau (X_1^{(\tau)} - \xi_1^{(\tau)})^2\right]. \tag{7}$$

As, by (2), (5) and (6),

$$X_1 - \xi_1 = \beta_x X_x - \beta, \tag{8}$$

[1]) A repeated Greek index with a dash should be summed over all values from 2 to K.

the logarithm of the likelihood function (7) is, apart from an additional constant, equal to

$$L \equiv -T \log \sigma - \frac{1}{2\sigma^2} \Sigma_\tau (\beta_\varkappa X_\varkappa^{(\tau)} - \beta)^2 . \tag{9}$$

The maximum likelihood estimates s', $b_{k'}$, b of the parameters σ, $\beta_{k'}$, β derived from the sample are those values of σ, $\beta_{k'}$, β for which L as defined by (9) reaches its maximum. This means, particularly, *that $b_{k'}$ and b are determined as those values of $\beta_{k'}$, β for which*

$$S \equiv \Sigma_\tau (X_1^{(\tau)} - \xi_1^{(\tau)})^2 = \Sigma_\tau (\beta_\varkappa X_\varkappa^{(\tau)} - \beta)^2, \tag{10}$$

the sum of squares of deviations of the first variable from the corresponding values in the right hand member of the explicit regression equation (5), *is a minimum — the well known principle of least squares.*

Completed by

$$b_1 = 1, \tag{11}$$

the b_k as determined from this principle have been called by Frisch (**13**) the coefficients of the *elementary regression equation* corresponding to the variable X_1.

The estimates s', $b_{k'}$, b are found by equating to zero the derivatives of L with respect to σ, $\beta_{k'}$, β:

$$\begin{cases} -Ts'^2 + \Sigma_\tau (b_\varkappa X_\varkappa^{(\tau)} - b)^2 = 0, \\ X_{k'}^{(\tau)} (b_\varkappa X_\varkappa^{(\tau)} - b) = 0, \\ b_\varkappa \Sigma_\tau X_\varkappa^{(\tau)} - Tb = 0. \end{cases} \tag{12}$$

Introducing the sample means (2.3) and square and product moments (2.2), this leads to

$$b = b_\varkappa \bar{X}_\varkappa , \tag{13}$$

where the $b_{k'}$ are determined by

$$m_{k'\lambda} b_\lambda = 0. \tag{14}$$

If M_{kl} denotes the cofactor of m_{kl} in the determinant $|\mathbf{m}|$, and if, as will be assumed,

$$M_{11} \neq 0, \tag{15}$$

the solution of the b_k from (11) and (14) is

$$b_k = \frac{M_{1k}}{M_{11}}. \tag{16}$$

Finally,

$$s'^2 = \frac{1}{T}\Sigma_\tau(b_\varkappa X_\varkappa^{(\tau)} - b)^2. \tag{17}$$

As a counterpart of (5) we may define the sample regression

$$x_1^{(t)} \equiv -b_{\varkappa'} X_{\varkappa'}^{(t)} + b, \tag{18}$$

and find

$$s'^2 = \frac{1}{T}\Sigma_\tau(X_1^{(\tau)} - x_1^{(\tau)})^2 \tag{19}$$

to be an estimate of σ^2 based on the deviations of the first variable from its sample regression values. On the other hand, we may derive from (17) two other forms for s'^2 which will prove to be special cases of corresponding formulae in part III. By (2.2) and (13), (17) assumes the form

$$s'^2 = \frac{1}{T} b_\varkappa m_{\varkappa\lambda} b_\lambda \tag{20}$$

and further, by (11), (14) and (16),

$$s'^2 = \frac{m_{1\varkappa} M_{1\varkappa}}{TM_{11}} = \frac{M}{TM_{11}}. \tag{21}$$

How are s'^2, $b_{k'}$, b distributed? The only elements of $\mathbf{m}$ subject to variation from one sample to the other are the m_{1k}, occurring in the first row and column only. But for m_{11}, which does not occur in any of the M_{1k}, these moments are linear functions of the spherically normal variables $z^{(t)}$. Therefore the $b_{k'}$, again depending linearly on the M_{1k}, are normally distributed themselves, their joint distribution being entirely characterized by their means, variances, and covariances.

In order to compute these quantities it is convenient first to derive from (3) some relations which are the counterparts to the formulae (14) and (16) in terms of the "true" variables and re-

gression coefficients. The square and product moments [1]) of the ξ_k

$$\mu_{kl} \equiv (\xi_k^{(\tau)} - \bar{\xi}_k)\, \xi_l^{(\tau)}, \quad \bar{\xi}_k \equiv \frac{1}{T} \Sigma_\tau \xi_k^{(\tau)}, \tag{22}$$

are, in consequence of (2), related to the sample moments m_{kl} by

$$\begin{cases} m_{11} = \mu_{11} + 2(\xi_1^{(\tau)} - \bar{\xi}_1)\, z^{(\tau)} + (z^{(\tau)} - \bar{z})\, z^{(\tau)}, \\ m_{k'1} = \mu_{k'1} + (\xi_{k'}^{(\tau)} - \bar{\xi}_{k'})\, z^{(\tau)}, \\ m_{k'l'} = \mu_{k'l'}. \end{cases} \tag{23}$$

Any corresponding minors of $|\mathbf{m}|$ and $|\boldsymbol{\mu}|$ not including elements of the first row or column are, therefore, equal. In particular [2]), remembering (15)

$$\mathcal{M}_{11} = M_{11} \neq 0. \tag{24}$$

From (3) and (22) we find

$$\mu_{k\lambda}\, \beta_\lambda = (\xi_k^{(\tau)} - \bar{\xi}_k)\, \xi_\lambda^{(\tau)}\, \beta_\lambda = \beta\, \Sigma_\tau (\xi_k^{(\tau)} - \bar{\xi}_k) = 0, \tag{25}$$

which is possible only if

$$\mathcal{M} \equiv |\boldsymbol{\mu}| = 0. \tag{26}$$

According to (24), $\boldsymbol{\mu}$ must therefore have the rank $K - 1$, so that β_k as normalized by (6) may be solved from (25):

$$\beta_k = \frac{\mathcal{M}_{1k}}{\mathcal{M}_{11}}. \tag{27}$$

Thus prepared, we may proceed to the computation of the parameters of the joint normal distribution of the $b_{k'}$. Ap-

[1]) It should be noted that $\mathbf{m}$ does not, as might be suggested by the notation, stand to $\boldsymbol{\mu}$ in the same relation which x_1, s'^2, $b_{k'}$ and b bear to ξ_1, σ^2, $\beta_{k'}$ and β. In the latter case the "latin" quantities are maximum likelihood estimates of the corresponding "greek" ones. The analogy between $\mathbf{m}$ and $\boldsymbol{\mu}$ is only that the elements of the former are the same functions of the X_k as those of the latter are of the ξ_k. To write $\mathbf{M}$ instead of $\mathbf{m}$, however, would have complicated the notation of the minors now denoted by M_{kl}.

Moreover it should be kept in mind that $\boldsymbol{\mu}$, the singular (see (26)) moment matrix of the "true" variables ξ_k, is different from $\| E(X_k^{(\tau)} - \bar{\xi}_k)\, (X_l^{(\tau)} - \bar{\xi}_l) \|$, the non-singular moment matrix of the parent distribution.

[2]) The symbol $\mathcal{M}$ replaces the Greek capital M.

plication of the calculus of mathematical expectation to (23) yields

$$(28)\qquad \begin{cases} Em_{k'1} = \mu_{k'1}, \\ E(m_{k'1} - \mu_{k'1})\,(m_{l'1} - \mu_{l'1}) = \sigma^2\,\mu_{k'l'}, \end{cases}$$

according to (22). Further we need the development of $M_{1k'}$ as a linear function of the $m_{k'1}$. If $M_{11.k'l'} = M_{11.l'k'}$ denotes the cofactor of $\begin{vmatrix} m_{11} & m_{1l'} \\ m_{k'1} & m_{k'l'} \end{vmatrix}$ in $|\mathbf{m}|$, this development is

$$(29)\qquad M_{1k'} = -\,m_{\varkappa'1}\,M_{11.\varkappa'k'}.$$

The minors in the right hand member are all constant and equal to the corresponding minors of $|\mu|$. As the same formula (29) may be written in terms of minors and elements of μ, it follows from (28) that

$$(30)\qquad \begin{cases} EM_{1k'} = \mathcal{M}_{1k'} \\ E(M_{1k'} - \mathcal{M}_{1k'})\,(M_{1l'} - \mathcal{M}_{1l'}) = \sigma^2\,\mathcal{M}_{11}\,\mathcal{M}_{11.k'l'}, \end{cases}$$

as

$$(31)\qquad \mu_{k'\lambda'}\,\mathcal{M}_{11.\lambda'l'} = \mathcal{M}_{11}\,\delta_{k'l'}.$$

Finally, we deduce from (16), (24) and (27)

$$(32)\qquad \boxed{\begin{aligned} Eb_{k'} &= \beta_{k'}, \\ E(b_{k'} - \beta_{k'})\,(b_{l'} - \beta_{l'}) &= \sigma^2\,\mu_{(11)}^{k'l'}, \end{aligned}}$$

where

$$(33)\qquad \mu_{(11)}^{k'l'} \equiv \frac{\mathcal{M}_{11.k'l'}}{\mathcal{M}_{11}}$$

are the elements of the inverse of the non-singular matrix $\|\,\mu_{k'l'}\,\|$ of order $K-1$.

We have now reached the result that, *in sampling from the parent distribution as specified by* (2), (3) *and* (4), *the elementary regression coefficients* $b_{k'}$ *of the sample are jointly normally distributed about the corresponding coefficients* $\beta_{k'}$ *of the parent distribution with variances and covariances given by the elements of the inverse of the matrix of square and product moments of the determining variables. Their distribution function is therefore*

$$(34)\qquad (2\pi\sigma^2)^{-\frac{1}{2}(K-1)}\,\mathcal{M}_{11}^{\frac{1}{2}}\exp\left[-\frac{1}{2\sigma^2}\,(b_{\varkappa'} - \beta_{\varkappa'})\,\mu_{\varkappa'\lambda'}\,(b_{\lambda'} - \beta_{\lambda'})\right].$$

The remaining problem of the distribution of s'^2 and b may be treated in connection with the *"analysis of variance"* of the problem. The deviations $X_1 - \xi_1$, the sum of squares of which constitutes the total variance S (see (10)) in the problem, may be broken up into three parts

$$(35)\quad X_1 - \xi_1 = (X_1 - x_1) + (x_1 - \xi_1 - \bar{x}_1 + \bar{\xi}_1) + (\bar{x}_1 - \bar{\xi}_1).$$

The first part is by (11), (13) and (18) equal to

$$(36)\qquad X_1 - x_1 = b_\varkappa\,(X_\varkappa - \overline{X}_\varkappa),$$

and the second part is, by (5), (6) and (18),

$$x_1 - \xi_1 - \bar{x}_1 + \bar{\xi}_1 = -\,(b_{\varkappa'} - \beta_{\varkappa'})\,(X_{\varkappa'} - \overline{X}_{\varkappa'}).$$

Both of these differences having mean 0, their cross terms with the third difference in the square of (35) vanish when summed over all values of t. The same holds for the remaining cross terms, in consequence of (2.2) and (14). Thus we are left with the square sums:

$$(37)\quad \begin{cases} S = S_1 + S_2 + S_3, \text{ where} \\ S_1 = \Sigma_\tau (X_1^{(\tau)} - x_1^{(\tau)})^2 = b_\varkappa\, m_{\varkappa\lambda}\, b_\lambda, \\ S_2 = \Sigma_\tau (x_1^{(\tau)} - \xi_1^{(\tau)} - \bar{x}_1 + \bar{\xi}_1)^2 = (b_{\varkappa'} - \beta_{\varkappa'})\,\mu_{\varkappa'\lambda'}\,(b_{\lambda'} - \beta_{\lambda'}), \\ S_3 = T\,(\overline{X}_1 - \bar{\xi}_1)^2, \end{cases}$$

$\bar{x}_1$ in S_3 being replaced by $\overline{X}_1$.

Each of these sums is a quadratic form in the spherically normal variables $z^{(t)}$. Writing

$$(38)\qquad S_p \equiv z^{(\tau)}\,\zeta_{\tau p}^{(\tau\sigma)}\,z^{(\sigma)}, \quad p = 1, 2, 3,$$

we find, firstly,

$$(39)\qquad \zeta_{33}^{(ts)} = \frac{1}{T}\,\delta^{(ts)}.$$

Further, from (16), (27), (29) and the corresponding formula for $\mathcal{M}_{1k'}$,

$$(40)\qquad b_{k'} - \beta_{k'} = -\,\frac{\mathcal{M}_{11.\varkappa'k'}}{\mathcal{M}_{11}}\,(m_{\varkappa'1} - \mu_{\varkappa'1}),$$

whence, according to (23) and (33),

$$(41)\qquad b_{k'} - \beta_{k'} = -\,\mu_{(11)}^{\varkappa'k'}\,(\xi_{\tau\varkappa'}^{(\tau)} - \bar{\xi}_{\varkappa'})\,z^{(\tau)}.$$

Consequently,

$$\zeta_2^{(ts)} = (\xi_{x'}^{(t)} - \bar{\xi}_{x'})\, \mu_{(11)}^{x'\lambda'}\, (\xi_{\lambda'}^{(s)} - \bar{\xi}_{\lambda'}). \tag{42}$$

The following relations are easily verified:

$$\begin{cases} \zeta_2^{(t\tau)}\, \zeta_2^{(\tau s)} = \zeta_2^{(ts)}, & \zeta_2^{(\tau\tau)} = K - 1, \\ \zeta_3^{(t\tau)}\, \zeta_3^{(\tau s)} = \zeta_3^{(ts)}, & \zeta_3^{(\tau\tau)} = 1, \\ \zeta_2^{(t\tau)}\, \zeta_3^{(\tau s)} = 0. \end{cases} \tag{43}$$

Therefore, as

$$S = \Sigma_\tau\, z^{(\tau)2},$$

it must, according to a proposition proved in section 2, be possible to find, by orthogonal transformation of the $z^{(t)}$, new spherically normal variables $z'^{(t)}$ of variance σ^2 such that

$$S_1 = \sum_1^{T-K} \tau\; z'^{(\tau)2}, \quad S_2 = \sum_{T-K+1}^{T-1} \tau\; z'^{(\tau)2}, \quad S_3 = z'^{(T)2}. \tag{44}$$

According to (19),

$$S_1 = Ts'^2,$$

and hence

$$Es'^2 = \frac{T-K}{T}\,\sigma^2. \tag{45}$$

An unbiased estimate of σ^2 *is therefore*

$$s^2 \equiv \frac{T}{T-K}\, s'^2 = \frac{1}{T-K}\, \Sigma_\tau\, (X_1^{(\tau)} - x_1^{(\tau)})^2. \tag{46}$$

According to (21), it may be computed from

$$\boxed{s^2 = \frac{M}{(T-K)M_{11}}}\,. \tag{47}$$

Being the mean of $T-K$ squares of spherically normal variables $z'^{(t)}$ of variance σ^2, s^2 *contains* $N \equiv T-K$ *"degrees of freedom", and* $\chi^2 \equiv Ns^2/\sigma^2$ *is distributed* [1]) *according to the well known "*χ^2*-distribution"*

$$\frac{1}{\Gamma(\frac{1}{2}N)} \left(\frac{Ns^2}{2\sigma^2}\right)^{\frac{1}{2}(N-2)} e^{-(Ns^2/2\sigma^2)}\, d\left(\frac{Ns^2}{2\sigma^2}\right), \tag{48}$$

[1]) For the proof see Fisher (**8**), or (**12**), or Derksen (**4**).

independently of the $b_{k'}$ *and* $\overline{X}_1$, which depend only on the $z'^{(t)}$ with $t > T - K$. As to $\overline{X}_1$ this follows from

$$\overline{X}_1 - \bar{\xi}_1 = T^{-\frac{1}{2}} z'^{(T)}, \tag{49}$$

resulting from a comparison of S_3 in (37) and in (44). That it holds, not only for the form S_2, but also for every individual $b_{k'}$, may be seen by the following argument. Every $b_{k'} - \beta_{k'}$ is a linear combination of the $z^{(t)}$ and hence of the $z'^{(t)}$:

$$b_{k'} - \beta_{k'} \equiv \eta_{k'}^{(\tau)} z'^{(\tau)},$$

so that, according to (37)

$$S_2 = z'^{(\tau)} \eta_{\kappa'}^{(\tau)} \mu_{\kappa'\lambda'} \eta_{\lambda'}^{(\sigma)} z'^{(\sigma)}.$$

A comparison with (44) yields

$$\eta_{\kappa'}^{(t)} \mu_{\kappa'\lambda'} \eta_{\lambda'}^{(s)} = 0 \text{ for } t, s = 1, 2 \ldots . T - K, T.$$

Putting $t = s$, this equation is, as $|| \mu_{k'l'} ||$ is positive definite and non-singular, only possible if

$$\eta_{k'}^{(t)} = 0 \text{ for } t = 1, 2 \ldots . T - K, T,$$

which proves the point.

Finally, it may be concluded from (49), that $\overline{X}_1$ *is normally distributed, independently of* $b_{k'}$, *and* s^2, *and with mean* $\bar{\xi}_1$ *and variance*

$$\sigma^2_{\overline{X}_1} = \frac{\sigma^2}{T}. \tag{50}$$

According to (13), b is a linear function of $\overline{X}_1$ and the $b_{k'}$, and, therefore, also normally distributed. In some cases it may be useful to make it identical with $\overline{X}_1$, and, for that reason, independent of the $b_{k'}$, by shifting the zero-points on the scales of the variables so as to make $\bar{\xi}_{k'} = 0$.

Readers who are acquainted with the simple and concise form (**6**, **8**) in which these results were originally published, will doubtless find the presentation in this section somewhat lengthy and complicated. This form has been deliberately chosen because we shall in later parts frequently draw on the present stock of formulae for comparison with new results.

The application of (32) requires the knowledge of σ^2. When, as in most of the practical cases, σ^2 is unknown, it can safely be

replaced by its estimate (47) if the number of observations is large, since, in consequence of (48),

$$[E(s^2 - \sigma^2)^2]^{\frac{1}{2}} = \left(\frac{2}{T-K}\right)^{\frac{1}{2}} \sigma^2. \tag{51}$$

If T is small, more accurate results are obtained by following a precept of Student (**28**), which, as Fisher has shown (**6**, **8**), can also be applied in the present situation. It consists in the use of the sampling distribution of the ratio

$$t_{k'} \equiv \frac{b_{k'} - \beta_{k'}}{s\,(\mu_{(11)}^{k'k'})^{\frac{1}{2}}} \tag{52}$$

of the normally distributed difference between estimated and true regression coefficient to the square root of the estimate of its sampling variance (32). This estimate being the mean of $N \equiv T - K$ squares of spherically normal variables with the same variance as $b_{k'} - \beta_{k'}$, of which it is independent, $t_{k'}$ is distributed according to the function

$$\frac{\Gamma\left(\frac{N+1}{2}\right)}{\sqrt{N\pi}\,\Gamma\left(\frac{N}{2}\right)} \left(1 + \frac{t^2}{N}\right)^{-(N+1)/2}, \tag{53}$$

as was anticipated by Student (**28**) and shown by Fisher (**8**).

A table, constructed by Fisher (**10**), gives percentiles of this distribution for values of N up to 30, while for larger values the difference between (53) and the normal distribution with variance 1 is negligible. Then, in testing according to a given level of significance θ the hypothesis that $\beta_{k'}$ has some specified value $\beta_{k'}^{(0)}$ — for instance 0 — the hypothesis is accepted if

$$\frac{|b_{k'} - \beta_{k'}^{(0)}|}{s\,(\mu_{(11)}^{k'k'})^{\frac{1}{2}}} < p_\theta, \tag{54}$$

where $-p_\theta$ and p_θ are the percentiles between which t falls with a probability $1 - \theta$. In general, if the statement is made that $\beta_{k'}$ lies within the range given by the inequality (54), it cannot be said that this statement has, in the individual case considered, a probability $1 - \theta$ to be true. For a definite measure of probability cannot be attached, in the individual case, to any

statement of this kind concerning $\beta_{k'}$ unless an a priori probability distribution of $\beta_{k'}$ is known. The relevant fact about the above statement is, however, that, if it is repeated again and again in a large number of cases in each of which the specification of this section applies, there will be a risk of error θ inherent in that procedure. For the considerations leading to this subtle distinction we may refer to Fisher (**7**, **9**, **11**) and Neyman (**18**, App. I).

The property which makes the ratio (52) particularly suited for the purpose it is used for is that its distribution function (53) is independent of σ and μ. It enables us to draw inferences about the $\beta_{k'}$ without actually knowing σ. Furthermore, owing to that property, the validity of (53) is not confined to cases where the $\xi_{k'}^{(t)}$ remain the same in repeated samples. This is an important point. Evidently, the restriction to cases where the values of the determining variables remain the same in repeated samples, made at the beginning of this section, is superfluous if use is made of the distribution (53) in judging the reliability of estimated regression coefficients.

As an example of the use of (53) we shall consider a relation, studied by Tinbergen (**33**), between the total sum of dividends of all corporations in the United States as dependent variable (X_1) and the reserves (X_2) and net profits (X_3) of these corporations as determining variables. The data for X_1 and X_3 have been compiled by the United States Treasury Department (Statistics of income for 1933, p. 234: Corporation income and excess profits tax return) from the profit and loss accounts of all corporations, and can, therefore, be expected to be practically without errors in the technical sense of the word. The same holds for the values for X_2, which have been computed from those for X_1 and X_3 in this way that

$$X_2^{(t)} - X_2^{(1)} = \sum_{1}^{t-1\tau} (X_3^{(\tau)} - X_1^{(\tau)})$$

the initial value $X_2^{(1)}$ at the year 1920 being arbitrary and irrelevant. Of course, appreciable differences may arise between net profits as appearing in the profit and loss accounts and real net profits, since any attempt to define the latter entails the difficulty of the valuation of capital stock. However, if we are interested only in the relation between the above mentioned

Table 3.1

t	X_1	X_2	X_3
	milliards of dollars		
1920	2.90	4.3	4.28
1921	2.69	5.7	— 0.05
1922	2.63	3.0	4.38
1923	3.30	4.7	5.83
1924	3.42	7.2	5.00
1925	4.01	8.8	6.97
1926	4.44	11.8	6.77
1927	4.76	14.1	5.88
1928	5.16	15.2	7.64
1929	5.76	17.7	8.08
1930	5.63	20.0	1.38
1931	4.18	15.7	— 3.14
1932	2.63	8.4	— 5.37
1933	2.20	0.4	— 2.36

Table 3.2

m_{kl}	$l = 1$	2	3
$k = 1$	17.54	82.61	33.45
2		470.3	66.31
3			248.09

Table 3.3

k	1	2	3
$-b_k$	— 1.000	0.162	0.091
s_{b_k}		0.015	0.020

quantities as they appear in the profit and loss accounts, it is legitimate to ascribe all erratic variation in the data to the influence on X_1 of neglected determining variables, taking X_2 and X_3 as being accurately known. Table 3.1. gives the values of the three variables for the years 1920—1933 inclusive and table 3.2 their moments m_{kl} (these figures have been taken from J. Tinbergen, **34**, p. 126). Finally, table 3.3 contains the coefficients $-b_{k'}$ of the elementary regression corresponding to X_1 as the dependent variable, and the estimated sampling standard deviations

$$s_{b_{k'}} \equiv [E(b_{k'} - \beta_{k'})^2]^{\frac{1}{2}} \tag{55}$$

of these coefficients, as given by (32) with σ replaced by s. The influence of X_3 on X_1, though much smaller than that of X_2, is doubtless significant. For $T - K = 11$ we find for the significance level $\theta = 0.05$ the percentile $p_\theta = 2.20$ in Fisher's table of t. This leads to the acceptance of any values of $\beta_{k'}$ satisfying

$$0.129 < -\beta_2 < 0.195 \quad \text{for } \beta_2,$$

$$0.047 < -\beta_3 < 0.135 \quad \text{for } \beta_3.$$

Even if a significance level als low as $\theta' = 0.01$ is adopted, the

influence of X_3 is found to be significant, the corresponding percentile in the table of t being $p_{\theta'} = 3.11$.

Figure 3.1 shows $X_1 - \overline{X}_1$ together with the linear combination

$$x_1 - \overline{X}_1 = 0.162\,(X_2 - \overline{X}_2) + 0.091\,(X_3 - \overline{X}_3)$$

by which it is approximated to, and, separately, each of the two terms in this combination.

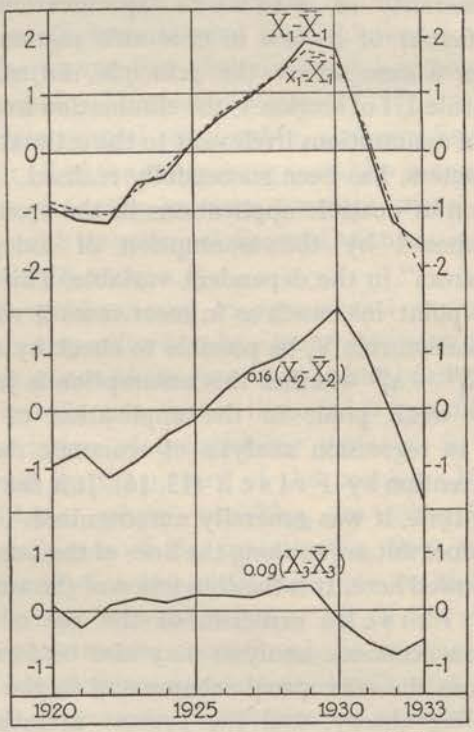

Figure 3. 1. The relation between dividends (X_1), reserves (X_2) and net profits (X_3) of United States corporations, 1920—1933. Taken from T i n b e r g e n, **34**, p. 126.

It is possible now to turn towards a discussion of the usefulness and applicability of the specification of the parent distribution as given by (2), (3) and (4) — for brevity, let us call it "F i s h e r's specification" — for the special kind of data studied here.

A conspicuous advantage is that this specification does not imply any assumption as to the distribution of $X_2 \ldots . X_K$.

If only the distribution (53) is used in interpreting the data, these observational variables need not be a random sample drawn from any probability distribution, but may as well be the values assumed by variables which develop in time by an, possibly unknown, causal mechanism; or they may be, as an intermediate case between these extremes, drawings from a series of distributions ordered in time, the next of which depends on the values drawn in the preceding ones, as were studied by Darmois (**3**). This generality of Fisher's specification is a point strongly in favour of its use in economic regression analysis. It constitutes a case where the principle, formulated in the specification rule III of section 1, the elimination from the parent distribution of assumptions irrelevant to the estimation of the required parameters, has been successfully realised.

A restriction of possible applications in the economic field is doubtless imposed by the assumption of independence of successive "errors" in the dependent variable. This is, however, not a serious point inasmuch as in most cases it will, according to the specification rule V, be possible to check by means of the deviations $X_1^{(t)} - x_1^{(t)}$ whether this assumption is justified.

The really weak point in the application of Fisher's specification to regression analysis of economic data has been brought to attention by Frisch (**13**, **16**). It is the more serious as, for a long time, it was generally unrecognized.

Frisch does not argue along the lines of the sampling theory which are followed here. It is the conviction of the author that the essentials of Frisch's criticism of the use of Fisher's specification in economic analysis may also be formulated and illustrated from the conceptual scheme and in the terminology of the sampling theory, and the present investigation is an attempt to do so. The loss in generality imposed by the assumptions involved in the construction of a parent distribution is then to some extent compensated by a gain in mathematical rigour in this respect, that by the sampling approach to the problem it is possible to attach definite risks of error to the test criteria reached, though of course on the hypothesis that the parent distribution was rightly specified.

Therefore, I venture to formulate and interpret Frisch's ideas in terms of the concepts of sampling theory. Thus transformed, his argument runs somewhat as follows.

The assumption that errors are present only in the dependent variate does not hold in most of the economic applications. If nevertheless we apply a specification which implies this assumption, we should, according to the fifth specification rule, make sure that slight deviations from this assumption could not vitiate the results obtained by the specification under consideration. There is, however, an important class of cases in which the significance of formula (32) for the sampling variances of the elementary regression coefficients as indicating the reliability of these coefficients is very closely linked to a rigorous fulfilment of the assumption that no errors are present in the independent variables. These are all cases in which, by chance or by some underlying causal relationship, a second linear relation between some or all of the *determining* variables is approximately satisfied by the data. Or, in more concrete terms, if M_{11} is small compared with its principal diagonal term $m_{22}\, m_{33} \ldots m_{KK}$ (see section 2 and **16**, section 1).

In biological problems and in agricultural experiments, in which F i s h e r's specification has found frequent application, such a situation is not likely to occur. It has been pointed out already in the introduction that in these fields as a rule adequate independent variation of determining variables is present or may be obtained by the nature of the problem. However, in economic problems, where variables are generally outside the sphere of influence of the investigator, they are often so closely interrelated that the difficulties due to high single or multiple correlations between some or all of the determining variables are by no means exceptional. As in many cases the number T of observations is relatively small, such a situation may in the first place arise "by chance" if at least one of the determining variables is influenced by irregular impulses of at least in part accidental character(crop yields, technical inventions etc.). More frequent are high correlations between determining variables as an effect of economic causation. For instance, in attempts to determine the coefficients of the regression equation of the demand for a commodity on the prices of that and competing commodities, these prices are frequently, and as a consequence just of the competing nature of these commodities, found to be highly correlated. Examples of this situation are found in S c h u l t z's investigations into the demand functions

of beef, pork and mutton (**27**) and of corn, barley, oats and hay (**26**). In these examples there are moreover common factors on the supply side which exert an influence towards high correlations between the different prices.

Let us consider the beef-pork-mutton case somewhat more closely. As a reasonably complete set of determining variables for the demand d_b for beef Schultz assumes real prices (p_b, p_p, p_m) of beef, pork and mutton, and real income i; if the relation can be taken as a linear one, we may write

$$d_b = \gamma_{bb}p_b + \gamma_{bp}p_p + \gamma_{bm}p_m + \gamma_{bi}i + z_b, \tag{56}$$

the variables being measured as deviations from equilibrium values. Here z_b is a small residual, containing the influence of neglected factors, and expected to be of accidental nature. In the current theory of price formation p_b is taken to be determined as the solution of the market equation

$$d_b = s_b \tag{57}$$

where s_b, the supply function of beef, is conceived as a function of p_b and other factors influencing the supply of beef, which we shall summarize by the symbol f_b; again assuming this supply function to be approximately linear, it may be written

$$s_b = \gamma_b p_b + \gamma'_b f_b + z'_b. \tag{58}$$

So p_b appears as a linear function of p_p, p_m, i, f_b and small accidental influences z_b and z'_b, and it is only in virtue of the latter and of the influence of f_b that the moment matrix of p_b, p_p, p_m and i fails to vanish. An additional tendency to approximate linear dependence between these variables is given by the circumstance that f_p and f_m, the summarized remaining factors affecting the supply of pork and mutton, entering into the corresponding expressions for p_p and p_m, will have important determining variables in common with f_b.

Finally, if equations are wanted estimating relations between the principal variables of the business cycle (see Tinbergen, **31**, **32**, **35**) as a tool of business cycle analysis or as a base for economic policy, it is a serious drawback that a great deal of these variables show cyclical oscillations, lagging, if at all, only a small fraction of their period, and are for that reason to a large extent intercorrelated.

Frisch illustrated his point by a sampling experiment [1]) in four normal variables

$$X_k^{(t)} = \xi_k^{(t)} + z_k^{(t)}, \tag{59}$$

$k = 1, 2, 3, 4$, $t = 1, 2 \ldots . 100$, where every $\xi_1^{(t)}$ and $\xi_2^{(t)}$ is an independent drawing from a normal distribution [2]) with mean 0 and variance 1, and where ξ_3 and ξ_4 are by

$$\begin{cases} \xi_3 = \xi_1 + \xi_2 \\ \xi_4 = \xi_1 - \xi_2 \end{cases} \tag{60}$$

made linearly dependent on ξ_1 and ξ_2. The $\xi_k^{(t)}$ are to be considered as "systematic components" or "true values", the $z_k^{(t)}$ as small errors. The latter are similar drawings divided by 10.

Now if, in fact, two linear relations exist between the "true values" of K variables, any attempt to fit one single regression equation to the observations is senseless. For every linear combination of these two relations would give a perfect fit to the scatter diagram of the $\xi_k^{(t)}$, and it would depend largely on the accidental errors as well as on the fitting procedure which of these linear combinations would be approximated by a single regression equation fitted to the $X_k^{(t)}$. Such a regression equation has no meaning because only a set of two equations can be descriptive of the systematic variation in the data. It will appear that this situation manifests itself by enormous discrepancies between the four elementary regressions more than by excessively large sampling variation of any of these.

The coefficients $b_k^{(l)}$ of each of the four elementary regressions ($l = 1, 2, 3, 4$) of the observations (59) are recorded by Frisch (**16**, p. 190) in *standard units*, (that is, in units such that

[1]) Another numerical example containing a high multiple correlation between determining variables is given by Richards (**24**). In this paper the points to the large effect of small errors of computation — due to the necessary limitation of the number of decimal places used — on the regression coefficients in cases of nearly linearly dependent determining variables. The conclusion that then eventual errors in the determining variables themselves must have a still larger effect is not drawn.

[2]) One independent drawing was obtained as the average of end digits in 100 consecutive drawings in the Norwegian State Lottery. The distribution of these averages did not show any significant skewness or curtosis.

$m_{11} = m_{22} = ... = m_{KK} = 1$), together with their "standard errors" $s_{b_k^{(i)}}$, the square roots of the expressions for their sampling variances derived in this section. Table 3.4 shows these quantities

Table 3.4

k	1	2	3	4
$b_k^{(1)}$	1.00	0.11 0.10	— 0.54 0.05	— 0.44 0.05
$b_k^{(2)}$	0.11 0.10	1.00	— 0.56 0.05	0.44 0.05
$b_k^{(3)}$	— 0.98 0.09	— 1.02 0.09	1.00	— 0.02 0.09
$b_k^{(4)}$	— 1.00 0.12	1.00 0.12	— 0.02 0.12	1.00

in the units underlying (60). In formula (32) it was unessential whether the moments of the $X_{k'}$ were written $m_{k'l'}$ or $\mu_{k'l'}$, and for formal reasons the $\mu_{k'l'}$ have been used. In discussing eventual applications of (32) to situations where the scheme (59) is a better approximation to reality, we must keep in mind that $m_{k'l'}$ and $\mu_{k'l'}$, computed respectively from $X_{k'}$ and $\xi_{k'}$, now differ, and that the sampling variance formula under discussion must be written

$$\text{(61)} \qquad \boxed{\text{est } E(b_{k'} - \beta_{k'})^2 = s^2 m_{(11)}^{k'k'} = \frac{MM_{11.k'k'}}{(T-K)\,M_{11}^2}.}$$

Here, in addition, σ^2 has been replaced by its unbiased estimate s^2 as given by (47), which is for the number of observations ($T = 100$) here considered quite legitimate, as the result of this procedure will not be appreciably different from that of the use of the ratio t given by (52).

Frisch gives the following comment on the figures in table 3.4 (**16**, p. 191):

> "If the standard errors should be reliable warning signals, they ought to tell us to keep away from *any* of these regression equations.... No statistician who is used to working with standard errors would hesitate to conclude that the last two regression coefficients ($b_3^{(1)}$ and $b_4^{(1)}$) are significant.

At least he would conclude that it is practically certain that both these coefficients are *positive*. From the way in which the example was constructed we know that this is sheer nonsense; a regression equation in the set $X_1 X_2 X_3 X_4$ has indeed no meaning at all."

It need hardly be stressed that these remarks do not imply a criticism of the logical consistency of the theory which led to the formula (32) for the sampling variance of regression coefficients. Strictly speaking, the set-up of Frisch's sampling experiment is not in conformity with the assumptions underlying Fisher's specification. The moment matrix (of rank 2) of the parent distribution from which the systematic components ξ_k were drawn is

$$\left\| \begin{matrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \\ 1 & 1 & 2 & 0 \\ 1 & -1 & 0 & 2 \end{matrix} \right\| . \tag{62}$$

Consequently, since the z_k are independent of the ξ_k, the variables X_k are 100 random drawings from the multivariate normal distribution with moment matrix

$$\mu^* = \left\| \begin{matrix} 1.01 & 0 & 1 & 1 \\ 0 & 1.01 & 1 & -1 \\ 1 & 1 & 2.01 & 0 \\ 1 & -1 & 0 & 2.01 \end{matrix} \right\| , \tag{63}$$

rather than a sample from the parent distribution assumed in Fisher's specification. Thus, it might be thought that larger values than those for the $s_{b_k^{(1)}}$ in table 3.1 would be obtained for the estimated sampling variances of the coefficients of the elementary regressions in sampling from the normal distribution given by (63). This is not the case. According to Bartlett (**1**, p. 269, form. (*20*)), the distribution function of $b_k^{(1)}$ in sampling from the normal distribution given by (63) is proportional to

$$\left[1 + \frac{\mathcal{M}_{11}^{*2}}{\mathcal{M}^* \mathcal{M}_{11.k'k'}^*} (b_{k'}^{(1)} - \beta_{k'}^{(1)})^2 \right]^{-\frac{1}{2}(T-K+2)} . \tag{64}$$

For $T - K = 96$, this distribution is not appreciably different from the normal distribution with variance

$$\frac{\mathcal{M}^* \mathcal{M}^*_{11.k'k'}}{(T - K + 1)\mathcal{M}^{*2}_{11}}, \tag{65}$$

and the determinants in this expression are estimated with a moderate sampling error by the corresponding determinants in (61). Thus, the estimated sampling variances of the elementary regression coefficients are virtually the same whether they are computed from (61) or, more closely in accordance with the type of sampling in Frisch's experiment, from an analogous expression estimating (65); and Frisch's experiment does not show anything which could not have been predicted from the results of sampling theory.

How can, then, the strikingly small values for $s_{b_k^{(t)}}$ in table 3.4 be in accordance with the fact that there is no sense in determining one single regression equation from the variables (59)? Evidently, in cases with highly interrelated determining variables which are themselves subject to error, there is no more a close connection between the sampling variances of the elementary regression coefficients and the reliability of these coefficients as indicating a systematic relationship. There is nothing disturbing in this situation. In fact, to attach significance to the coefficients of the first elementary regression, or to consider their estimated sampling variances (61) as a measure of their reliability, implies, that, according to Fisher's specification, variation in all directions perpendicular to that of the first coordinate axis is taken as systematic variation. It is not amazing that (61) loses the meaning it has on that assumption if in one or more of these directions the erratic variation is comparable in size with the systematic variation.

The point is clearly perceived if we consider the almost trivial case

$$\xi_k^{(t)} = 0, \quad k = 1, 2, 3, 4, \quad t = 1, 2 \ldots . 100, \tag{66}$$

where all systematic variation in the variables (59) has disappeared. If the $z_k^{(t)}$ are the same as before, the moment matrix of the spherically normal distribution of the X_k is now one hundredth of the unit matrix $\boldsymbol{\delta}$. The four elementary regression

vectors of this distribution coincide with the four coordinate axes. The three last ones of the coefficients

(67) $$1, \quad 0, \quad 0, \quad 0,$$

of the first elementary regression are, according to (64), unbiasedly estimated by the corresponding coefficients of the sample with a standard deviation which, owing to (65), equals

$$\frac{1}{\sqrt{97}} = 0.10 \ldots .$$

though, as we know from (66), any set of regression coefficients "fits" to the systematic components. *The elementary regression fitting procedure,* therefore, *picks out one particular set of all possible sets of regression coefficients that fit to the systematic components,* namely (67), *and estimates that one with remarkable stability in repeated samples.*

Similarly, in the other case, the adjoint of (63) being approximately (Frisch, **16**, p. 82, table 13.9)

(68)
$$\left\| \begin{array}{rrrr} 0.061 & 0.000 & -0.030 & -0.030 \\ 0.000 & 0.061 & -0.030 & 0.030 \\ -0.030 & -0.030 & 0.030 & 0.000 \\ -0.030 & 0.030 & 0.000 & 0.030 \end{array} \right\| ,$$

sets of coefficients proportional to those in the rows of (68) are unbiasedly estimated by the elementary sample regressions with sampling variances which are easily computed from (63) and (65). Frisch's experiment simply consists in the actual computation, for one sample [1]) of 100, of these regressions and of the expressions (61) which, for that number of observations, closely estimate their sampling variances (65). It exemplifies *that in cases of highly linearly interrelated determining variables the sampling variance formula* (61) *loses its meaning as a measure of the reliability of an empirical regression as an indicator of a system-*

[1]) Properly speaking, Frisch's sample was not a random one, but was restricted by the identities $\Sigma_\tau \xi_1^{(\tau)2} = \Sigma_\tau \xi_2^{(\tau)2} = 100\, \Sigma_\tau z_k^{(\tau)2}$, $k = 1, 2, 3, 4$. For simplicity, this restriction has been ignored in the above formulations, since it will modify the resulting coefficients only within the range of their sampling variation.

atic relation, if one of the assumptions underlying Fisher's specification — absence of errors in the determining variables — is not rigorously fulfilled.

It could be argued that the economist should avoid the regression analysis of any set of variables in which more than one economically significant relation may be expected to exist. However, the deficiency of the sampling variance formula (61) in cases where some or all of the determining factors are subject to error is not confined to the extreme situation, present in Frisch's sampling experiment, in which two linear relations between the "true values" of the variables are exactly satisfied.

In Parts III and IV this point will be considered more in detail. There will appear to be at least three causes at work which make (61) the more deficient as an indicator of the reliability of estimated regression coefficients the more the "true values" of the variables are interrelated.

4. *R. Frisch — the diagonal regression.*

A third line of attack on the problem has been chosen by Frisch (**16**). His starting point is the deficiency of the standard error formula (3.61) — or, in small samples, of the use of the t-test with t given by (3.52) — as found by means of Fisher's specification, in cases of approximate linear dependence between the determining variables where these are themselves subject to error. He therefore considers the whole set of elementary regressions of the sample, that is, the set of regression equations containing besides (3.18) $K-1$ analogous equations obtained by interchanging in (3.2) the subscript 1 with any of the remaining subscripts 2, 3 .... K. In terms of the scatter diagram in K-dimensional cartesian space, the n-th elementary regression is the equation of the hyperplane having minimum sum of squares of deviations from scatter points measured parallel to the n-th coordinate axis.

The main tool in Frisch's analysis is a comprehensive and detailed study of the spread in the K values

$$(1) \qquad b_{kl}^{(n)} \equiv -\frac{M_{nl}}{M_{nk}}, \quad n = 1, 2 \ldots . K,$$

obtained for the "net" regression coefficient of the k-th on the

l-th variable in each of the K elementary regressions. The whole of the argument is concerned with quantities computed from the sample; as to the specification of a probability distribution as a drawing from which the sample could be considered, Frisch says (**16**, p. 88):

> "It is on purpose that I have not attempted to give any formal and rigorous definition of the "probability" for a specified result obtained by the different minimalisations. Such a formal definition may indeed be obtained by starting from many *different* types of abstract schemes. Each scheme will lead to a particular definition of the probability in question. By focussing too much attention on the exact definiton of the probability there is some risk that one will forget the very relative and limited meaning which must always attach to such a numerical computation of a "probability". It is indeed only in a very special meaning that any such probability can be said to measure the "significance" of the results. At least, to start with, I believe it will be a better application of time and energy to work experimentally with the method and rely on one's intuitive judgement of whether a given spread in the various determinations of a given regression coefficient is reasonable or not."

The first problem to be studied by means of the coefficients (1) is the selection of the set of variables to which a regression equation will be fitted. Usually this problem presents itself in the form of the question whether some new variable, which is thought to take part in the relation under study, can safely and effectively be added to an already accepted set.

If between K variables a linear regression is approximately satisfied, the scatter diagram (in standard units!) in K-dimensional space S_K will consist of a swarm of points concentrated in the neighbourhood of a hyperplane S_{K-1} of $K-1$ dimensions. However, it will be possible to determine approximately the coefficients of the equation of S_{K-1} only if there is no second linear relation, defining another hyperplane S'_{K-1} considerably different from S_{K-1}, for whatever reason approximately satisfied by the variables. For in that case, as was pointed out already on p. 33, the scatter points would be concentrated around the intersection S_{K-2} of S_{K-1} and S'_{K-1}, and it would depend largely

on the random errors in the data and on the fitting procedure which linear combination of the equations of S_{K-1} and S'_{K-1} would result from any attempt to determine one single regression equation.

Therefore, it should be found out whether the scatter diagram of a given set of K variables is approximately singly — as opposed to multiply — linearly dependent. Thus, in trying to improve the fit in a set of variables by tentatively adding a new variable, two questions must be considered:

1. *Is the fit of a linear regression equation improved by inclusion of the new variable?* If this is the case the data do not contradict the assumption that the new variable is systematically connected with the other ones — or, in the terminology of the introduction, that it ranks among the relevant determining variables of the dependent variable. However, it still depends on the answer to the second question whether the inclusion of the new variable is really a step forward.

2. *Is the advantage of an improved fit not annihilated by a simultaneous introduction of approximate multiple linear dependence in the set?*

In order to find answers to these questions with regard to each of the variables the coefficients (1) are determined not only in the complete set of K variables which are expected to be involved in the relation under consideration, but also in every subset of every smaller number of variables selected from the complete set.

Numerical work is effected in "standard units". The moment matrix $\mathbf{m}$ then equals the matrix $\mathbf{r}$ of single correlation coefficients r_{kl}, and (1) assumes the form

$$b_{kl}^{(n)} = -\frac{R_{nl}}{R_{nk}}. \tag{2}$$

The necessary computations are performed by the *"tilling technique"*, a thoroughly rationalised method of computing in a well chosen succession the adjoints of all principal minors of $\mathbf{r}$ of every order. The corresponding coefficients are then graphically exhibited in the *"bunch map"*, containing one graph for every intercoefficient b_{kl} in every subset of every number — say K_1 — out of the K variables. Each graph contains a „bunch" of K_1 "beams" connecting the origin with one of the

K_1 points with rectangular coordinates R_{nk} and $-R_{nl}$ (or $-R_{nk}$ and R_{nl}, if $R_{nk} < 0$), the index n successively indicating all of the K_1 variables in the subset. This representation has the advantage of showing, together with the coefficients (1) as indicated by the slope of the beams, the absolute sizes of the principal minors (R_{kk} etc.) of the correlation matrix, called the *scatterances*.

Frisch's answer to both questions mentioned above is guided by the study of the "bunch map", especially by the comparison of the bunches related to the same intercoefficient before and after the introduction of a new variable. An improved fit without approaching multiple linear dependence will be indicated by a diminished spread of the values (2) for b_{kl} corresponding to the different directions of deviation square sum minimalization. In the map this manifests itself by a tightening of the bunches after adding the new variable. An additional though not obligatory indication is a decided change in the general slopes of the bunches. However, if by addition of the new variable multiple linear dependence is introduced, at least some of the bunches will "explode", that is, show beams scattered in all directions, according to the circumstance that the K different fitting procedures will approximate widely different linear combinations of the equations of the above mentioned S_{K-1} and S'_{K-1}.

A set of variables in which a good fit to a linear relation exists, but to which no additional variable, related to the problem studied, can be added without introducing multiple linear dependence, is called by Frisch *a closed set*. If, by means of criteria of which the principal ones are indicated, a closed set has been reached, the second problem remains: how to choose the fitting procedure and how to judge the precision of the result.

The starting point of this part of Frisch's analysis forms a discussion of the two-variable problem in which very simplifying assumptions are made. Considering the variables $X_k^{(t)}$, $k = 1, 2$ as the sums of a systematic and an erratic component

$$X_k^{(t)} = \xi_k^{(t)} + z_k^{(t)}, \tag{3}$$

the moments of the variables are connected to those of the systematic components by

$$m_{kl} = \mu_{kl} + \xi_k^{(\tau)} z_l^{(\tau)} + z_k^{(\tau)} \xi_l^{(\tau)} + z_k^{(\tau)} z_l^{(\tau)} \tag{4}$$

(For simplicity the means $\overline{X}_k$, $\bar{\xi}_k$, $\bar{z}_k$ are taken equal to zero). As a base for the argument, the following assumptions are made (**16**, p. 52):

$$(5)\qquad \begin{cases} z_k^{(\tau)} z_l^{(\tau)} = 0 \text{ if } k \neq l, \\ \xi_k^{(\tau)} z_l^{(\tau)} = 0 \text{ if } k \neq l, \\ 2\,\xi_k^{(\tau)} z_k^{(\tau)} + z_k^{(\tau)} z_k^{(\tau)} \geq 0. \end{cases}$$

This means that there is supposed in the sample rigorous absence of correlation between (a) any two different erratic components and (b) any systematic component and the erratic components of other variables. Finally, it is assumed that the systematic and erratic components of the same variable are not so highly *negatively* correlated that the square moment of any variable should be exceeded by that of its systematic component.

On these assumptions we have

$$(6)\qquad \begin{cases} m_{kl} = \mu_{kl} \text{ if } k \neq l, \\ m_{kk} \geq \mu_{kk}. \end{cases}$$

Now, as the $\xi_k^{(t)}$ exactly satisfy a linear relation,

$$(7)\qquad \begin{Vmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{Vmatrix} = 0,$$

and the relation may be written

$$(8)\qquad \mu_{22}\xi_1^{(t)} - \mu_{21}\xi_2^{(t)} = 0.$$

Introducing the *"disturbing intensities"* (**16**, p. 52)

$$(9)\qquad i_k \equiv \frac{m_{kk} - \mu_{kk}}{\mu_{kk}},$$

we find from (6) and (7)

$$(10)\qquad \begin{vmatrix} m_{11}(1 - i_1) & m_{12} \\ m_{21} & m_{22}(1 - i_2) \end{vmatrix} = 0,$$

whence i_1 and i_2 must satisfy

$$(11)\qquad \begin{cases} (1 - i_1)(1 - i_2) = \dfrac{m_{12}^2}{m_{11} m_{22}} \equiv r_{12}^2, \\ i_1 \geq 0, \\ i_2 \geq 0. \end{cases}$$

The only thing which by (5) is assumed about the disturbing intensities i_k is that they are restricted by these relations. This implies, however, that *the "true" regression coefficient*

$$(12) \qquad \gamma_{12} \equiv \frac{\mu_{21}}{\mu_{22}} = \frac{m_{21}}{m_{22}(1-i_2)} = \frac{m_{11}(1-i_1)}{m_{12}}$$

of ξ_1 *on* ξ_2 *in* (8) *is, by*

$$(13) \qquad \begin{cases} b_{12}^{(1)} \leq \gamma_{12} \leq b_{12}^{(2)} & \text{if } m_{12} > 0, \\ -b_{12}^{(1)} \leq -\gamma_{12} \leq -b_{12}^{(2)} & \text{if } m_{12} < 0, \end{cases}$$

included between the two elementary regression coefficients

$$(14) \qquad b_{12}^{(1)} \equiv \frac{m_{21}}{m_{22}}, \qquad b_{12}^{(2)} \equiv \frac{m_{11}}{m_{12}},$$

of the sample. Therefore, Frisch proposes to adopt the geometric mean of $b_{12}^{(1)}$ and $b_{12}^{(2)}$, the *diagonal* regression coefficient (**13**, p. 74)

$$(15) \qquad d_{12} \equiv (\operatorname{sgn} m_{12}) \left(\frac{m_{11}}{m_{22}}\right)^{\frac{1}{2}}$$

as an estimate of γ_{12} and to supply it with a *"significance factor"*

$$(16) \qquad f_{12} \equiv \left(\frac{m_{12}^2}{m_{11}m_{22}}\right)^{\frac{1}{2}} = |r_{12}|$$

indicating the factor by which d_{12} should be respectively multiplied and divided in order to obtain the limits (14).

The results of this argument are heuristically extended to the general case of K variables. As probable boundaries for γ_{kl}, the "true" regression coefficient of the k-th on the l-th variable, are adopted the values

$$(17) \qquad b_{kl}^{(k)} \equiv -\frac{M_{kl}}{M_{kk}}, \qquad b_{kl}^{(l)} \equiv -\frac{M_{ll}}{M_{lk}},$$

and as an estimate of the "true" regression equation the *diagonal* regression

$$(18) \qquad d_\varkappa x_\varkappa = 0,$$

where

$$(19) \qquad d_k^2 \equiv M_{kk};$$

if the signs in the rows of M_{kl} are compatible the same set of signs should be given to the d_k, and if one or more rows show divergencies in the signs, an investigation of the bunch map can hardly leave any doubt as to the right signs for a set of variables which, by satisfying all the above criteria, was recognized as a closed set. The estimated regression coefficient of X_k on X_l,

$$d_{kl} = -\frac{d_l}{d_k}, \tag{20}$$

is, again, the geometric mean of the boundary values (17). Its reliability is — on the condition that the set of variables was found to be a closed set in the above sense — indicated by a "significance factor"

$$f_{kl} \equiv \left(\frac{M_{kl}^2}{M_{kk}M_{ll}}\right)^{\frac{1}{2}} \tag{21}$$

which "happens" to equal the absolute value of the partial correlation coefficient between X_k and X_l.

Having resulted from the criticism of the use of Fisher's specification in the analysis of linear relations between economic variables, set out in section 3, Frisch's approach to the problem has the apparent advantage that the dangers due to high intercorrelations between determining variables are recognized and avoided, and that the variables are dealt with in a symmetrical way: all of them are allowed to contain an erratic component. As a result of this symmetrical treatment it is clearly shown that the estimation of the "true regression" from data in which all of the variables are subject to error involves an error which arises from absence of knowledge on the relative sizes of the variances of the erratic components. This error is, indeed, *the* relevant point which must be considered in any attempt at a theory of regression which admits errors in each of the variables. Limits for this error are derived.

On the other hand, Frisch's method shows some deficiences, not present in Fisher's specification, which are due to the oversimplified set of assumptions on which the argument is based. Full light may be thrown on these points only by the developments of Parts III and IV. If, therefore, some of these results

are anticipated in the remarks that follow here, their foundation and elucidation must be postponed.

Without assuming a probability distribution of the errors as a tool guiding the analysis, it is rather difficult to get at an adequate appreciation of the elaborate system of checks and cross checks which should be passed by a set of variables in order to be accepted as a "closed set". Certainly, a small spread in the elementary regression hyperplanes is an indication that the swarm of scatter points is both highly concentrated in the neighbourhood of some hyperplane and highly scattered in all directions parallel to that hyperplane. Nevertheless, it is difficult to get an idea as to how much weight should, in cases of doubt or even of contradiction, be given to the various elements of the extensive body of criteria — the more where these criteria, for lack of significance limits, are formulated in rather vague terms. There is no warrant that this vagueness of the tests, conditioned by the very method of approach, may be compensated for by increasing their number; one is left with the feeling that the various parts of the system of tests are to a large extent interdependent.

Still more the need for a probability distribution of the errors in the variables is felt in the argument concerning the precision of the estimated regression equation. It will appear in section 10 that, if all variables contain an erratic component, an estimated regression equation is subject to two quite different kinds of error, which are, under the conditions of that section, in first approximation superposed. Only one of them is considered by F r i s c h, and is due to absence of knowledge on the ratio's of the variances of the errors in the individual variables. The other one is the usual sampling error, and is, in cases of sufficient independent variation of all of the determining variables, approximately given by an expression which bears close resemblance to the formula (3.32), found by means of F i s h e r's specification. It arises from the fact that the errors in the variables, even if being uncorrelated, mutually and to the systematic components, in the parent distribution, will in general fail to be so in a sample. Therefore, the assumptions (5) are tantamount to the complete neglect of this sampling error. It depends largely on the circumstances of the problem which of these two different errors in the regression coefficients is quantitatively the most important one. In general the sampling error is relatively the more im-

portant the smaller the number of observations and the closer the fit. In fact, in a two-variable problem with correlation $r_{12} = 0.95$ some 37 observations are needed in order to obtain, according to (3.61), a sampling standard error of $b_{12}^{(1)}$ not exceeding one half of the difference $| b_{12}^{(1)} - b_{12}^{(2)} |$ between the limits assumed by Frisch.

It is beyond doubt that any concrete specification of a parent distribution of errors considerably restricts the number of cases in which the results deduced from that specification may, with some approximation to reality, be applied. However, in my opinion, a far more one-sided and severe restriction of possible applications is imposed by the plain neglect of all errors of sampling.

In another respect (21) may induce an unfounded underestimation of the precision of an estimated regression equation. Such a situation may be met with if a close fit has been obtained in a set of variables by a regression equation in which the regression coefficient(s) of the dependent variable on one or more of the determining variables, *measured in standard units*, is small compared with unity, though perhaps significantly different from zero — a case which cannot occur if, as is supposed in the argument leading to (16), only two variables are present. In connection with a definition of "close fit" which seems to be reasonable, it will be shown in section 12 that in such cases the elementary regression hyperplanes corresponding to the determining variables of small influence on the dependent variable fail to show anything which might be called a close fit to the scatter. Consequently, the diagonal regression coefficients corresponding to these variables of small influence, obtained as geometric averages partly based on these divergent elementary regression coefficients, are biased in a direction so as to exaggerate the influence of the variables concerned on the dependent variable. In particular, the use of the diagonal regression might suggest a significant influence exerted on the dependent variable by other variables, which is not warranted by the data. On the other hand, the divergent elementary regression coefficients also enter into the expressions (21), and would lead one to admit, as possible, regression hyperplanes which do not at all fit to the scatter.

5. *M. J. van Uven — the weighted regression.*

The argument given in section 3 suggests that, in cases where all variables are subject to error, a fitting procedure is needed which deals with all of the variables on an equal footing. A procedure having this property was introduced by K. Pearson (**22**). He defines the *orthogonal* regression hyperplane by the requirement *that the sum of squares of deviations of scatter points from the hyperplane, measured in a direction perpendicular to the hyperplane itself, should be a minimum.*

Let

$$\alpha\mathbf{x} - \alpha \equiv \alpha_\varkappa x_\varkappa - \alpha = 0 \tag{1}$$

be the equation of any hyperplane, the coefficients of which may be normalized by

$$\alpha_\varkappa \alpha_\varkappa = 1 . \tag{2}$$

The sum of squared distances of scatter points from this hyperplane is

$$S(\alpha, \alpha) \equiv \Sigma_\tau (\alpha \mathbf{X}^{(\tau)} - \alpha)^2 . \tag{3}$$

According to their definition, the coefficients of the orthogonal regression equation

$$a_\varkappa x_\varkappa - a = 0 \tag{4}$$

as normalized by

$$a_\varkappa a_\varkappa = 1 \tag{5}$$

are found as those values of α_k, α, satisfying (2), for which $S(\alpha, \alpha)$ is minimal. By differentiating (3) with respect to α it is easily seen that, for given values of the α_k, $S(\alpha, \alpha)$ is minimal if α assumes the value

$$\hat{\alpha} = \alpha\overline{\mathbf{X}} . \tag{6}$$

Consequently,

$$a = \mathbf{a}\overline{\mathbf{X}} , \tag{7}$$

where the a_k are found, according to (2.2), as those values of the α_k, restricted by (2), for which

$$S(\alpha) \equiv \alpha\mathbf{m}\alpha \tag{8}$$

is a minimum. The solution of this classical problem has been mentioned in section (2). $\boldsymbol{\alpha}$ is the characteristic vector of $\mathbf{m}$ corresponding to the smallest characteristic value m_1, and is uniquely determined unless the two smallest characteristic values m_1 and m_2 coincide. In geometrical language: the vector $\mathbf{a}$ indicates the direction of the shortest principal axis of the ellipsoid of inertia of the scatter points. Equation (7) shows that the orthogonal regression hyperplane passes through the "centre of gravity", with coordinates $\bar{X}_1 \ldots . \bar{X}_K$, of the scatter points.

In the set of variables $X_1 \ldots . X_K$ quantities of different nature may be present: prices, quantities of goods, ratio's of prices or quantities, etc. In making use of a presentation of the variables in K-dimensional space we should take care not to be influenced by the suggestion (not: assumption!) of homogeneity of the variables, inherent in that manner of presentation. Indeed, the implications of the postulate which served as a definition of the orthogonal regression are not immediately obvious. Distance, perpendicularity, the principal axes of an ellipsoid, are all concepts which derive their geometric meaning from invariance for orthogonal transformations. Such transformations have no sense when qualitatively different variables are involved. For that reason, it may be useful to deprive the definition of the orthogonal regression of its seeming simplicity by formulating it in terms of differences between variables measured in the same units. This can be done by introducing the foot points [1])

$$x_1^{(t)}(\boldsymbol{\alpha}) \ldots . x_K^{(t)}(\boldsymbol{\alpha}), \quad t = 1, 2 \ldots . T, \tag{9}$$

of the perpendiculars to the hyperplane (1), passing through the corresponding scatter points $X_k^{(t)}$. After van Uven (**36**) they will be called points obtained by adjustment of the scatter points to the plane, or, briefly, *adjusted scatter points*. The coordinates of the t-th adjusted point may be defined as those values of $x_1 \ldots . x_K$, restricted by (1), for which

$$S^{(t)} \equiv \Sigma_\varkappa (X_\varkappa^{(t)} - x_\varkappa)^2 , \tag{10}$$

the squared distance between scatter point and foot point, assumes its minimum value $S^{(t)}(\boldsymbol{\alpha})$. It is easily seen that the solution $x_k^{(t)}(\boldsymbol{\alpha})$ satisfies

[1]) Though the foot points, of course, depend also on α, for brevity only $\boldsymbol{\alpha}$ has been entered into the notation.

$$X_k^{(t)} - x_k^{(t)}(\alpha) = \alpha_k \, (\alpha \mathbf{X}^{(t)} - \alpha) \tag{11}$$

if the common sign factor (+ 1 or — 1) in α and α, not determined by (2), is adequately chosen. The second factor in (11) indicates the length of the t-th perpendicular. The first one represents, for $k = 1, 2 \ldots . K$, the set of direction cosines, common to all perpendiculars.

As an illustration fig. 4.1 shows the foot points of perpendiculars to the plane

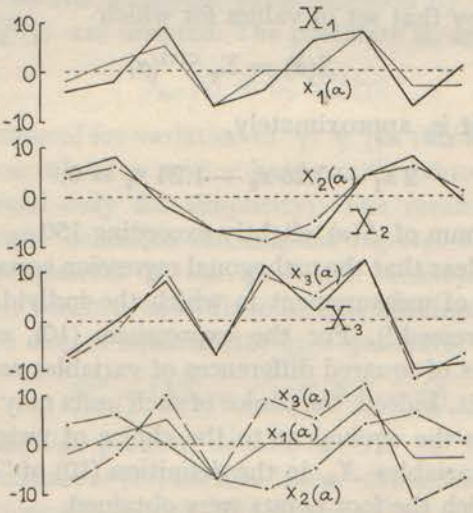

Figure 4. 1. A constructed example of orthogonal adjustment of scatter points to the plane $2x_1 - x_2 - x_3 = 0$.

$$2x_1 - x_2 - x_3 = 0 \tag{12}$$

passing through the scatter points of the series

$$\left\{ \begin{array}{rrrrrrrrrr} -4 & -2 & 9 & -7 & -4 & 6 & 8 & -7 & 1, \\ 5 & 8 & -2 & -7 & -7 & -8 & 4 & 8 & -1, \\ -7 & 0 & 8 & -7 & 11 & 2 & 12 & -10 & -9. \end{array} \right. \tag{13}$$

These series have been deliberately chosen so that the corresponding scatter points lie close to some plane without being concentrated in the neighbourhood of any line within that plane. The deviations (11) are

$$\begin{array}{rrrrrrrrr} -2 & -4 & 4 & 0 & -4 & 6 & 0 & -4 & 4, \\ 1 & 2 & -2 & 0 & 2 & -3 & 0 & 2 & -2, \\ 1 & 2 & -2 & 0 & 2 & -3 & 0 & 2 & -2, \end{array}$$

the sum of their squares is 180.

The plane (12) is not the orthogonal regression hyperplane. We can imagine "foot point series" $x_k^{(t)}(\alpha)$ computed for a number of sets of values given to α_1, α_2, α_3 in (1) (α may be taken 0 as in (13) $\overline{X}_1 = \overline{X}_2 = \overline{X}_3 = 0$). The orthogonal regression plane is indicated by that set of values for which

$$S(\alpha) = \Sigma_\tau S^{(\tau)}(\alpha) \tag{14}$$

is minimal. It is, approximately,

$$2\,x_1 - 0.26\,x_2 - 1.24\,x_3 = 0,$$

with a minimum of $S(\alpha)$ slightly exceeding 150.

It will be clear that the orthogonal regression equation depends on the units of measurement in which the individual variables X_k are expressed [1]). For the expressions (10), entering into (14), are sums of squared differences of variables each measured in its own unit. Indeed, the choice of such units may in a sense be considered as the equivalent to the choice of weights given to each of the variables X_k in the definition (10) of "distance" by means of which the foot points were obtained.

This idea leads to a generalization of the postulate defining the orthogonal regression in which such weights are explicitly introduced. This generalization was given by Rhodes (**23**). His fitting procedure may be formulated like this. In defining the "adjusted points" (9) we now require that, instead of (10), the weighted sum of squared differences

$$S_{sw}^{(t)} \equiv \varepsilon^{\varkappa\varkappa}\,(X_\varkappa^{(t)} - x_\varkappa)^2 \tag{15}$$

should be minimal for values of x_k restricted by a linear equation

$$\gamma\mathbf{x} - \gamma = 0. \tag{16}$$

The solution $x_k^{(t)}(\gamma)$ are foot points of perpendiculars to the

[1]) A systematic study of the behaviour of different regressions under some types of linear transformations of the variables was given by Frisch (**13**).

hyperplane (16) drawn through scatter points only if the scatter diagram is constructed after changing the units of measurement of the variables X_k in the ratio's $1 : \sqrt{\varepsilon^{kk}}$. Or, as an alternative interpretation, they are intersections of the hyperplane and straight lines drawn through the scatter points in a direction conjugate to that of the hyperplane with respect to the quadratic form

$$\gamma_\varkappa \varepsilon^{\varkappa\varkappa} \gamma_\varkappa . \tag{17}$$

Let $S_{sw}^{(t)}(\boldsymbol{\gamma})$ be the minimum value assumed by (15) when the solutions $x_k^{(t)}(\boldsymbol{\gamma})$ are inserted. The postulate is, again, that

$$S_{sw}(\boldsymbol{\gamma}) \equiv \Sigma_\tau S_{sw}^{(\tau)}(\boldsymbol{\gamma}) \tag{18}$$

should be minimal for variation of $\boldsymbol{\gamma}$, γ. (In this formulation no normalization rule of $\boldsymbol{\gamma}$ and γ has been introduced; above, the rule (2) served only for simplicity). The resulting regression equation might be called *the weighted regression.*

Still more general is the postulate introduced by van Uven (**36**), which is equivalent to that obtained by replacing the weighted sum of squares (15) by a quadratic form

$$S_w^{(t)} \equiv (X_\varkappa^{(t)} - x_\varkappa)\, \varepsilon^{\varkappa\lambda} (X_\lambda^{(t)} - x_\lambda) \tag{19}$$

with non singular positive definite matrix $||\varepsilon^{kl}|| \equiv \boldsymbol{\epsilon}^{-1}$.

The counterpart of (11) for this "skew" definition of the adjusted points is found as follows. The values $x_k^{(t)}(\boldsymbol{\gamma})$ of x_k, restricted by (16), for which (19) is minimal, may be found as the values minimizing

$$U \equiv S_w^{(t)} - 2\varphi\, (\gamma_\varkappa x_\varkappa - \gamma), \tag{20}$$

where the Lagrange multiplier φ is determined so that these values satisfy (16). This leads to the equations

$$\frac{\partial U}{\partial x_k} = 2\varepsilon^{k\lambda} (X_\lambda^{(t)} - x_\lambda) - 2\varphi\gamma_k = 0,$$

which are solved by forming the linear combinations

$$\tfrac{1}{2}\, \varepsilon_{k\varkappa} \frac{\partial U}{\partial x_\varkappa} = X_k^{(t)} - x_k - \varphi\, \varepsilon_{k\varkappa}\, \gamma_\varkappa = 0$$

by means of the inverse matrix $||\varepsilon_{kl}|| \equiv \boldsymbol{\epsilon}$ of $\boldsymbol{\epsilon}^{-1}$. Combined with (16) this leads to

$$\varphi = \frac{\boldsymbol{\gamma}\mathbf{X}^{(t)} - \gamma}{\boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma}},$$

whence the adjusted points $x_k^{(t)}(\boldsymbol{\gamma})$ satisfy

$$X_k^{(t)} - x_k^{(t)}(\gamma) = \varepsilon_{k\nu}\gamma_\nu \cdot \frac{\boldsymbol{\gamma}\mathbf{X}^{(t)} - \gamma}{\boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma}}, \tag{21}$$

and the minimum of (19) is

$$S_w^{(t)}(\boldsymbol{\gamma}) = [\mathbf{X}^{(t)} - \mathbf{x}(\boldsymbol{\gamma})]\,\boldsymbol{\epsilon}^{-1}[\mathbf{X}^{(t)} - \mathbf{x}(\boldsymbol{\gamma})] = \frac{(\boldsymbol{\gamma}\mathbf{X}^{(t)} - \gamma)^2}{\boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma}}. \tag{22}$$

Equation (21) shows that the adjusted points are intersections of the hyperplane (16) with straight lines drawn through scatter points in the direction indicated by the direction numbers $\varepsilon_{k\nu}\gamma_\nu$, $k = 1 \ldots K$, and therefore conjugate to the direction of the hyperplane with regard to the quadratic form $\boldsymbol{\gamma}\boldsymbol{\epsilon}^{-1}\boldsymbol{\gamma}$. Again, the coefficients $\mathbf{c}$, c of the regression equation fitted according to v a n U v e n's principle are found as the values of $\boldsymbol{\gamma}$, γ for which

$$S_w(\boldsymbol{\gamma}) \equiv \Sigma_\tau\, S_w^{(\tau)}(\boldsymbol{\gamma}) \tag{23}$$

assumes its minimum value. The resulting equation will also be called the weighted regression. If necessary it may be distinguished from R h o d e s's regression by calling it the *general*, as distinct from the *special, weighted regression.*

As (22) is homogeneous of degree zero in the γ_k, γ, the solutional values c_k, c are determined but for a common factor, and are proportional to the values of γ_k, γ minimizing

$$\Sigma_\tau\,(\boldsymbol{\gamma}\mathbf{X}^{(\tau)} - \gamma)^2 \tag{24}$$

subject to the restriction

$$\boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma} = 1. \tag{25}$$

As γ does not occur in (25), differentiation of (24) with respect to γ shows that in the minimum

$$\gamma = \boldsymbol{\gamma}\overline{\mathbf{X}}. \tag{26}$$

Inserting this into (24) we find, owing to (2.2), that $\mathbf{c}$ is proportional to that vector $\boldsymbol{\gamma}$, restricted by (25), which minimizes

$$\boldsymbol{\gamma}\mathbf{m}\boldsymbol{\gamma}. \tag{27}$$

So we are led to the minimum problem of section 2 in its general formulation. *The weighted regression coefficients* c_k *satisfy*

$$(m_{k\lambda} - l\varepsilon_{k\lambda})\, c_\lambda = 0, \tag{28}$$

where l *is the smallest root of*

$$|\mathbf{m} - l\boldsymbol{\epsilon}| = 0, \tag{29}$$

and indicates the minimum of (23). *They are uniquely determined (but for normalization) if* l *is a single root,* which will be assumed in what follows. *The "level coefficient"* c *is, according to* (26), *determined by*

$$c = \mathbf{c}\overline{\mathbf{X}}. \tag{30}$$

V a n U v e n's principle was put forward in connection with the hypothesis that the errors in the variables X_k are normally correlated with a moment matrix proportional to $\boldsymbol{\epsilon}$. In addition he derived expressions for the standard errors of the regression coefficients of the general weighted regression for different normalization rules of these coefficients. In this part of his work, however, the value of his results is reduced by errors arising from neglect of a careful distinction between the parameters of the parent distribution and the statistics computed from the observations in order to estimate these parameters.

More detailed consideration of these points is reserved for part III. The weighted regression will be the main tool in the developments in these sections. After a study of v a n U v e n's problem of the sampling variance of the weighted regression coefficients on the hypothesis that the matrix $\boldsymbol{\epsilon}$ is known a priori, an attempt will be made to remove the limitations to possible applications imposed by that hypothesis by investigating how the weighted regression coefficients and the expressions approximating their sampling variances depend upon assumed values of the elements of $\boldsymbol{\epsilon}$.

PART III

Synthesis — deductive part

6. *Comment on some comments.*

Many writers have called attention to the difficulties for an error theory of regression analysis of economic time series which are caused by the cyclical nature of these series as opposed to the random series assumed in many applications of sampling theory. E. J. Working (**37**, p. 148), speaking of the cyclical character of demand series, states:

> "It is largely because of this fact that it seemed necessary to point out earlier in this paper that care should be used in selecting an index which might be expected to move in accord with changes in the demand of a commodity. Indexes of industrial production, employment, payrolls, and consumer income, as well as various price indexes, are of a cyclical nature, and they all have moved fairly closely together in the past fifteen years. The fact that a given one of these indexes seems to be closely related to fluctuations of demand during the past fifteen years is of itself of little value in indicating how that index will be related to the cyclical shifts in demand for a particular commodity during future years. The close relation between two series of data through one complete cycle cannot be considered to be of much significance as empirical evidence. Even though the data may cover a period of a dozen years, this does not mean that we have that many independent observations. There may, indeed, be question whether data which cover one cycle should be considered the equivalent of as many as two independent observations".

An interesting illustration of these difficulties is given by Bartlett (**2**, p. 543):

> "....I may mention an examination of the correlation

obtained by Thomas between a business index and the death-rate in her very interesting book *Social Aspects of the Business Cycle* (**29**). In her investigation of the relationship between the pre-war trade cycle and the death-rate lagging one year, she was puzzled to find an apparently significant *positive* correlation of .31 between prosperity and mortality. I found on examining her figures that the correlation for the period 1890—1900, which I had reason to select owing to the death-rate showing strong cyclical fluctuations at this time due to the influx of influenza, was .71; and .11 for the remainder of the period 1863—1893. The most important contribution to the correlation was due, when the death-rate was lagging one year, to the coincidence of the large influenza fluctuations with one of the unemployment cycles. The correlation is thus probably spurious; and at least in view of its dependence on this particular coincidence loses any statistical significance it appeared to possess."

This and many other examples show that high correlations between series of cyclical nature need not be an indication of causal relationship. With regard to the problem of testing significance of correlation between time series from statistical evidence Bartlett (**2**, p. 542) concludes:

"If neither series is random, no valid test can be recommended, for it is not likely that the dependence of the observations can be specified in any satisfactory statistical way. (In practice the hypothesis of serial correlations [1]) considered earlier in this note is certainly not in itself an adequate modification). These conditions occur when the series contain short-term cyclical fluctuations that are the source of any correlation. The usual test of significance will then admit, as significant, correlations that should not be, — all the more if lagged correlations are also to be considered. In some problems it may not be possible to regard the observed correlation as more than a summary of the relationship, real or otherwise, that actually existed for the period examined. This might be said, for example, of the correlation of —.36 between unemployment and fluctuations in the level of *real* wages from 1860 to 1914 [2]), (it may be

[1]) As introduced by Yule (**40**).

[2]) M. S. Bartlett, unpublished essay.

noted that this correlation is negative in spite of a natural negative correlation between unemployment and prices). If any attempt of assessing significance with such cycles is made, it may be noted that it is the number of cycles rather than the number of observations which is the more relevant factor."

Such criticism might induce serious doubt as to the value and reliability of regression equations fitted to economic time series. At any rate it points to the necessity of adapting the assumptions on which an error theory of regression is based to the special situation present in economic applications. Considering the problem of trend fitting, which may formally be taken as a problem of regression of a variable on time, H. Working and Hotelling (**38**) proposed to evade the difficulty of interdependence of successive observations by grouping the data, each group containing a constant number of years large enough to leave only insignificantly small serial correlations in the series of group means. To this procedure the objection could be made that a series devoid of serial correlation is not for that reason a random series [1]). Further, if one should take up the idea (not proposed by Working and Hotelling) of applying this precept in fitting a regression equation on variables other than time, one would in many cases find a great deal of relevant variation in the data eliminated by the replacement of groups of observations by group means.

In my opinion the way out of the difficulties set out in the above quotations is as indicated by E. J. Working in the remarks immediately following the passage cited before:

"This brings us again to the point which has often been stressed before; the need for the analysis of demand in terms of causation. This is especially important where we are studying changes in demand which are primarily of a cyclical nature, and for which we have only one or two cycles of evidence. Closest attention must in such cases be given to a priori reasoning concerning the causal relations which may be involved, and to the many scattered bits of inductive evidence which may bear upon the problem. Even where

[1]) The series .... — 1, + 1, + 1, — 1, — 1, + 1...., having zero serial correlation, is, instead of being random, constructed by a definite mathematical law.

data are not of a cyclical nature, statistical methods alone are incapable of yielding definite evidence of causation. It is only when we combine statistical evidence in a closely knit reasoning process that we can hope to arrive at causal, and hence permanent, relationships between factors."

To use a concrete picture, the problem should be treated by the economist and the student of mathematical statistics in collaboration, each dealing with his own part of the problem; or, alternatively, one investigator should play both roles as distinct parts of his activity. The economist — or, in general, the expert in the field to which the dependent variable belongs — should by economic reasoning and general economic experience — or by his knowledge of the special branch of science concerned — devise a set of determining variables which he expects to be a complete set in the sense indicated on p. 6. The statistician cannot test the correctness of this expectation. Causation, determination are concepts outside the domain of statistics. However good the fit of a regression equation to the data combining the dependent variable with the supposed complete set of determining variables, the possibility remains that the supposed complete set, though differing from the real one in at least one variable, figures in the regression equation as a representative of it, and is able to do so because of close interrelations, "by chance" or otherwise conditioned, between variables of the two sets.

Nevertheless it is clear that a regression equation fitted by means of an erroneously supposed complete set of determining variables should not necessarily be devoid of meaning, if only the relations enabling the supposed complete set to represent the real one were themselves of a stable and economically significant nature. The regression equation could then be considered as an estimate of a relation obtained as the result of the elimination of one or more variables from two or more economically significant relations. Stability, and, with that, prognostic value, may be assigned to such a relation only if we may expect that the circumstances conditioning any of the relations from which it has been obtained by an elimination process will remain the same. The safest way to ascertain whether such a situation prevails is, indeed, to break up the intricate system of causal relations into elements, each consisting of one relation expressing one of the variables in terms of its complete set of immediately

determining variables. If the analysis has to serve as a base for discussing economic policy, such a decomposition into elementary causal links is indispensable. For then, it is required to study the effect of measures affecting only one or a few of these elementary link relations (see for instance T i n b e r g e n , **31**, **32**, **35**).

Returning to the division of labour between our two collaborators, it may be stated that to ask the statistician for a test affirming the validity of the economist's supposed complete set of determining variables on purely statistical evidence would mean, indeed, to turn the matter upside down. For, in order to be able to tell what may occur "by chance", he must know the laws, causal or in terms of probability, governing the development in time of the variables concerned. Stable frequency distributions, as in biology, do not exist in economic variables. His task would, in fact, require knowledge of the causal relationships determining the variables which he is asked to test.

The statistician, though unable to confirm the validity of the economist's contribution, may in some cases tell that he is wrong; namely in those cases in which a regression equation expressing the dependent variable in terms of the determining variables of the supposed complete set, fitted to the data, leaves large residuals, especially if these are of cyclical nature. In the opposite case, that is, if the residuals are small and do not exhibit systematic variation, *the task of the statistician is confined to a study of the reliability of the empirical regression coefficients on the hypothesis that the economist indicated the right set of determining variables, or, at least, that he did not omit important determining variables from his list.* Such a study is, indeed, possible. Since F i s h e r's specification does not imply any hypothesis as to the distribution of the systematic components, this study may be performed by means of the formulae of section 3 if the underlying assumptions of that specification are satisfied. In the present work, a generalization which admits errors in all of the variables will be considered.

The second, more inclusive, version of the above hypothesis which necessarily underlies the statistician's work has been added in order to cover another case in which it may be possible for the statistician to correct the economist's indications. If the latter performed his task so conscientiously as to include in his list, besides all relevant determining variables, one or more other

variables, the statistician may be able to tell him that the data do not warrant a significant influence of these superfluous variables on the dependent one. It will appear, however, that such a correction will only be possible if these superfluous variables are not highly intercorrelated with the relevant determining variables.

These considerations may be illustrated by the two examples reported by B a r t l e t t. A medical expert studying the factors affecting mortality would not think of the business cycle as the most important, and by no means as the only determining variable of the death-rate.

He should therefore not be disturbed in finding the sign of the single regression coefficient of the death-rate on employment opposite to that expected for the corresponding partial regression coefficient computed by means of a complete set of determining variables of the death-rate. From such a result he should conclude that the business cycle, itself being a factor of minor importance, must in the period under study have been correlated to a considerable extent, for whatever reason, with one or more of the more important determining variables of mortality; and, indeed, such a situation was detected by B a r t l e t t.

Things are different in the other example. Not only the economic expert, but every worker knows by experience the significance of the influence of unemployment on the rate of change of wages. Even if the single correlation between these variables in some cases proved to be zero instead of —0.36, it would be quite legimate to try to estimate the net influence of unemployment on the rate of change in wages by looking for other determining variables forming together with unemployment a complete set. Such an attempt has certainly failed so long as cyclical oscillation is left in the residuals from the fitted regression equation. If residuals are small and random, the reliability of the result is bound up with the validity of the supposed complete set of determining variables in the light of economic analysis and experience of factors affecting wage changes.

7. *Specification.*

The elements of the specification of the parent distribution which is put forward in this section are already present in the work reviewed in part II. They are, however, spread over the

contributions of different investigators. Our procedure will be to retain from each of these contributions just those elements which appear adequate to the requirements of the present field of applications, and to drop other features which have turned out to lead, under some circumstances, to erroneous conclusions.

To start with, the quotations in the preceding section strongly advocate the advantage of Fisher's specification, in which no assumption is made about the distribution of the systematic components of the variables. Since it is known a priori, by general experience, that a stable distribution exists only in very rare cases, an infringement against our third specification rule would be committed by the introduction of a specified probability distribution for the systematic components.

On the other hand, Frisch's criticism of the arbitrary use of the results of that specification in the analysis of economic time series clearly shows that we should try to avoid asymmetry in the treatment of the several variables. For that reason, Frisch's decomposition of each of the variables into systematic and erratic components will be adopted as an essential feature of the specification. The interpretation of the erratic components may then be taken as indicated on p. 6.

Thus, writing

$$X_k^{(t)} = \xi_k^{(t)} + z_k^{(t)} \tag{1}$$

with

$$\gamma_\varkappa \xi_\varkappa - \gamma = 0, \quad \gamma_1 \neq 0, \tag{2}$$

we introduce, besides the required parameters $\boldsymbol{\gamma}$, γ a large number, precisely $T(K-1)$, of other unknown parameters

$$\xi_{k'}^{(t)}, \quad k' = 2, 3 \ldots . K, \quad t = 1, 2 \ldots . T, \tag{3}$$

representing all values, assumed in the different years by the systematic components of the determining variables. By (2) the corresponding values $\xi_1^{(t)}$ for the dependent variable may be expressed in terms of $\boldsymbol{\gamma}$, γ and the parameters (3).

Finally, it will be assumed that the erratic components $z_k^{(t)}$ may be considered, for every value of t, as independent drawings from a multivariate normal distribution, with a non singular matrix of variances and covariances given by

$$|| \sigma^2 \varepsilon_{kl} ||, \tag{4}$$

where ϵ is normalized by some rule, for instance

$$\epsilon_{xx} = 1. \tag{5}$$

So far, as many as $T(K-1) + K + \frac{1}{2}K(K+1)$ independent unknown parameters have been ihtroduced ($\boldsymbol{\gamma}$ and γ also being normalized by some conventional rule), against only TK observational values as source of information concerning these parameters. This is a situation somewhat unusual in sampling theory, and at first sight it might be doubted whether this set up leads to any result. By what follows it will be seen, however, that the present specification is well suited to the attainment of what is our only aim, viz. *the accurate estimation of the "required parameters"* $\boldsymbol{\gamma}$, γ.

It will appear in the subsequent developments that the parameters fall into three groups:

I. *The quantities* ϵ_{kl}, *indicating the ratio's of the variances and covariances of the erratic components in the individual variables. Attempts to estimate them from the data are opposed by difficulties of a fundamental nature.* On the other hand, unbiased estimation of the "required parameters" $\boldsymbol{\gamma}$ is only possible if $\boldsymbol{\epsilon}$ is given a priori. In this situation, we shall start with the (unrealistic) assumption that $\boldsymbol{\epsilon}$ is a priori given. By means of the results obtained by that assumption we shall proceed to a study of the bias introduced in the estimation of $\boldsymbol{\gamma}$ in cases where the values, tentatively adopted for the elements of $\boldsymbol{\epsilon}$, differ from the true ones.

II. *The values* $\xi_{k'}^{(t)}$ *assumed by the systematic components of the determining variables.* Even if $\boldsymbol{\epsilon}$ is known, *estimation of the* $\xi_{k'}^{(t)}$ *is possible only with a very low accuracy.* In fact, the sampling error in estimating the point $\boldsymbol{\xi}^{(t)}$ is of the same order of magnitude as is the corresponding erratic displacement $\mathbf{z}^{(t)}$. Therefore, the estimation of the $\xi_{k'}^{(t)}$ lacks a property which is typical for the usual estimation processes of sampling theory, viz., that the accuracy of estimation may be raised up to any level by a sufficient increase of the number T of observations.

III. *The required parameters* $\boldsymbol{\gamma}$, γ *and the common factor* σ^2 *in the variances and covariances of the erratic components.* Under favourable conditions, *these parameters can be estimated with an accuracy considerably exceeding that of the individual observations.* Indeed, the sampling variances of the estimates of these para-

meters tend to zero if the number T of observations is increased beyond any limit, provided that in the infinite series of additional observations so much independent variation of the systematic components $\xi_k^{(t)}$ of the determining variables occurs, that the matrix $T^{-1} \| \mu_{k'l'} \|$ of their variances and covariances tends to some non singular limit.

It may be useful to illustrate the situation by means of a geometrical picture of the case where the number of variables is $K = 3$. The regression equation (2) of the parent distribution then represents a plane Π not parallel to the ξ_1-axis. The systematic components $\xi_k^{(t)}$ are indicated by a swarm of points within Π, one point for every year. No assumption is made about the distribution of these points over the surface of π. As a representation of the probability distribution of the observations we may imagine small "normal" clouds of probability density, each centered at one of the T points $\boldsymbol{\xi}^{(t)}$ in Π. All the clouds show the same ellipsoidal shape, of which the relative dimensions are indicated by ϵ, the absolute extension by σ. The integral of the density over anyone of the clouds is equal to unity. The observations are conceived as a series of T drawings, one from every cloud of probability density.

This picture already suggests the validity of what has been said above about the estimation of $\boldsymbol{\gamma}$ and γ. If the points $\boldsymbol{\xi}^{(t)}$ are both sufficiently numerous and sufficiently scattered in all directions within their plane, the deviations $\mathbf{X}^{(t)} - \boldsymbol{\xi}^{(t)}$ will approximately balance even within groups of years corresponding to points $\boldsymbol{\xi}^{(t)}$ localized only in parts of the region on the surface of Π which is covered by the scatter points. Hence an adequate fitting procedure using the assumed a priori knowledge of ϵ may be expected to yield a rather accurate estimation of $\boldsymbol{\gamma}$ and γ.

The specification of the parent distribution, as given by (1), (2), (4), may now be considered in the light of the remarks and conclusions contained in the preceding section. As in Fisher's specification, the systematic components of the variables may be interdependent in successive observations in an arbitrary way without affecting the validity of the conclusions drawn from the data. The hypothesis of independence in successive years is — also as in Fisher's specification — maintained only with regard to the erratic components in the variables. To some extent the validity of this hypothesis may be tested from the

data themselves. If it holds, no degree of interdependence in successive values of systematic components should prevent us from considering the observational values of every year as yielding an independent contribution to the estimation procedure of $\boldsymbol{\gamma}$ and γ. (This does not mean, of course, that observations of different years give equal contributions to the accuracy of estimation of $\boldsymbol{\gamma}$. The importance of the observational values of one particular year in increasing the accuracy of estimation of any of the γ_k largely depends on the position of the corresponding point $\boldsymbol{\xi}^{(t)}$ relative to the swarm of the other similar points in Π — a well known feature also of Fisher's specification). With regard to the problem of judging the accuracy of estimation of $\boldsymbol{\gamma}$ and γ, or, in particular, in testing the significance of the deviations from zero of one or more of the γ_k, there is no reason to consider the number of cycles as the relevant factor instead of calculating the joint effect of all individual observations.

As in all applications of sampling theory, results reached in the study of these problems have a conditional validity. They can be trusted in so much as the specification of the parent distribution can be deemed to represent adequately the essential features of the object of inquiry. In our problem this does not only mean that the decomposition (1) and the assumed normal distribution of the errors $z_k^{(t)}$, independent and stable for different years, should yield a reasonable approximation to reality, but also that a linear relation (2) should connect the systematic components.

So the specification (1), (2), (4) — like that of section 3 — already implies the hypothesis, indicated in the preceding section as the basis for the statistician's work: *that no relevant determining variable was omitted from the set of variables.* If this hypothesis is open to doubt, the formulae deduced from either of these specifications may lose their meaning, even if they happen to suggest a very accurate estimation of the regression coefficients.

8. *Estimation.*

Considering the matrix $\boldsymbol{\epsilon}$ as given a priori, we shall derive *maximum likelihood estimates* for the parameters $\boldsymbol{\gamma}$, γ, σ^2, $\xi_{k'}^{(t)}$.

The superiority of this method of estimation over other possible

methods in cases where the parent distribution is known to have some specified mathematical form has been established for large samples by Fisher (**7**) in proving that, if n is the number of observations in a sample, the limit for $n \to \infty$ of n times the sampling variance of the maximum likelihood estimate is equal to or smaller than the corresponding limit of any other consistent estimate of the same parameter. The reciprocal value of this limit was called by Fisher the *amount of information* present in one single observation and can be expressed by an integral which is, but for its sign, equal to the mean value of the second derivative, with respect to the parameter to be estimated, of the logarithm of the distribution function of one independent observation.

In this theory, the sample (1.1) is to be considered as one single "observation", every "repeated sample" constituting an additional "observation", and, as was pointed out in section 1, in time series analysis only one such "observation" can be obtained for a given time period. In the theory of small samples, into which we are thus forced, results of great formal beauty have been obtained by Fisher (**7**) by an extension of the definition of the amount of information to a number of observations and to statistics derived from such observations. The information integral for n independent observations proved to be n times that for a single observation, while the information integral for a statistic or a set of statistics cannot exceed that for the observations it is computed from. If a statistic exists which contains all the information present in the observations, it can be obtained by the method of maximum likelihood. In connection with these properties, Fisher considers as the aim of the theory of estimation the establishment of such statistics which together contain all information present in the data, and are chosen in such a way as to admit statements in terms of *fiducial probability* (**7**, **9**, **11**) concerning the unknown parameter(s), while he declines (see the discussion of Neyman's paper, **18**, p. 618) to assign unique validity to such statements from statistics conveying only a part of the information in the sample.

This view is not shared by Neyman (**18**) who introduced statements of this kind in terms of *confidence intervals* irrespective of the amount of information in the statistic on which the statement is based. Since the validity of these latter statements can formally be proved, the point at issue can only concern their logical meaning.

As I see it, the adequacy of the above mentioned conception of the aim of statistical estimation in cases where the number of observations is small owing to the very nature of the problem, would be considerably cleared up if some proposition could be proved establishing some connection between the amount of information in a statistic or a set of statistics and the statements in terms of confidence intervals, to be made about the parameter(s) from these statistics — in such a way that the statistics containing more of the information would admit, in the average, more "accurate" statements on the parameter(s). Such

a proposition could not, as one might think, be formulated in terms of the average length of confidence intervals in repeated samples, since this length is changed if the concerned parameter is replaced by a monotonic function of it — a substitution which must be irrelevant to the problem. Yet, I must confess that, in the absence of such a proposition in a more adequate formulation, I have some difficulty n understanding the implications of the endeavour to derive also from small samples statistics having an information integral as large as possible.

In view of these difficulties, there will not be attempted a discussion of the information integrals regarding the several parameters under consideration. Such an attempt would, moreover, involve a technical mathematical problem for itself; a generalization would be required of the concept of information to the simultaneous estimation of more than one parameter if, in addition, a number of other, irrelevant, unknown parameters occur in the parent distribution function. Thus, maximum likelihood estimation is here adopted simply because it seems to lead to useful statistics.

Owing to (7.4), the distribution function of the erratic components $z_k^{(t)}$ relating to one particular year is

$$(1) \qquad f^{(t)} \equiv (2\pi\sigma^2)^{-\frac{1}{2}K} |\varepsilon_{kl}|^{\frac{1}{2}} \exp - \frac{z_\varkappa^{(t)} \varepsilon^{\varkappa\lambda} z_\lambda^{(t)}}{2\sigma^2},$$

whence the logarithm

$$\Sigma_\tau \log f^{(\tau)}$$

of the likelihood function differs only by a constant term from

$$(2) \qquad L \equiv - TK \log \sigma - \frac{1}{2\sigma^2} (\mathbf{X}^{(\tau)} - \boldsymbol{\xi}^{(\tau)})\, \boldsymbol{\epsilon}^{-1}\, (\mathbf{X}^{(\tau)} - \boldsymbol{\xi}^{(\tau)}).$$

We must find such values s'^2, $\mathbf{c}$, c, $\mathbf{x}^{(t)}$ for σ^2, $\boldsymbol{\gamma}$, γ, $\boldsymbol{\xi}^{(t)}$ satisfying (7.2) for which L assumes its maximum value $\hat{L}$. This maximalization will be performed in two steps. First arbitrary values $\boldsymbol{\gamma}$, γ will be imposed for a determination of those values $s'^2(\boldsymbol{\gamma})$, $\mathbf{x}^{(t)}(\boldsymbol{\gamma})$ of σ^2, $\boldsymbol{\xi}^{(t)}$, restricted by (7.2) with the imposed values of $\boldsymbol{\gamma}$, γ inserted, for which L reaches its restricted maximum $\hat{L}(\boldsymbol{\gamma})$. Then $\mathbf{c}$ and c will be found as those values of $\boldsymbol{\gamma}$, γ for which $\hat{L}(\boldsymbol{\gamma})$ is maximal.

Now, both steps have already been performed in section 5, the only complication being the occurrence of σ^2 in (2). The second term in the right hand member of (2) consists of a factor

$1/2\sigma^2$ multiplied into the opposite of a sum of T terms of the form (5.19) each depending on the $\xi_k^{(t)}$ for one value of t only. Consequently, the first step requires that each of these terms is separately made a minimum by adequate choice of the corresponding point $\mathbf{x}^{(t)}(\gamma)$. These latter quantities are, therefore, identical with the adjusted points according to van Uven's definition, as given by (5.21). Further, by differentiation of L with respect to σ, according to (5.22),

(3)
$$s'^2(\gamma) = \frac{1}{TK}\,[\mathbf{X}^{(\tau)} - \mathbf{x}^{(\tau)}(\gamma)]\,\boldsymbol{\epsilon}^{-1}\,[\mathbf{X}^{(\tau)} - \mathbf{x}^{(\tau)}(\gamma)] = \frac{\Sigma_\tau\,(\gamma\mathbf{X}^{(\tau)} - \gamma)^2}{TK\;\gamma\boldsymbol{\epsilon}\gamma},$$

whence

(4)
$$\hat{L}(\gamma) = -\tfrac{1}{2}\,TK\,[\log s'^2(\gamma) - 1].$$

It ensues that, as to the second step, we may, instead of maximizing $\hat{L}(\gamma)$, as well minimize $s'^2(\gamma)$, which bears a constant ratio to (5.23). This problem has also been treated in section 5. The results may be recapitulated as follows. *Maximum likelihood estimation of the parameters of the parent distribution as specified by* (7.1), (7.2), (7.4) *with an a priori given matrix* $\boldsymbol{\epsilon}$ *yields as estimates* $\mathbf{c}$, c of γ, γ *the coefficients of the general weighted regression, defined by* (5.28), (5.29), (5.30); σ^2 *is estimated by*

(5)
$$s'^2 = \frac{l}{TK},$$

l being defined as the smallest root of (5.29), *and, owing to* (5.28), *equalling*

(6)
$$l = \frac{\mathbf{cmc}}{\mathbf{c\epsilon c}};$$

the systematic components $\xi_k^{(t)}$ are estimated by the coordinates $x_k^{(t)}$ of the points obtained by adjustment of the scatter points to the general weighted regression hyperplane according to van Uven's principle of adjustment, which are given by

(7)
$$X_k^{(t)} - x_k^{(t)} = \varepsilon_{k\lambda}\,c_\lambda \cdot \frac{\mathbf{c}(\mathbf{X}^{(t)} - \overline{\mathbf{X}})}{\mathbf{c\epsilon c}}.$$

With regard to the estimation of σ^2 the situation is somewhat

peculiar. It may best be understood by a comparison with the cases in which all or some of the other parameters are given a priori. If — besides $\boldsymbol{\epsilon}$ — all the points $\boldsymbol{\xi}^{(t)}$, and hence also $\boldsymbol{\gamma}$ and γ, were known a priori, the observations $\mathbf{X}^{(t)}$ would define T drawings $\mathbf{z}^{(t)}$ from the distribution (1) in which σ^2 is the only unknown parameter. Its maximum likelihood estimate would be

$$s^2(\boldsymbol{\xi}) \equiv \frac{1}{TK} (\mathbf{X}^{(\tau)} - \boldsymbol{\xi}^{(\tau)})\, \boldsymbol{\epsilon}^{-1} (\mathbf{X}^{(\tau)} - \boldsymbol{\xi}^{(\tau)}), \tag{8}$$

with the mathematical expectation

$$Es^2(\boldsymbol{\xi}) = \frac{\sigma^2}{K} \varepsilon_{\varkappa\lambda}\, \varepsilon^{\varkappa\lambda} = \sigma^2, \tag{9}$$

while $\chi^2 \equiv TKs^2(\boldsymbol{\xi})$ has the distribution function (3.48) with N replaced by TK (this follows from the fact that $\mathbf{z}\boldsymbol{\epsilon}^{-1}\mathbf{z}$ is the sum of squares of K spherically normal variables, as may be seen by a skew linear transformation of the z_k which brings both $\boldsymbol{\epsilon}^{-1}$ and $\boldsymbol{\epsilon}$ to the form of the unit matrix $\boldsymbol{\delta}$).

Next, the case may be considered where, besides $\boldsymbol{\epsilon}$, only $\boldsymbol{\gamma}$ and γ are known a priori. The maximum likelihood estimate of σ^2 is now given by (3), which may, on account of (7.2), also be written

$$s'^2(\boldsymbol{\gamma}) = \frac{\Sigma_\tau [\boldsymbol{\gamma}(\mathbf{X}^{(\tau)} - \boldsymbol{\xi}^{(\tau)})]^2}{TK\, \boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma}}, \tag{10}$$

having the mathematical expectation

$$Es'^2(\boldsymbol{\gamma}) = \frac{\sigma^2}{K}, \tag{11}$$

while

$$\chi^2 \equiv TKs'^2(\boldsymbol{\gamma})$$

is distributed according to (3.48) with $N = T$, since the T variables $\boldsymbol{\gamma}\mathbf{z}^{(t)}$ are spherically normal with variance $\sigma^2\boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma}$.

Evidently $s'^2(\boldsymbol{\gamma})$ is a consistent estimate of σ^2/K instead of σ^2, which is unbiasedly estimated by

$$s^2(\boldsymbol{\gamma}) \equiv Ks'^2(\boldsymbol{\gamma}), \tag{12}$$

only, however, with the accuracy of an estimate of the χ^2-type based on T degrees of freedom, as compared with TK degrees

in the case of $s^2(\xi)$. While every year contributes K independent squares (viz. of the K above mentioned skew combinations of the z_k) to $s^2(\xi)$, only one squared linear combination per year enters into $s^2(\gamma)$. The remaining degrees of freedom are consumed in the estimation of the $\xi^{(t)}$.

This all follows from the analysis of variance of the problem. The total variance

$$(13) \qquad S \equiv (\mathbf{X}^{(\tau)} - \xi^{(\tau)})\, \epsilon^{-1} (\mathbf{X}^{(\tau)} - \xi^{(\tau)}) = \mathbf{z}^{(\tau)}\, \epsilon^{-1}\, \mathbf{z}^{(\tau)}$$

consists of T independent sums

$$(14) \qquad S^{(t)} \equiv \mathbf{z}^{(t)}\, \epsilon^{-1}\, \mathbf{z}^{(t)},$$

each of which is the sum of squares of K spherically normal variables with variance σ^2. When

$$\mathbf{z} = \mathbf{X} - \xi = [\mathbf{X} - \mathbf{x}(\gamma)] + [\mathbf{x}(\gamma) - \xi]$$

is inserted into (14), the cross term vanishes on account of (5.21), as both points $\mathbf{x}(\gamma)$ and ξ lie within the hyperplane (7.2). Consequently S splits up into the $T + 1$ sums

$$(15) \qquad \begin{cases} S &= S_I + \Sigma_\tau S_{II}^{(\tau)}, \\ S_I &= [\mathbf{X}^{(\tau)} - \mathbf{x}^{(\tau)}(\gamma)]\, \epsilon^{-1}\, [\mathbf{X}^{(\tau)} - \mathbf{x}^{(\tau)}(\gamma)], \\ S_{II}^{(t)} &= [\mathbf{x}^{(t)}(\gamma) - \xi^{(t)})]\, \epsilon^{-1} [\mathbf{x}^{(t)}(\gamma) - \xi^{(t)}]. \end{cases}$$

The first one is, according to (3), (10), (12) equal to

$$S_I = T s^2(\gamma) = \frac{\Sigma_\tau (\gamma \mathbf{z}^{(\tau)})^2}{\gamma \epsilon \gamma},$$

the sum of squares of T spherically normal variables

$$\frac{\gamma \mathbf{z}^{(t)}}{(\gamma \epsilon \gamma)^{\frac{1}{2}}}$$

of variance $\sigma^2 \,.\, \Sigma_\tau S_{II}^{(\tau)}$ therefore equals a remaining sum of $T(K-1)$ such squares, which can be taken in such a way that every $S_{II}^{(t)}$ contains $(K-1)$ squared linear combinations of the corresponding $z_k^{(t)}$ only.

Without explicitly writing down the distribution of the $x_k^{(t)}(\gamma)$ it will be clear that they are scarcely better "estimates" of the $\xi_k^{(t)}$ than the observations $X_k^{(t)}$ themselves, only a fraction $1/K$

of the total variance being eliminated if the latter are replaced by the $x_k^{(t)}(\gamma)$, or "adjusted" to the hyperplane. The process of adjustment only eliminates the component $\mathbf{X} - \mathbf{x}(\gamma)$ of the erratic displacement $\mathbf{z}$ in the direction indicated by the direction numbers $\varepsilon_{k\lambda}\gamma_\lambda$, leaving $K - 1$ other independent components within the hyperplane.

That a maximum likelihood estimate of an unknown variance is biased and must be corrected for loss of degrees of freedom in the estimation procedure is by no means new. In section 3, for instance, this correction was made in (3.46). The present situation is unusual only in so far that, owing to the large number of unknown parameters, for every degree of freedom retained in the estimation of σ^2, $K - 1$ other ones are lost. Because of that, the "bias" amounts to $K - 1$ times the estimate itself, and the "correction" consists of a multiplication by K.

Finally, the case where also $\boldsymbol{\gamma}$ and γ are unknown. The expression (5) is no more a sum of squares of spherically normal variables. Nevertheless, it will appear in the next sections that, under some conditions, it can be approximated to by such a sum. Similarly the c_k can be approximated to by linear combinations of normal variables. It is then possible to consider the analysis of variance for these comparison "estimates" (which, in fact, depend on the unknown parameters), to which the real maximum likelihood estimates are approximately equal. Here a situation analogous to that encountered in section 3 arises. K further degrees of freedom are used in the estimation of $\boldsymbol{\gamma}$ and γ, so that

$$s^2 \equiv \frac{TK}{T-K}\, s'^2 = \frac{l}{T-K} \tag{16}$$

is approximated to by an "estimate" of the χ^2-type based on $T - K$ degrees of freedom and having mathematical expectation σ^2.

The question arises whether it would be possible to derive from the data estimates also of the ε_{kl}, by pushing the maximalization procedure still one step further. This would require the determination of values e_{kl} for ε_{kl} which, restricted by some normalization rule, make l as defined by (5.29) a minimum. Yet, such a procedure would hardly make sense. All solutions of equations of maximum likelihood, derived in the preceding pages,

show the property of not depending on the units of measurement of the individual variables. This feature of the formulae has been obtained just by the assumption of a priori given values of ε_{kl}, which enabled us to build quadratic forms (of the type of the last term in (2)) invariant for a change in the units of measurement, or which — in geometrical terminology — introduced a metric in the space of the variables $\mathbf{X}$. Now, on closer inspection, it turns out to be impossible to choose a useful normalization rule of $\boldsymbol{\epsilon}$ such that also $\mathbf{e}$ is independent of the units of measurement. Thus, at least in small samples, the sampling distribution of any such values e_{kl} cannot be expected to have a systematic connection with the unknown parameters ε_{kl}.

More light is thrown upon the question by considering another parent distribution than that used here, which is obtained if the $\boldsymbol{\xi}^{(t)}$, instead of remaining the same in repeated samples, are supposed to be random drawings from a normal distribution with a singular matrix

$$T^{-1}\,\boldsymbol{\mu}$$

of variances and covariances. In that case the $\mathbf{X}^{(t)}$ are random drawings from a normal distribution with a matrix of variances and covariances

$$\boldsymbol{\pi} \equiv T^{-1}\,\boldsymbol{\mu} + \boldsymbol{\epsilon}.$$

From this matrix — and only this one can be estimated from the sample — $\boldsymbol{\epsilon}$ cannot be reconstructed by the sole condition that $\boldsymbol{\mu}$ should be singular.

This analogy leads one to expect accurate estimation of $\boldsymbol{\epsilon}$ to be impossible, if the points $\boldsymbol{\xi}^{(t)}$ in the specification of section 7, however numerous, are scattered in their plane in such a way that they could have constituted a not improbable sample from a normal distribution. There are, however, indications that in other cases[1]), if $\boldsymbol{\epsilon}$ is known to be a diagonal matrix, estimation of the ε_{kk} must be possible from a sufficiently large number T of observations.

Though, therefore, much is left unsettled as to the possibility of estimating $\boldsymbol{\epsilon}$, this problem may be expected to become im-

---

[1]) Consider for instance the case with $K = 2$ where one half of the points $\boldsymbol{\xi}^{(t)}$ coincide with $\boldsymbol{\xi}^{(1)}$, the other half with $\boldsymbol{\xi}^{(2)}$, and where

$$|\,\xi_k^{(1)} - \xi_k^{(2)}\,| \gg \sigma^2 \varepsilon_{kk}.$$

portant only if T is considerably larger than it is in most of the applications for which this study is intended. For that reason, it seems better to deal with ϵ by the procedure indicated on p. 61.

Though van Uven did not maintain in his formulae the distinction between parameters and statistics, he was apparently guided by a conceptual scheme of the nature of the variables similar to that exposed in section 7. The expression (16) for the estimation of σ^2 is given by van Uven (**36**) in formula (*46*), p. 154. An inconsistency in his set-up is that, on the one side, the adjusted points $\mathbf{x}^{(t)}$ are defined (**36**, p. 145) by a principle equivalent to that applied in section 7, leading to mutually parallel adjusting displacements, or to values of

$$X_k^{(t)} - x_k^{(t)}, \quad k = 1, 2, \ldots . K, \tag{17}$$

proportional for all values of t — and, on the other, probability is introduced into the problem by the assumption that the displacement components (17) are subject to a joint normal distribution with a matrix of variances and covariances proportional to ϵ. Evidently, van Uven's symbol for the displacements (17) is used in two different senses.

The way in which the expressions for the maximum likelihood estimates are derived in this section presupposes that ϵ is non singular. It is easily ascertained that these expressions preserve their meaning if some of the variables X_k are supposed to be accurately known — that is, if some rows of ϵ are supposed to consist of zero's only, while a non singular positive definite matrix is left after omitting these rows and the corresponding columns — provided only that

$$\gamma\epsilon\gamma > 0. \tag{18}$$

In particular, the present specification of the parent distribution becomes identical with Fisher's specification in the special limiting case

$$\begin{cases} \varepsilon_{11} = 1, \\ \varepsilon_{kl} = 0 \text{ unless } k = l = 1, \end{cases} \tag{19}$$

and corresponding formulae become identical if (19) holds. Firstly (7.1) changes into (3.2). Then, the last $K - 1$ equations in (5.28) pass into (3.14), whence $\mathbf{c}$, if normalized by $c_1 = 1$,

will be the same as $\mathbf{b}$. Further, a comparison of (16), l being derived from (6), and (3.46), with s'^2 as given by (3.20), shows the identity of the estimates s^2 of σ^2 in both cases (though the expressions for s'^2 are different!). Finally, (7) shows that the adjusting displacements become parallel to the X_1-axis, while the adjusted points come to lie on the sample regression hyperplane (3.18), since the first equation in (7) passes into (3.36).

9. *Distribution.*

It would probably be a hard task to find the exact sampling distribution of the coefficients $\mathbf{c}$ and c of the weighted regression equation in sampling from the parent distribution specified by (7.1), (7.2), (7.4). Complications would doubtless arise in consequence of the definition of $\mathbf{c}$ and c as constituting a solution of

$$(m_{k\lambda} - l\varepsilon_{k\lambda})c_\lambda = 0\,, \tag{1}$$

$$\mathbf{c}\overline{\mathbf{X}} - c = 0, \tag{2}$$

if for l the *smallest* root l_1 of

$$|\,\mathbf{m} - l\boldsymbol{\epsilon}\,| = 0 \tag{3}$$

is inserted. For even if $T\sigma^2$ were a very small fraction of the smallest *positive* [1]) root λ_2 of

$$|\,\boldsymbol{\mu} - \lambda\boldsymbol{\epsilon}\,| = 0\,, \tag{4}$$

the circumstance that the normal distribution of the $z_k^{(t)}$ extends to infinity would make for a finite, though small, probability for samples to be drawn, in which the two smallest roots l_1 and l_2 of (3) are nearly equal. In such samples a slight variation of only one of the $z_k^{(t)}$ may suffice to produce a large discontinuous jump in $\mathbf{c}$, if only l_1 and l_2 coincide at some intermediate point during that variation.

In the face of these difficulties approximation seems to be the most efficient tool. According to (7.1) we may generalize (3.23) to

$$\mathbf{m} = \boldsymbol{\mu} + \mathbf{m}' + \mathbf{m}'', \tag{5}$$

---

[1]) The absolutely smallest root is $\lambda_1 = 0$, owing to (3.26).

where

$$m'_{kl} = (\xi_k^{(\tau)} - \bar{\xi}_k) z_l^{(\tau)} + z_k^{(\tau)} (\xi_l^{(\tau)} - \bar{\xi}_l) , \tag{6}$$

$$m''_{kl} = (z_k^{(\tau)} - \bar{z}_k) z_l^{(\tau)} . \tag{7}$$

Writing, again, $l_1 = l$, analogous developments

$$l = 0 + l' + l'' + \ldots . \tag{8}$$

$$\mathbf{c} = \boldsymbol{\gamma} + \mathbf{c}' + \mathbf{c}'' + \ldots . \tag{9}$$

of l and $\mathbf{c}$ may be defined by the requirement that, in the equations

$$[\mu_{k\lambda} + m'_{k\lambda} + m''_{k\lambda} - (l' + l'' + \ldots) \varepsilon_{k\lambda}] (\gamma_\lambda + c'_\lambda + c''_\lambda + \ldots) = 0 , \tag{10}$$

obtained by inserting these developments into (1), sums of terms of equal order of magnitude (that is: terms bearing the same total number of dashes) are separately equal to zero:

$$\mu_{k\lambda} \gamma_\lambda = 0 , \tag{11}$$

$$\mu_{k\lambda} c'_\lambda + m'_{k\lambda} \gamma_\lambda = l' \varepsilon_{k\lambda} \gamma_\lambda , \tag{12}$$

$$\mu_{k\lambda} c''_\lambda + m'_{k\lambda} c'_\lambda + m''_{k\lambda} \gamma_\lambda = l' \varepsilon_{k\lambda} c'_\lambda + l'' \varepsilon_{k\lambda} \gamma_\lambda , \text{ etc.} \tag{13}$$

Since equation (11) holds on account of (7.2) (compare (3.25)), it supplies the justification for the first terms (0 and $\boldsymbol{\gamma}$) in the developments of l and $\mathbf{c}$ assumed in (8) and (9) respectively. For the corresponding development of c,

$$c = \gamma + c' + c'' + \ldots . \tag{14}$$

we find from (5.30)

$$c' = \boldsymbol{\gamma} \bar{\mathbf{z}} + \mathbf{c}' \bar{\boldsymbol{\xi}} , \tag{15}$$

$$c'' = \mathbf{c}' \mathbf{z} + \mathbf{c}'' \boldsymbol{\xi} , \text{ etc.} \tag{16}$$

The question of the conditions for a rapid decrease in the value of successive terms in the developments deserves some consideration. In (5) every next term is one degree higher in the $\mathbf{z}^{(t)}$, and one degree lower in the $\boldsymbol{\xi}^{(t)} - \bar{\boldsymbol{\xi}}$ than the preceding one. This suggests that a rapid convergence in the developments may be expected if the dimensions of the "clouds of probability density" from which the points $\mathbf{X}^{(t)}$ are drawn, as given by the matrix

$\sigma^2\epsilon$

amount to only a small fraction of the dimensions of the swarm of points $\xi^{(t)}$ within the hyperplane (7.2), in broad outline described by the matrix

$$T^{-1}\mu.$$

of the variances and covariances of the $\xi_k^{(t)}$. In this form, the statement is rather unsatisfactory, for the same reason which led us in section 5 to comment on the orthogonal regression. There is not much sense in the notion of the dimensions of a swarm of points, indicating by cartesian representation variables measured in incomparable units. So, choosing a formulation which implies only comparison of quantities measured in the same unit, we will consider, *provisionally*, as a condition [1]) for rapid convergence in the developments, that

$$\frac{T\sigma^2\epsilon_{kk}}{\mu_{kk}} \ll 1 \tag{17}$$

for every value of k. This means that, if our method of approach is not to be meaningless, the variances of the erratic components should be relatively small fractions of the variances of the corresponding systematic components.

The condition does not impose an important restriction on the applicability of the method. If in any series an erratic component occurs which is about as variable as the systematic component, this series will not even permit, by whatever method, a rough estimation of the regression coefficient which attaches to the systematic component in a relation with other variables, however accurately the latter may be known. In general, if accurate estimation of regression coefficients is possible, it may be performed by a few terms in the developments. The more the accuracy of estimation is diminished by the presence of large erratic components, the more terms will be needed to evaluate that amount of accuracy which is left. In this study we shall include in the analysis only terms up to the second order.

According to (7.2) and (6)

$$\gamma\mathbf{m}'\gamma = 0\,. \tag{18}$$

[1]) As we are dealing with positive definite matrices, it is sufficient to formulate the condition in terms of the diagonal elements.

Consequently, by multiplying both members of (12) by γ_k and summing over k,

$$l' = 0\,, \tag{19}$$

in virtue of (8.18). The same operation, applied to (13), yields

$$l''\boldsymbol{\gamma\epsilon\gamma} = \boldsymbol{\gamma m' c'} + \boldsymbol{\gamma m''\gamma}\,. \tag{20}$$

The rank of $\boldsymbol{\mu}$ being $K-1$, $\mathbf{c}'$ is determined but for an arbitrary additive vector $\varphi\boldsymbol{\gamma}$ by the K equations (12). The factor φ must then be determined by a conventional rule normalizing the first approximation $\boldsymbol{\gamma}+\mathbf{c}'$ to $\mathbf{c}$. The solution of $\mathbf{c}'$ from (12) and from the normalization rule may conveniently be performed by means of a matrix $\boldsymbol{\mu}_{(1)}^{-1} \equiv \| \mu_{(1)}^{kl} \|$ which might be called the *partial inverse* of $\boldsymbol{\mu}$ and is defined by the equation

$$\left(\mu_{(1)}^{k\varkappa} + \frac{\gamma_k\gamma_\varkappa}{\gamma_\lambda\gamma_\lambda}\right)\left(\mu_{\varkappa l} + \frac{\gamma_\varkappa\gamma_l}{\gamma_\nu\gamma_\nu}\right) \equiv \delta_{kl}\,. \tag{21}$$

This definition is equivalent to the following one: $\boldsymbol{\mu}_{(1)}^{-1}$ is that matrix which, by the orthogonal transformation which brings $\boldsymbol{\mu}$ into the form

$$\left\|\begin{matrix} 0 & 0 & \dots & 0 \\ 0 & \mu_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mu_K \end{matrix}\right\|, \text{ itself assumes the form } \left\|\begin{matrix} 0 & 0 & \dots & 0 \\ 0 & 1/\mu_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/\mu_K \end{matrix}\right\|. \tag{22}$$

For $\boldsymbol{\gamma}$, being the characteristic vector belonging to the characteristic value $\mu_1 = 0$ of $\boldsymbol{\mu}$, passes by that transformation into a vector with components

$$(\gamma_\lambda\gamma_\lambda)^{\frac{1}{2}},\ 0,\ \dots\dots\ 0, \tag{23}$$

whence the two matrices whose product consitutes the left hand side in (21) assume forms differing from those in (22) only in that the 1-1-elements are 1 instead of 0. Evidently, $\boldsymbol{\mu}_1^{-1}$ is also symmetrical, and of rank $K-1$. Moreover, it can be seen from (23) and (22) that, besides (11), also

$$\mu_{(1)}^{k\lambda}\,\gamma_\lambda = 0. \tag{24}$$

Consequently, we find from (21)

$$\mu_{(1)}^{k\varkappa}\,\mu_{\varkappa l} = \delta_{kl} - \frac{\gamma_k\,\gamma_l}{\gamma_\lambda\,\gamma_\lambda}\,, \tag{25}$$

whence, in particular,

$$\mu_{(1)}^{k\varkappa}\,\mu_{\varkappa\lambda}\,\mu_{(1)}^{\lambda l} = \mu_{(1)}^{kl}\,. \tag{26}$$

Now, by premultiplying (12) by $\mu_{(1)}^{-1}$ and using (25),

$$c'_k - \frac{\gamma_\varkappa\, c'_\varkappa}{\gamma_\lambda\,\gamma_\lambda}\,\gamma_k = -\,\mu_{(1)}^{k\varkappa}\, m'_{\varkappa\lambda}\,\gamma_\lambda\,.$$

The left hand member in this equation is not changed by addition of an arbitrary vector $\varphi\boldsymbol{\gamma}$ to c'. Hence, as

$$c'_k = -\,\mu_{(1)}^{k\varkappa}\, m'_{\varkappa\lambda}\,\gamma_\lambda$$

is the particular solution satisfying $\gamma_\varkappa c'_\varkappa = 0$ (according to (24)), the general solution is

$$c'_k = -\,\mu_{(1)}^{k\varkappa}\, m'_{\varkappa\lambda}\,\gamma_\lambda + \varphi\gamma_k\,. \tag{27}$$

Only the first term of the expression (6) for m'_{kl} enters into (27), the second one vanishing when summed over λ:

$$c'_k = -\,\mu_{(1)}^{k\varkappa}\,(\xi_\varkappa^{(\tau)} - \bar{\xi}_\varkappa)\,u^{(\tau)} + \varphi\gamma_k\,, \tag{28}$$

where the

$$u^{(t)} \equiv \boldsymbol{\gamma}\mathbf{z}^{(t)} \tag{29}$$

are spherically normal variables satisfying

$$E\,u^{(t)}u^{(s)} = \sigma^2\,.\,\boldsymbol{\gamma\epsilon\gamma}\,.\,\delta^{(ts)}. \tag{30}$$

Let

$$\boldsymbol{\theta}\mathbf{c}' = 0, \quad \text{where} \quad \boldsymbol{\theta\gamma} \neq 0, \tag{31}$$

be the general rule indicating the normalization of $\boldsymbol{\gamma} + \mathbf{c}'$ if that of $\boldsymbol{\gamma}$ is given; $\boldsymbol{\theta}$ may, again, be normalized by

$$\boldsymbol{\theta\gamma} = 1. \tag{32}$$

Then, (28) inserted into (31) yields

$$\varphi = \boldsymbol{\theta}\,\mu_{(1)}^{-1}(\boldsymbol{\xi}^{(\tau)} - \bar{\boldsymbol{\xi}})u^{(\tau)}. \tag{33}$$

Consequently, the c'_k are linear combinations of the normal variables $u^{(t)}$, and the joint distribution of any independent group of them is also normal. In particular,

$$Ec_{k'} = 0. \tag{34}$$

The variances and covariances of the c'_k will now be determined for two different special normalization rules for $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma} + \mathbf{c}'$. The first one

$$\gamma_\varkappa \gamma_\varkappa = 1, \qquad \gamma_\varkappa c'_\varkappa = 0, \qquad \text{so} \qquad \boldsymbol{\theta} = \boldsymbol{\gamma}, \tag{35}$$

is somewhat arbitrary. It may be derived from the requirement that, in first approximation, also $\boldsymbol{\gamma} + \mathbf{c}'$ is normalized by

$$(\gamma_\varkappa + c'_\varkappa)(\gamma_\varkappa + c'_\varkappa) = 1$$

(neglecting the squares of the c'_k), and is, therefore, associated with the special units in which the X_k happen to be measured. Yet, the rule will be considered here since there can be attached to this type of normalization an instructive geometrical illustration which may help in grasping the essentials of the situation. According to (24), (33) and (35), $\varphi = 0$, and we find from (28) by means of (30), (2.2), (26)

$$E\, c'_k c'_l = \sigma^2 . \boldsymbol{\gamma} \boldsymbol{\epsilon} \boldsymbol{\gamma} . \mu^{kl}_{(1)} . \tag{36}$$

As the rank of $\boldsymbol{\mu}^{-1}_{(1)}$ is $K - 1$, only one linear relation — the normalization rule — exists between the c'_k in all samples.

Let $K = 3$, and consider a linear transformation $\boldsymbol{\xi} \to \boldsymbol{\xi}^*$ consisting of the above mentioned orthogonal transformation which makes

$$\boldsymbol{\mu}^* = \begin{Vmatrix} 0 & 0 & 0 \\ 0 & \mu_2 & 0 \\ 0 & 0 & \mu_3 \end{Vmatrix}, \tag{37}$$

followed by a translation owing to which

$$\bar{\boldsymbol{\xi}}^* = \mathbf{0}.$$

Since the components of $\boldsymbol{\gamma}^*$ are given by

$$\boldsymbol{\gamma}^* = (1, \quad 0, \quad 0), \tag{38}$$

the "true" regression plane Π is now $\xi^*_1 = 0$, and the shape of the swarm of scatter points $\boldsymbol{\xi}^{*(l)}$ within that plane is roughly indicated by the ellipse

$$\begin{cases} \xi^*_1 = 0\,, \\ \mu_2^{-1}\, \xi^{*2}_2 + \mu_3^{-1}\, \xi^{*2}_3 = 1\,, \end{cases} \tag{39}$$

whose axes have lengths $\mu_2^{\frac{1}{2}}$, $\mu_3^{\frac{1}{2}}$ and coincide with the ξ_2^*- and ξ_3^*-axis respectively. The direction of the perpendicular to the sample regression hyperplane P is approximated to by the direction of the radius vector connecting the new origin O* with the point C′ having new coordinates 1, $c_2'^*$, $c_3'^*$ and lying within the plane (Π_1) $\xi_1^* = 1$. The normal sampling distribution of the coordinates $c_2'^*$, $c_3'^*$ of this point is indicated by the ellipse

$$(40) \qquad \begin{cases} \xi_1^* = 1 \\ \mu_2 c_2'^{*2} + \mu_3 c_3'^{*2} = \sigma^2 . \gamma\epsilon\gamma \end{cases}$$

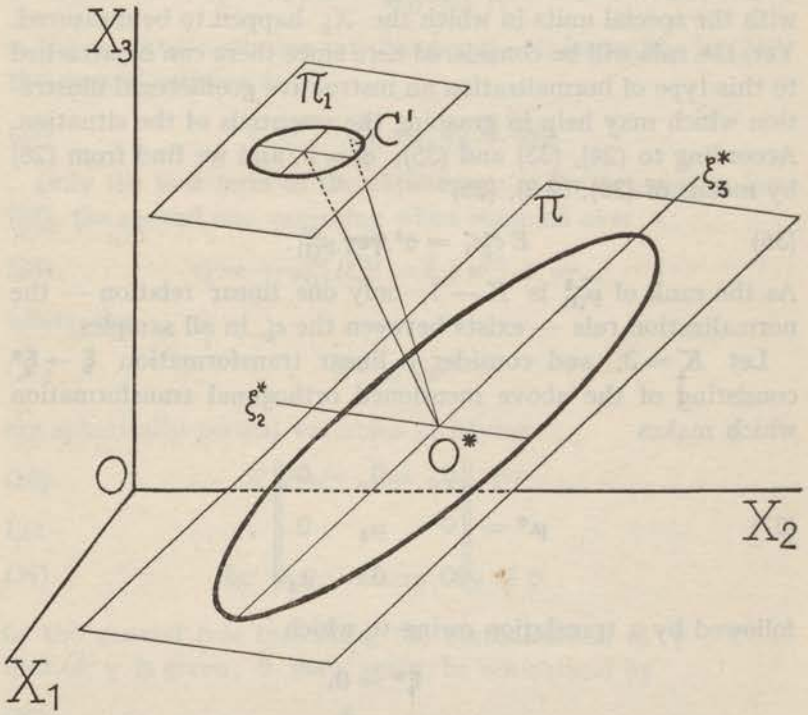

Figure 9.1. The large ellipse indicates the second degree moments of the scatter points $\xi^{(i)}$ in the plane Π, corresponding to the "true values" of the variables; the small ellipse indicates the second degree moments of the approximate sampling distribution of the components of the weighted regression vector $\overline{O^*C'}$ parallel to Π.

consisting of points of equal probability density, and having axes parallel in direction but inversely proportional in length to those of (39). Fig. 9.1 represents the situation of the planes Π and Π_1 in the original ξ-space. Fig. 9.2 is drawn in Π and shows the ellipse

(39) together with the orthogonal projection of the ellipse (40) on Π.

The absolute size of the ellipse (40) is governed by the factor $\sigma^2.\boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma}$ which, according to (38), equals

$$\sigma^2.\boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma} = E(\boldsymbol{\gamma}\mathbf{z})^2 = Ez_1^{*2} . \tag{41}$$

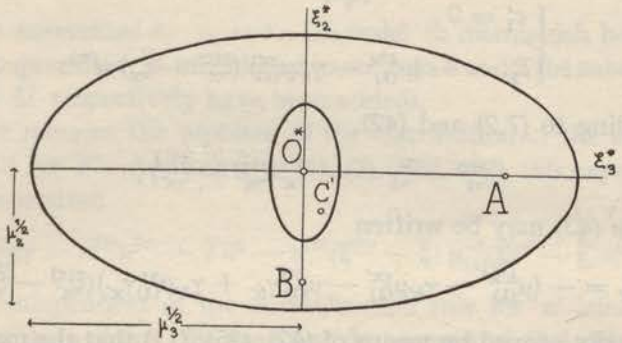

Figure 9.2. The ellipses of fig. 9.1, projected on Π.

Thus, as could be expected, *the only quantity in the distribution* (8.1) *of the errors* z_k *that matters for the first approximation terms to the sampling variances and covariances of the* c_k *is the variance of the component of the vector* $\mathbf{z}$ *in the direction of* $\boldsymbol{\gamma}$. Quite natural is also the inverse proportionality of the axes of the two ellipses. An erratic displacement $\mathbf{z}^{(t)}$ given to a point $\boldsymbol{\xi}^{(t)}$ located in A (fig. 9.2) will, by its larger distance from O*, have in the average a smaller influence on c_3' than a similar displacement to a point $\boldsymbol{\xi}^{(t)}$ located in B could exert upon $c_2'^*$. So, if the points $\boldsymbol{\xi}^{(t)}$ are more widely scattered in the direction of the ξ_3^*-axis than in that of the ξ_2^*-axis, the sampling variance of $c_2'^*$ will exceed that of $c_3'^*$.

Another normalization rule,

$$\gamma_1 + c_1' = \gamma_1 = 1, \quad \text{so} \quad c_1' = 0, \quad \boldsymbol{\theta} = (1, 0 \ldots . 0), \tag{42}$$

facilitates comparison with the results of section 3. The remaining coefficients $\gamma_{k'} + c_{k'}'$ now approximate the $c_{k'}$ if these are also normalized [1]) by

$$c_1 = 1 .$$

[1]) In fig. 9.1 the point $(1, \gamma_2 + c_2', \gamma_3 + c_3')$ is now represented by the intersection of the plane $\xi_1 = 1$ and a line drawn through O parallel to O*C'. Therefore, the use of this rule presupposes that $\gamma_1/(\gamma_\kappa\gamma_\kappa)^{\frac{1}{2}}$ is so large that samples in which this line passes the plane (or, in general, the hyperplane) $\xi_1 = 0$ occur only with a negligible probability. X_1 being the dependent variable, this condition will be satisfied in most of the applications.

If (42) holds, (33) becomes

$$\varphi = \mu_{(1)}^{1\varkappa}\,(\xi_{\varkappa}^{(\tau)} - \bar{\xi}_{\varkappa})\,u^{(\tau}$$

and (28) passes into

$$(43)\qquad \begin{cases} c_1' = 0\,, \\ c_{k'}' = -\,(\mu_{(1)}^{k'\varkappa} - \gamma_{k'}\mu_{(1)}^{1\varkappa})\,(\xi_{\varkappa}^{(\tau)} - \bar{\xi}_{\varkappa})\,u^{(\tau)}\,. \end{cases}$$

According to (7.2) and (42),

$$\xi_1^{(t)} - \bar{\xi}_1 = -\,\gamma_{\varkappa'}(\xi_{\varkappa'}^{(\tau)} - \bar{\xi}_{\varkappa'})\,,$$

whence (43) may be written

$$(44)\quad c_{k'}' = -\,(\mu_{(1)}^{k'\varkappa'} - \gamma_{k'}\mu_{(1)}^{1\varkappa'} - \mu_{(1)}^{k'1}\gamma_{\varkappa'} + \gamma_{k'}\mu_{(1)}^{11}\gamma_{\varkappa'})(\xi_{\varkappa'}^{(\tau)} - \bar{\xi}_{\varkappa'})u^{(\tau)}\,.$$

It is easily proved by means of (42), (25), (11) that the matrix of order $K-1$ in the right hand member of (44) is inverse to $\|\,\mu_{k'l'}\,\|$ and therefore equals

$$(45)\qquad \mu_{(1)}^{k'l'} - \gamma_{k'}\mu_{(1)}^{1l'} - \mu_{(1)}^{k'1}\gamma_{l'} + \gamma_{k'}\mu_{(1)}^{11}\gamma_{l'} = \mu_{(11)}^{k'l'}$$

as defined by (3.33). Thus,

$$(46)\qquad c_{k'}' = -\,\mu_{(11)}^{k'\varkappa'}(\xi_{\varkappa'}^{(\tau)} - \bar{\xi}_{\varkappa'})u^{(\tau)}\,,$$

and hence

$$(47)\qquad \boxed{Ec_{k'}'c_{l'}' = \sigma^2\,.\,\boldsymbol{\gamma\epsilon\gamma}\,.\,\mu_{(11)}^{k'l'}\,.}$$

In the special case

$$(48)\qquad \varepsilon_{11} = 1,\quad \varepsilon_{kl} = 0 \text{ unless } k = l = 1,\quad \text{so}\quad z_{k'}^{(t)} = 0,$$

$u^{(t)}$ becomes, according to (29), identical with the error $z^{(t)}$ in the dependent variable, introduced in section 3. Then, (46) passes into (3.41), (47) into (3.32). In this case, therefore, the first approximations $\gamma_k + c_k'$ are exactly equal to the coefficients c_k; their distribution was already found in section 3.

In the general case, (47) as compared with (3.32) shows that *the variances and covariances of the normally distributed first approximations c_k' to the sampling errors in the weighted regression coefficients c_k in sampling from the parent distribution as specified in section 7 are proportional to those of the elementary regression*

coefficients in sampling from the parent distribution according to Fisher's specification, with a proportionality factor

$$\gamma\epsilon\gamma \cdot \left(\frac{\sigma_U}{\sigma_F}\right)^2 \tag{49}$$

if γ *is normalized by* $\gamma_1 = 1$. (In order to distinguish between the two quantities σ introduced in sections 3 and 7 the subscripts F and U respectively have been added).

There remains the problem of the distribution of the expression (20) for l''. According to (6), (7), (28), (29), this expression may be written

$$l''\gamma\epsilon\gamma = u^{(\tau)}u^{(\tau)} - T\bar{u}^2 - u^{(\tau)}(\xi^{(\tau)} - \bar{\xi})\,\mu_{(1)}^{-1}(\xi^{(\sigma)} - \bar{\xi})u^{(\sigma)} \tag{50}$$

and is independent of the normalization rule for $\mathbf{c}'$ since the terms with φ from (28) drop out on account of (24). The sum

$$S \equiv u^{(\tau)}u^{(\tau)}$$

differs only by the constant factor $\gamma\epsilon\gamma$ from the sum S_I in (8.15). This part of the total variance, entering there into the estimation of σ^2, now splits up into three parts

$$\begin{cases} S &= S_1 + S_2 + S_3, \quad \text{where} \\ S_p &= u^{(\tau)}v_p^{(\tau\sigma)}u^{(\sigma)}, \qquad p = 1, 2, 3, \\ v_2^{(ts)} &= (\xi^{(t)} - \bar{\xi})\,\mu_{(1)}^{-1}(\xi^{(s)} - \bar{\xi}), \\ v_3^{(ts)} &= \frac{1}{T}\,\delta^{(ts)}, \end{cases} \tag{51}$$

which may be treated in exactly the same way as the sums in (3.37). Consequently,

$$s''^2 \equiv \frac{l''}{T - K} \tag{52}$$

is an unbiased "estimate" of σ^2, *distributed according to* (3.48) *with* $N = T - K$, *independently of* $\mathbf{c}'$ *and* c. *It can be approximated to by the estimate*

$$\boxed{s^2 \equiv \frac{l}{T - K}} \tag{53}$$

which may be computed from the data.

Since s''^2 is the mean of $T - K$ squares of spherically normal variables of variance σ^2,

$$E(s''^2 - \sigma^2)^2 = \frac{2\sigma^4}{T - K}. \tag{54}$$

Here, instead of 2, v a n U v e n [1]) (**36**) obtains (form. (*55*)) 4 for the numerical factor in the right hand member. The way in which this difference comes in is not simple. V a n U v e n computes variances and covariances of the errors in the various estimates, these errors being denoted by writing Δ before the symbol denoting the estimate. As v a n U v e n uses the same symbols for parameters and for the corresponding statistics, a comparison of his formulae with those of this section is only possible up to that order of magnitude which the former are intended to cover. If I have not misunderstood his argument, the comparison runs as follows. The starting point is the formula at the top of p. 151, which, in our notation, runs

$$\Delta \mathbf{m} = \mathbf{m}', \tag{55}$$

$\mathbf{m}''$ being neglected as a second order quantity. This neglect of $\mathbf{m}''$ should lead to

$$\Delta l = l' = \boldsymbol{\gamma}\mathbf{m}'\boldsymbol{\gamma} = 0,$$

showing that second order quantities need to be considered for a study of the sampling variance of l. This does, however, not appear from v a n U v e n's formula (*48*), which, owing to (*33*), may be written

$$\mathbf{c}\epsilon\mathbf{c} \,.\, \Delta l = \mathbf{c}\mathbf{m}'\mathbf{c}, \tag{56}$$

because $\boldsymbol{\gamma}$ has been replaced by $\mathbf{c}$. By that, as (56) can be developed into

$$\mathbf{c}\epsilon\mathbf{c} \,.\, \Delta l = 0 + 2\,\boldsymbol{\gamma}\mathbf{m}'\mathbf{c}' + \ldots,$$

Δl is again made to differ from zero by an amount which has no relation to the sampling error in l, and which is of the same order of magnitude as the expression (20) for l'' which, for neglect of $\mathbf{m}''$ in (55), could not appear in v a n U v e n's formulae.

From (55) one would not expect the final result to differ only by a factor 2 from (53). This is caused by another error of the same kind: in (*49*) the missing terms containing $\mathbf{m}''$ have been introduced by replacing $\boldsymbol{\mu}$ by $\mathbf{m}$ in expressions of the type $\mu_{k\lambda}c_\lambda$.

The formulae (*51*) and (*55*) being erroneous, the resulting expressions for the sampling variances of the c_k do not apply. It should be stressed that the approximate computation of sampling variances by means of simple variational calculus is bound to fail if applied to statistics which have the property that, in their development into powers of the elementary observational errors z_k, the terms of first degree are missing — as is the case with l. Such a situation cannot

---

[1]) Formulae from v a n U v e n's paper will be quoted in italics.

pass unnoticed when the distinction between parameters and statistics is explicitly introduced into the formulae.

A reconsideration of the conditions for a rapid convergence in the developments of $\mathbf{c}$ and l is now due. If the same method of solution by which $\mathbf{c}'$ was found is also applied to terms of higher orders, the components of the p-th order term $\mathbf{c}^{(p)}$ become, indeed, as was foreseen on p. 73, homogeneous functions, on the one hand, of the differences $\xi_k^{(t)} - \bar{\xi}_k$, with a degree $-p$, on the other, of the $z_k^{(t)}$, with a degree p; a similar situation prevails with regard to the higher terms in l. The condition (17) is, therefore, certainly necessary for a rapid convergence (though, apparently, only with $k > 1$ if the normalization rule (42) is applied to terms of every order). Owing to the special way in which the $\xi_k^{(t)} - \bar{\xi}_k$ occur in $c_k^{(p)}$, however, this condition is not sufficient. In fact, some terms in $c_k^{(p)}$ are of degree p in the elements of $||\mu_{(11)}^{k'l'}||$ (or of $||\mu_{(1)}^{kl}||$, according to the normalization rule), the balance being restored by a factor of the same degree in the $\xi_k^{(t)} - \bar{\xi}_k$ themselves. Because of that, a rapid convergence in the series can be expected only unless, for any value of k',

$$\mu_{k'k'}\,\mu_{(11)}^{k'k'} \gg 1 .$$

According to (3.33), this inequality can also be written

$$\mu_{k'k'}\,\mathcal{M}_{11.k'k'} \gg \mathcal{M}_{11} . \tag{57}$$

This means that our method of approach breaks down in cases in which the systematic components of the determining variables themselves are nearly linearly dependent. The more such a situation is approached, the larger the number of terms that must be considered. And, in order to get an idea of the rapidity of convergence of the developments, it is better, instead of relying on (17), to see whether

$$T\sigma^2\,\varepsilon_{k'k'}\,\mu_{k'k'}(\mu_{(11)}^{k'k'})^2 \ll 1 \tag{58}$$

is satisfied for every value of k'.

This modification being accepted, a remark made on p. 74 regarding the influence of large erratic components on the adequacy of the present method can be supplemented by a similar remark concerning the influence of mutual linear dependence of determining variables: the less the sampling varian-

ces of the $c'_{k'}$, as given by (47), are increased by intercorrelation of determining variables, the better these quantities approximate to the total sampling variances of the c'_k.

We are now in a position to indicate two of the *three causes*, mentioned at the end of section 3, by which the use of formula (3.61), criticized by Frisch, leads to an *underestimation* of the real sampling variance of the regression coefficients in cases where the present specification of the parent distribution is a better approximation to reality. The first cause relates to the factor of proportionality (49). σ_F^2 is in (3.61) unbiasedly estimated by s_F^2, which, in virtue of (3.20) and (3.46), equals

$$s_F^2 = \frac{\mathbf{bmb}}{T-K}. \tag{59}$$

According to (53),

$$\sigma_U^2 \boldsymbol{\gamma\epsilon\gamma} \tag{60}$$

is estimated by

$$\frac{l}{T-K}\,\mathbf{c\epsilon c}, \tag{61}$$

or, owing to (1), by

$$\frac{\mathbf{cmc}}{T-K}. \tag{62}$$

The first term in the development

$$l\,.\,\mathbf{c\epsilon c} = l''\,.\,\boldsymbol{\gamma\epsilon\gamma} + \ldots.$$

of the numerator in (61) yields, if divided by $T-K$, an unbiased "estimate" of (60). Neither is a bias introduced by the terms of one order higher, which, being of degree 3 in the $z_k^{(t)}$, have mean values equal to zero, and, therefore, only affect — probably enlarge — the sampling variance of (61). Thus, the bias in (61) as an estimate of (60), to which the first contribution can be expected only from the fourth order terms, will, if (58) holds, be small compared with the estimate itself, the latter being a quantity of the second order.

Therefore, a comparison of (59) and (62) may give an idea of the proportionality factor (49). To that end, we observe that the minimum of the quadratic form

$$\mathbf{pmp} = p_\varkappa m_{\varkappa\lambda} p_\lambda,$$

where $p_1 = 1$, is reached for

$$p_{k'} = \frac{M_{1k'}}{M_{11}} = b_{k'} .$$

Therefore, as soon as $\mathbf{c}$ differs from $\mathbf{b}$, (59) falls short of (62); at present, we may content ourselves with the statement that, accordingly, the factor s^2 in (3.61) makes for a tendency to underestimate the sampling variances of the $c_{k'}$. The quantitative importance of this tendency, as indicated by the ratio

$$\frac{\mathbf{cmc}}{\mathbf{bmb}}, \tag{63}$$

depends on $\boldsymbol{\epsilon}$, and will be studied in section 13 when more is known about the range of variation of $\mathbf{c}$ when $\boldsymbol{\epsilon}$ is allowed to assume different values. It will turn out to be the more important, the larger the spread in the elementary regressions.

A second cause [1]) which makes (3.61) unreliable under the present conditions is the use of

$$m_{(11)}^{k'k'} \equiv \frac{M_{11.k'k'}}{M_{11}} \tag{64}$$

as an estimate of

$$\mu_{(11)}^{k'k'} \equiv \frac{\mathcal{M}_{11.k'k'}}{\mathcal{M}_{11}} . \tag{65}$$

Like the first cause, its influence is especially large when the determining variables are highly intercorrelated. If (57) holds for some value of k', the first term in the development

$$M_{11} = \mathcal{M}_{11} + (m_{\varkappa'\lambda'} - \mu_{\varkappa'\lambda'})\, \mathcal{M}_{11.\varkappa'\lambda'} + \ldots . \tag{66}$$

need not to be much larger than the terms of one order higher. M_{11} is in such cases considerably influenced by the errors $z_k^{(t)}$, which in itself constitutes a reason to suspect (3.61), because the denominator M_{11} of (64) will then be subject to considerable variation from one sample to another. No less serious is the fact that, owing to these influences, M_{11} is biased as an estimate of $\mathcal{M}_{11}$. Unless, for any value of k',

$$m_{k'k'}\, M_{11.k'k'} \ll m_{22} m_{33} \ldots . m_{KK} , \tag{67}$$

[1]) Readers acquainted with Frisch's work (**13**, **14**, **16**) will recognize his argument in what follows.

that is, if there is only one linear relation approximately satisfied by the determining variables but not by any subset of these variables, the remaining terms in (66) can be neglected under the condition (17). The mean value of M_{11} then exceeds $\mathcal{M}_{11}$ approximately by the amount

$$\mathcal{M}_{11.\varkappa\lambda'} E(m_{\varkappa\lambda'} - \mu_{\varkappa\lambda'}) = (T-1)\sigma^2 \mathcal{M}_{11.\varkappa\lambda'} \varepsilon_{\varkappa\lambda'}$$

(owing to (5), (6), (7)), which must be positive as it is at the same time the mean value of the positive definite form

$$\frac{T-1}{T} z_{\varkappa'}^{(\tau)} \mathcal{M}_{11.\varkappa\lambda'} z_{\lambda'}^{(\tau)};$$

on the other hand $M_{11.k'k'}$ approximately equals $\mathcal{M}_{11.k'k'}$. So (64) is biased as an estimate of (65), approximately underestimating it in the average in the ratio

$$\frac{1}{1 + (T-1)\sigma^2 \mu_{(11)}^{\varkappa\lambda'} \varepsilon_{\varkappa\lambda'}}. \tag{68}$$

If, however, (67) holds for at least one value of k', the situation is still worse. Then, for such a value of k', both numerator and denominator of (64) are biased and are determined by the errors $z_k^{(t)}$ more than by the quantities they should estimate. Consequently (3.61) has lost all value as a source of information regarding the reliability of the c_k.

Both of these effects of the errors in the determining variables on the sampling variances of the weighted regression coefficients gain in importance the more the determining variables approach mutual linear dependence. The same holds for a third unfavourable effect of errors in the determining variables on the reliability of estimated regression coefficients, which will be studied in the next section. Each of these effects would have been sufficient to justify Frisch's criticism of the arbitrary use of (3.61) in economic analysis.

It has already been pointed out, however, that the approximations, by means of which these effects are found, are the more inaccurate, the more the same critical limiting case is approached. For that reason, it should be emphasized that the results, derived in this investigation, constitute only a first step in the correction of (3.61) for cases in which the present specification applies. The more important this correction becomes owing to interrelated

determining variables, the more desirable is a completion of the present results by an actual study of the terms of higher order. Such a study would, moreover, supply a welcome substitute for the rather provisional and superficial discussion of the conditions for a rapid convergence in the developments given in this section.

A final remark may be made on the circumstance that, if the method of solution of this section is also used to express in terms of the $z_k^{(t)}$ higher degree terms $\mathbf{c}^{(p)}$, $l^{(p)}$ in the developments of $\mathbf{c}$ and l, only the second degree moments μ_{kl} of the systematic components $\xi_k^{(t)}$ occur. It would be erroneous to conclude from this fact that the sampling distributions of these higher terms — or even those of $\mathbf{c}$ and l themselves — depend only on the μ_{kl}. It is, for instance, easily seen that, in general, in the mathematical expectations of terms $\mathbf{c}^{(p)}$ and l of even degree in the $z_k^{(t)}$, moments of the same degree in the $\xi_k^{(t)} - \bar{\xi}_k$ will occur.

10. *The error of weighting.*

Hitherto we have assumed that the matrix $\boldsymbol{\epsilon}$ was known a priori. We are now confronted with the difficulty that, on the one hand, this assumption is not satisfied in most of the applications, and, on the other, we did not find means to estimate $\boldsymbol{\epsilon}$ from the data.

In the Inductive Part, in section 13, this problem will be considered more in detail. At present, it is already possible to detect one of the features which are relevant to this problem, by means of the method of approximation introduced in the preceding section. Besides (9.1), (9.2), (9.3), which define the *correctly weighted* regression coefficients $\mathbf{c}$, we consider similar equations obtained by replacing the unknown matrix $\boldsymbol{\epsilon}$ by an assumed matrix $\mathbf{e}$ which also satisfies

$$\boldsymbol{\gamma}\mathbf{e}\boldsymbol{\gamma} \neq 0. \tag{1}$$

These equations are

$$|\mathbf{m} - l(\mathbf{e}) \,.\, \mathbf{e}| = 0, \tag{2}$$

$$[m_{k\lambda} - l(\mathbf{e}) \,.\, e_{k\lambda}]\, c_\lambda(\mathbf{e}) = 0, \tag{3}$$

$$c_\varkappa(\mathbf{e})\bar{X}_\varkappa - c(\mathbf{e}) = 0. \tag{4}$$

$l(\mathbf{e})$ is again the smallest root of (2), and is supposed to be single,

whence $\mathbf{c}(\mathbf{e})$ is uniquely defined but for normalization. Then

$$l(\boldsymbol{\epsilon}) = l, \quad \mathbf{c}(\boldsymbol{\epsilon}) = \mathbf{c}, \tag{5}$$

if the latter quantities are normalized by the same rule. The $c_k(\mathbf{e})$ will be called the *tentatively weighted* regression coefficients. Still considering them as estimates of the γ_k we may study how their distribution depends on the unknown matrix $\boldsymbol{\epsilon}$ and on the assumed matrix $\mathbf{e}$. In particular, we are interested in the size of the bias in $\mathbf{c}(\mathbf{e})$ due to the more or less arbitrary choice of $\mathbf{e}$. This bias will be called the *error of weighting*.

The distribution of $l(\mathbf{e})$ and $\mathbf{c}(\mathbf{e})$ can be studied by the same procedure as was applied in the preceding section. Indeed, the equations (9.8) to (9.16), (9.19), (9.20), (9.27), (9.28), (9.34), can be repeated for $l(\mathbf{e})$ and $\mathbf{c}(\mathbf{e})$ with only $\boldsymbol{\epsilon}$ replaced by $\mathbf{e}$. Changed in this way, these equations will be quoted as (9.8*e*) etc. So, in particular, according to (9.27*e*), *the first order term* $\mathbf{c}'(\mathbf{e})$ *in the error of sampling in* $\mathbf{c}(\mathbf{e})$ *is independent of the assumed matrix* $\mathbf{e}$ *by means of which* $\mathbf{c}(\mathbf{e})$ *is computed*. According to the equation

$$Ec'_k(\mathbf{e})c'_l(\mathbf{e}) = \sigma^2 \cdot \boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma} \cdot \mu^{kl}_{(1)} \tag{6}$$

corresponding to (9.36), the sampling variances of the $c_k(\mathbf{e})$ depend only, in first approximation, on the unknown matrix $\boldsymbol{\epsilon}$. Further, by a comparison of (9.20) and (9.20*e*).

$$l''(\mathbf{e}) \cdot \boldsymbol{\gamma}\mathbf{e}\boldsymbol{\gamma} = l'' \cdot \boldsymbol{\gamma}\boldsymbol{\epsilon}\boldsymbol{\gamma}. \tag{7}$$

Evidently, terms of higher order in the developments must be considered in order to find the above mentioned bias. In the same way in which (9.27) was obtained from (9.12) we may get from (9.13*e*)

$$c''_k(\mathbf{e}) = \mu^{k\varkappa}_{(1)}[-m'_{\varkappa\lambda}\, c'_\lambda(\mathbf{e}) - m''_{\varkappa\lambda}\,\gamma_\lambda + l''(\mathbf{e}) \cdot e_{\varkappa\lambda}\,\gamma_\lambda] + \psi\gamma_k, \tag{8}$$

or, expressed in terms of the $z^{(t)}_k$ by means of (9.6), (9.7), (9.27),

$$\left\{\begin{aligned} c''_k(\mathbf{e}) &= \mu^{k\varkappa}_{(1)}[(\xi^{(\tau)}_\varkappa - \bar{\xi}_\varkappa)z^{(\tau)}_\lambda + z^{(\tau)}_\varkappa(\xi^{(\tau)}_\lambda - \bar{\xi}_\lambda)]\cdot \\ &\cdot [\mu^{\lambda\nu}_{(1)}(\xi^{(\sigma)}_\nu - \bar{\xi}_\nu)z^{(\sigma)}_\mu\gamma_\mu - \varphi\gamma_\lambda] - \mu^{k\varkappa}_{(1)}(z^{(\tau)}_\varkappa - \bar{z}_\varkappa)z^{(\tau)}_\lambda\gamma_\lambda + \\ &+ l''(\mathbf{e}) \cdot \mu^{k\varkappa}_{(1)}\, e_{\varkappa\lambda}\gamma_\lambda + \psi\gamma_k, \end{aligned}\right. \tag{9}$$

where $l''(\mathbf{e})$ can also be written, by means of (7) and (9.50), remembering (9.29), as a quadratic expression in the $z^{(t)}_k$. In order to avoid unessential complications, we shall take

$$\varphi = \psi = 0,$$

that is, we shall adopt the normalization rules

$$\gamma_\varkappa c'_\varkappa(\mathbf{e}) = 0, \quad \gamma_\varkappa c''_\varkappa(\mathbf{e}) = 0, \tag{10}$$

simply because they are most suited to our present purpose. Then, the mean value of $c''_k(\mathbf{e})$ can easily be found from (9) by some calculations, using (2.2), (9.25) and (9.26). The result is

$$\boxed{Ec''_k(\mathbf{e}) = (T-K)\sigma^2 \cdot \boldsymbol{\gamma\epsilon\gamma} \cdot \left(\frac{\mu^{k\varkappa}_{(1)} e_{\varkappa\lambda}\gamma_\lambda}{\boldsymbol{\gamma e\gamma}} - \frac{\mu^{k\varkappa}_{(1)}\varepsilon_{\varkappa\lambda}\gamma_\lambda}{\boldsymbol{\gamma\epsilon\gamma}} \right) - \sigma^2 \mu^{k\varkappa}_{(1)}\varepsilon_{\varkappa\lambda}\gamma_\lambda.} \tag{11}$$

The first term vanishes if $\mathbf{e} = \boldsymbol{\epsilon}$. The second term, then remaining, equals Ec''_k. If the conditions for a rapid convergence in the series (9.9) are satisfied, it must, consequently, be a small bias [1]), negligible compared with the standard deviation

$$[E\{c'_k(\mathbf{e})\}^2]^{\frac{1}{2}} \tag{12}$$

of $c'_k(\mathbf{e})$, as given by (6). With regard to σ and the $\xi^{(t)}_k - \bar{\xi}_k$, the first term is of the same order of magnitude as the second one. Hence, it would have been equally negligible if the factor $T - K$ had not been present. Compared with (12), the right hand member of (11) is one order higher in the quantities [2])

$$(T\sigma^2 \varepsilon_{kk}\, \mu^{kk}_{(1)})^{\frac{1}{2}}, \tag{13}$$

but one order lower in

$$T^{-\frac{1}{2}}.$$

In any applications, therefore, *the relative importance of the error of weighting compared with the error of sampling depends on the circumstances of the problem.* It follows from (11) and (6) that *the error of weighting is, both absolutely and relatively to the error*

---

[1]) In fact, the second term is due to the arbitrary normalization rules (10), and would not have appeared if, instead of (10), the rules

$$\boldsymbol{\gamma\epsilon c'}(\mathbf{e}) = 0, \qquad \boldsymbol{\gamma\epsilon c''}(\mathbf{e}) = 0,$$

had been adopted.

[2]) The difference between (13) and the square root of the left hand member in (9.58) is due to the fact that (9.58) has been written down in connection with (9) (for $\mathbf{e} = \boldsymbol{\epsilon}$), whereas, in the derivation of (11) from (9), one degree in the $\mu^{kl}_{(1)}$ has cancelled against the μ_{kl} on account of (9.25).

of sampling, the smaller, the less the systematic components of the determining variables are intercorrelated, and the smaller the part of the variances of the variables X_k *due to the erratic components. On the other hand, a large number of observations reduces the error of sampling, absolutely, and relatively to the error of weighting.* If the number of observations could be increased beyond limit in such a way that the matrix μ/T would tend to a limit matrix of rank $K - 1$, the error of sampling would tend to zero, while the error of weighting would approach a finite limit, depending on ϵ and $\mathbf{e}$.

Owing to (9.34*e*), the mean value $E\mathbf{c}(\mathbf{e})$ of the tentatively weighted regression vector is approximated to by

$$\gamma + E\mathbf{c}''(\mathbf{e}) \tag{14}$$

up to second order terms in the quantities (13). Neglecting the second term in (11), we shall study the range of variation of the vector (14) if $\mathbf{e}$ is allowed to equal any positive definite diagonal matrix

$$|| e_{kl} || = || e_k \delta_{kl} ||, \qquad e_k \geq 0. \tag{15}$$

If ϵ satisfies a similar condition, that is, if in the parent distribution the erratic components of different variables are independently distributed, the approximated mean value

$$\gamma + E\mathbf{c}''$$

of the correctly weighted regression vector $\mathbf{c}$ must be included within this range.

Let $\gamma_k \neq 0$ for $k = 1, 2 \ldots K$ and let, if necessary, the signs of the variables X_k be changed in such a way that

$$\gamma_k > 0. \tag{16}$$

The vector η with components

$$\eta_k \equiv \frac{e_k \gamma_k}{\gamma \mathbf{e} \gamma} \tag{17}$$

then satisfies both

$$\gamma \eta = 1 \tag{18}$$

and

$$\eta_k \geq 0, \tag{19}$$

and, therefore, connects the origin O with a point H lying in that part of the hyperplane (18) which lies within or on the borders of the angle of K-dimensional space indicated by (19) — one of the 2^K similar angles into which the entire space is divided by the coordinate hyperplanes. For $K = 3$ (fig. 10.1), H lies within or on the border of the triangle $H_1H_2H_3$; in general H

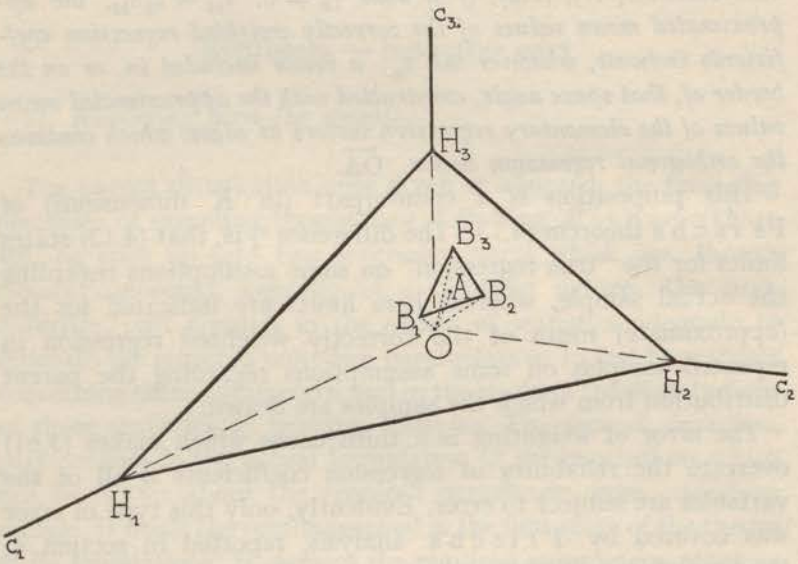

Figure 10.1. Limits for the mean value of the weighted regression vector if errors in different variables are uncorrelated in the parent distribution.

lies within or on the border of a multiplex extending into $K - 1$ dimensions and having edge points $H_1H_2 \ldots H_K$, one on every positive coordinate axis. Because of that, the vector (14) whose components are linear functions of the η_k connects O with a point C within or on the border of a multiplex $B_1B_2 \ldots B_K$ into which $H_1H_2 \ldots H_K$ passes by a linear transformation. Owing to (10), this second multiplex is also a part of the hyperplane (18) if

$$\gamma_x\gamma_x = 1. \tag{20}$$

The n-th edge point H_n is obtained from (17) by putting

$$e_n = 1, \quad e_k = 0 \text{ for } k \neq n. \tag{21}$$

Consequently, the coordinates of B_n approximate to

$$Ec_k(e_n)$$

if $\mathbf{e}_n$ represents the matrix $\mathbf{e}$ in the special case defined by (14) and (21). Now it is easily seen from (2) that

$$\mathbf{c}(\mathbf{e}_n) \equiv \mathbf{b}^{(n)}$$

equals the n-th elementary regression vector of the sample. Thus, we reach the result that, *in sampling from the parent distribution* (7.1), (7.2), (7.4) *with* $\gamma_k \neq 0$, $\varepsilon_{kl} = \varepsilon_k \delta_{kl}$, *the approximated mean values of the correctly weighted regression coefficients indicate, whatever the* ε_k, *a vector included in, or on the border of, that space angle, constructed with the approximated mean values of the elementary regression vectors as edges, which contains the orthogonal regression vector* $\overline{OA}$.

This proposition is a counterpart (in K dimensions) of Frisch's theorem (4.13). The difference [1]) is, that (4.13) states limits for the "true regression" on some assumptions regarding the actual sample, whereas here limits are indicated for the (approximate) mean of the correctly weighted regression in repeated samples on some assumptions regarding the parent distribution from which the samples are drawn.

The error of weighting is a third cause which makes (3.61) overrate the reliability of regression coefficients if all of the variables are subject to error. Evidently, only this type of error was covered by Frisch's analysis, reported in section 4, the error of sampling being eliminated ab initio by the assumption that the erratic components are uncorrelated in the actual sample. Thus, if a close fit is reached by means of not considerably intercorrelated determining variables, where the number of observations is small, the significance factors (4.21), proposed by Frisch, will yield too favourable an impression of the reliability of the regression coefficients, the sampling error dominating the error of weighting. On the other hand, if T is large, the error of weighting may be the most important feature in assessing the precision of the results.

---

[1]) The extension to K dimensions leads to a slight correction on the anticipation of Frisch in (4.17), that only $b_{kl}^{(k)} \equiv -b_l^{(k)}/b_k^{(k)}$ and $b_{kl}^{(l)}$ determine the range of variation of the intercoefficient $c_{kl} \equiv -c_l/c_k$; in fact, this range extends between the largest and smallest of all $b_{kl}^{(n)}$ for $n = 1, 2 \ldots . K$. The practical importance of this modification is small, as, in most cases, of all $b_{kl}^{(n)}$ just $b_{kl}^{(l)}$ will be largest, $b_{kl}^{(k)}$ smallest in absolute value.

PART IV

Synthesis — inductive part

11. *Reasoning from the sample.*

The parent distribution once given or adopted, the reasoning processes of sampling theory may, following Fisher (**11**, p. 40), be split up into two successive parts which are, though closely connected, logically of a different nature. The first, *deductive*, part consists in the choice of statistics adequate to estimate the required unknown parameters or to test hypotheses concerning these parameters, and in the study of the distributions of these statistics in repeated samples. The second, *inductive*, part consists in the actual formulation of the conclusions which can be drawn about the required parameters from one given sample. It needs the results reached in the first stage of the theory as its foundations. If, besides the required parameters, other — *irrelevant* — parameters figure in the parent distribution, the inductive inference can, if possible, best be based on those results of deductive theory which concern "studentized" statistics, that is, statistics, which, together with their distribution function, depend on the sample and on the required parameters only. As an example, the formulation of the t-test (3.54), used in section 3 in testing assumed values of unknown regression coefficients, rests upon the distribution (3.53) of the "studentized" statistic (3.52).

The logical problems involved in the inductive reasoning have only been touched upon in this study in sections 3 and 8. Nor will they be considered more in detail in this Part. Still the distinction between deductive and inductive theory has been adopted as a principle for arranging the sections 6—14 in two Parts, mainly so since the argument concerning the *error of weighting* must be thrown into a new form if it is to involve only quantities computed from the sample.

Owing to the presence of the parameters ϵ which are not estimated, the inductive argument is, similarly to the Deductive Part, developed in two steps. In section 12 there is indicated, on the assumption of an a priori given matrix ϵ, a generalization of the t-test of section 3 which is shown to have approximate validity. This section, therefore, merely contains considerations pertaining to that part of deductive theory on which the inductive inference can be based so long as ϵ is known. Section 13, finally, rigorously considers the sample as only source of information concerning the parameters. It will then be seen that, though we decided in section 8 not to attempt an estimation of ϵ, there will, nevertheless, be frequently occasion *to draw inferences regarding* ϵ *from the sample*, and that from a perhaps unexpected consideration.

The specification of the parent distribution, given in section 7, was preceded by a discussion of the special exigencies of the field of application aimed at in this study. Later sections have only been concerned with the mathematical consequences of that specification. The inference regarding ϵ, to be drawn from the sample will, however, be derived not as a consequence of the mathematical form of the parent distribution but from a consideration recalling the discussion of the general setting of the problem of regression analysis of economic time series, on which that specification was also based.

Two restrictions will in that section be introduced in order to facilitate mathematical treatment. In the first place, we shall assume, as in section 10, *that the errors* z_k *in different variables are uncorrelated in the parent distribution*. It appears that, in most cases, this restricting assumption will not seriously affect the trustworthiness of the results to which it leads. We must be prepared, however, to meet with exception, as for instance in cases where different variables have been obtained by a division of different time series by the same price index number. Secondly, *we shall only consider cases where the signs of the coefficients in each of the* K *elementary regressions are compatible*, which will hold in most cases where a close fit has been obtained by means of a set of determining variables which all have a decidedly significant influence on the dependent variable.

12. *The t-test as an approximation.*

To apply formula (9.47) for the approximated sampling variance of the weighted regression coefficients one should know, besides $\boldsymbol{\epsilon}$, also $\boldsymbol{\mu}$, σ and, strictly speaking, even the γ_k themselves. There now arises the problem: what conclusions about $\boldsymbol{\gamma}$ can be drawn from the data if none of the above mentioned quantities is known?

In this section we shall still assume that $\boldsymbol{\epsilon}$ is known. In the next section, even this assumption will be dropped.

Only cases where

$$|\mathbf{m}| > 0 \tag{1}$$

need to be considered. For in the trivial case (which will, moreover, hardly ever occur in practice) that $\mathbf{m}$ is singular (of rank $K-1$) the sample regression plane gives a perfect fit to the scatter, while, since then $l = 0$, the estimated sampling variances of its coefficients are zero. And, if the rank of $\mathbf{m}$ is still lower, our problem entirely loses its meaning.

The condition (9.58) for the application of the present method must now be written in terms of estimates of the quantities occurring in (9.58):

$$\frac{T}{T-K}\, l\, \varepsilon_{k'k'}\, m_{k'k'} (m_{(11)}^{k'k'})^2 \ll 1. \tag{2}$$

The adequacy of $m_{(11)}^{k'k'}$ as an estimate of $\mu_{(11)}^{k'k'}$ will be considered below. Like $m_{k'k'}$ as an estimate of $\mu_{k'k'}$, its sampling variance and its bias are small just if (2) holds.

In that case, a fairly good estimate of (9.47) is given by

$$\text{est } Ec_{k'}'^2 = m_{(11)}^{k'k'} \frac{l}{T-K}\, \mathbf{c\epsilon c}, \tag{3}$$

if $T-K$ is sufficiently large. If, however, $T-K$ is small, it is more accurate to account for the sampling variance of l by considering the correspondingly approximated distribution of the ratio

$$t_{k'} \equiv \frac{c_{k'} - \gamma_{k'}}{\left(m_{(11)}^{k'k'} \dfrac{l}{T-K}\, \mathbf{c\epsilon c} \right)^{\frac{1}{2}}} \tag{4}$$

obtained, in the same way as (3.52), by dividing the differences $c_{k'} - \gamma_{k'}$ by the square roots of the estimates (3) of their sampling variances.

The first term in the development of (4) in powers of the $z_k^{(t)}$ is

$$t'_{k'} \equiv \frac{c'_{k'}}{\left(\mu_{(11)}^{k'k'} \frac{l''}{T-K} \gamma\epsilon\gamma\right)^{\frac{1}{2}}}. \tag{5}$$

It follows from (9.47) and from the analysis of variance based on (9.50) that $t'_{k'}$ is distributed according to the t-distribution (3.53) with $N \equiv T - K$. Now, if (2) is satisfied, *an approximate test of supposed values of* $\gamma_{k'}$ is *obtained by treating* $t_{k'}$, which only depends on the sample and on the value of $\gamma_{k'}$ to be tested, *as if it were itself a quantity distributed according to* (3.53), entering the table of t (as explained in section 3) with the value of $t_{k'}$ given by (4).

Some provisional remarks may be made on the accuracy of that procedure. Owing to (10.11) (with $\mathbf{e} = \epsilon$), only a small *bias* in $c_{k'}$ as an estimate of $\gamma_{k'}$ is due to the term $c''_{k'}$, which is of an order of magnitude $1/(T-K)$ relative to the error of weighting. Apart from that, the terms c''_k and c'''_k in the development of the numerator in (4) can only make for a minor addition to the sampling variance of $t_{k'}$ as compared with that of $t'_{k'}$. The same will be the effect of the terms of third and fourth degree in the $z_k^{(t)}$ in the development of the denominator in (4), where the fourth degree terms may be as important as those of third degree, since the fourth degree terms are likely to yield a bias in the denominator in (4) as an estimate of that in (5), to which a negative contribution can be expected from the second degree terms in $m_{(11)}^{k'k'}$.

A complete discussion of these effects would require the evaluation of l''' and l^{IV}, and the computation of mean values of a number of polynomials in the $z_k^{(t)}$. At present, I have no such results to offer. The situation may, however, be illustrated by a study of the first terms in the development of

$$m_{(11)}^{k'k'} \equiv \frac{M_{11.k'k'}}{M_{11}} \tag{7}$$

alone. The denominator expands into

$$(8)\qquad \begin{cases} M_{11} = \mathcal{M}_{11} + (m_{\varkappa'\lambda'} - \mu_{\varkappa'\lambda'})\,\mathcal{M}_{11.\varkappa'\lambda'} + \\ + \begin{vmatrix} m_{\varkappa'\lambda'} - \mu_{\varkappa'\lambda'} & m_{\varkappa'\pi'} - \mu_{\varkappa'\pi'} \\ m_{\nu'\lambda'} - \mu_{\nu'\lambda'} & m_{\nu'\pi'} - \mu_{\nu'\pi'} \end{vmatrix} \mathcal{M}_{11.\varkappa'\lambda'.\nu'\pi'} + \cdots, \end{cases}$$

where $\mathcal{M}_{11.k'l'.n'p'} = 0$ as soon as $k' = n'$ or $l' = p'$, and a similar development can be written for the numerator. Putting for simplicity $k' = 2$, it follows from (9.5) and (9.6) that, if the ensuing development of $m^{22}_{(11)}$ is written

$$(9)\qquad m^{22}_{(11)} = \mu^{22}_{(11)} + (m^{22}_{(11)})' + (m^{22}_{(11)})'' + \cdots,$$

the term of first degree in the $z^{(t)}_k$ equals

$$(10)\qquad (m^{22}_{(11)})' = \mu^{22}_{(11)}\, m'_{\varkappa'\lambda'}\,(\mu^{\varkappa'\lambda'}_{(11.22)} - \mu^{\varkappa'\lambda'}_{(11)}),$$

where

$$(11)\qquad \mu^{k'l'}_{(11.22)} \equiv \frac{\mathcal{M}_{11.22.k'l'}}{\mathcal{M}_{11.22}}\ (= 0 \text{ if } k' = 2 \text{ or } l' = 2).$$

The mean value of (10) being zero, its variance is found from the formula

$$(12)\quad E\, m'_{kl}\, m'_{np} = \sigma^2\,(\mu_{kn}\,\varepsilon_{lp} + \mu_{kp}\,\varepsilon_{ln} + \mu_{ln}\,\varepsilon_{kp} + \mu_{lp}\,\varepsilon_{kn}),$$

which follows from (9.6). Since, according to (11)

$$(13)\qquad \mu_{\varkappa'n'}\,\mu^{\varkappa'l'}_{(11.22)} = \delta_{n'l'} \text{ if } n' \neq 2,$$

the result is

$$(14)\qquad E(m^{22}_{(11)})'^2 = 4\sigma^2(\mu^{22}_{(11)})^2\,\varepsilon_{\varkappa'\lambda'}(\mu^{\varkappa'\lambda'}_{(11)} - \mu^{\varkappa'\lambda'}_{(11.22)}).$$

By similar computations it can be found that the main term in the mean value

$$(15)\qquad E(m^{22}_{(11)})''$$

of the second degree term in (9) is

$$(16)\qquad \begin{cases} \mu^{22}_{(11)}\, E\, m''_{\varkappa'\lambda'}(\mu^{\varkappa'\lambda'}_{(11.22)} - \mu^{\varkappa'\lambda'}_{(11)}) = \\ \qquad = -\,(T-1)\sigma^2\,\mu^{22}_{(11)}\,\varepsilon_{\varkappa'\lambda'}\,(\mu^{\varkappa'\lambda'}_{(11)} - \mu^{\varkappa'\lambda'}_{(11.22)}), \end{cases}$$

the remaining terms in (15) being of equal order in the μ_k and $\sigma^2\varepsilon_{kl}$ but not containing the factor $T - 1$.

The *negative* contribution (16) to the bias in the denominator of (4) as an estimate of that in (5) constitutes the second cause, indicated [1]) on p. 85, by which the use of the classical formula (3.61) may be misleading if the present specification applies. In cases where this bias becomes important, it might be more accurate to allow for it by dividing $t_{k'}$ in (4) by the factor

$$(17) \qquad q_{k'} \equiv [1 - \frac{T-1}{T-K} \, l \, \varepsilon_{\varkappa\lambda'} \, (m^{\varkappa'\lambda'}_{(11)} - m^{\varkappa'\lambda'}_{(11.k'k')})]^{-\frac{1}{2}}$$

before comparing it with the percentiles in the table of t. Whether such a correction is a relevant improvement can hardly be decided without the actual study of all of the third and fourth degree terms in the development of (4).

It might be thought that, where the distribution of $t_{k'}$ is only approximately that of S t u d e n t's ratio, there is hardly any gain in substituting the use of the table of t for the simple computation from (3) of the "standard error" of $c'_{k'}$. In fact, it depends on the problem whether or not there is reason to consider the ratio's $t_{k'}$, since the error avoided by the use of the table of t depends on $T - K$ only, while the degree of approximation of (4) by (5) is governed by the ratio's in (9.58) or (10.13).

13. *Limits for the weighted regression.*

Finally, we shall even drop the assumption that $\boldsymbol{\epsilon}$ is given a priori. As in section 10, we shall, however, restrict the generality of the parent distribution by assuming that $\boldsymbol{\epsilon}$ is diagonal:

$$(1) \qquad \varepsilon_{kl} = \varepsilon_k \, \delta_{kl}, \qquad \varepsilon_k \geq 0.$$

Given a significance level θ, and hence the corresponding percentiles $\pm p_\theta$ in the t-distribution, we consider the interval on the $\gamma_{k'}$-axis defined by

$$(2) \qquad |t_{k'}| < p_\theta,$$

with $t_{k'}$ as given by (12.4), possibly corrected by means of (12.17). This interval, called by N e y m a n (**18**, App. I) the *region*

---

[1]) In the discussion on p. 85, the terms with $\mu^{k'l'}_{(11.22)}$ were supposed to be small compared with those with $\mu^{k'l'}_{(11)}$, and were therefore neglected.

of acceptance for the parameter $\gamma_{k'}$, consists of those values of $\gamma_{k'}$ which are not rejected from the sample in applying the t-test. It is completely defined by two quantities, its mid point $c_{k'}$ and the denominator

$$s_{e_{k'}} \equiv \left(m_{(11)}^{k'k'} \frac{l}{T-K} \mathbf{c}\epsilon\mathbf{c} \right)^{\frac{1}{2}}. \tag{3}$$

in (12.4).

Now, replacing the unknown quantities (1) by assumed values

$$e_{kl} = e_k \, \delta_{kl}, \quad e_k \geq 0, \tag{4}$$

our first problem is to study the range of variation of $\mathbf{c}(\mathbf{e})$ as defined by (10.3), if the e_k are allowed to assume all possible non negative values. This problem will be considered for the important case in which, possibly after changing the signs of some of the variables X_k,

$$m^{kl} > 0, \quad k, l = 1, 2 \ldots . K, \tag{5}$$

that is, the case in which the signs of the coefficients in the different elementary regressions are compatible.

It follows from (4) and (10.3) that, for values of the e_k for which the two smallest roots of (10.2) differ,

$$l_1(\mathbf{e}) < l_2(\mathbf{e}), \tag{6}$$

and for which $\mathbf{c}(\mathbf{e})$ is, therefore, uniquely defined but for normalization, the vector $\mathbf{g}$ with components

$$g_k \equiv m_{k\lambda} \, c_\lambda(\mathbf{e}) \tag{7}$$

satisfies

$$(e_k \, m^{k\lambda} - l_1^{-1}(\mathbf{e}) \, . \, \delta_{k\lambda}) g_\lambda = 0 \tag{8}$$

where $l_1^{-1}(\mathbf{e})$ is the largest of the roots of

$$| e_k \, m^{kl} - l^{-1}(\mathbf{e}) \, . \, \delta_{kl} | = 0, \tag{9}$$

which are inverse to those of (10.2).

It will first be supposed that the e_k are *positive:*

$$e_k > 0. \tag{10}$$

Let $\mathbf{h}^{(1)}$ be an arbitrary vector satisfying

$$h_k^{(1)} > 0, \tag{11}$$

and consider the infinite series of transformations [1])

$$(12)\qquad \begin{cases} \mathbf{h}^{(1)} \to \mathbf{h}^{(2)} \to \ldots . \mathbf{h}^{(p)} \to \ldots . \text{ defined by} \\ h_k^{(p+1)} = \varphi^{(p)} e_k m^{k\lambda} h_\lambda^{(p)}, \text{ with } \varphi^{(p)} \text{ such that} \\ \mathbf{h}^{(p+1)} \mathbf{e}^{-1} \mathbf{h}^{(p+1)} = \mathbf{h}^{(p)} \mathbf{e}^{-1} \mathbf{h}^{(p)} > 0, \end{cases}$$

where the normalizing factor $\varphi^{(p)}$ serves only to prevent $\mathbf{h}^{(p)}$ from vanishing or becoming infinite for $p \to \infty$. We shall prove that, whether or not (6) holds, the sequence $\mathbf{h}^{(p)}$ converges, for $p \to \infty$, towards a solution $\mathbf{g}$ of (8), unless $\mathbf{h}^{(1)}$ happens to satisfy

$$(13)\qquad \mathbf{h}^{(1)} \mathbf{e}^{-1} \mathbf{g} = 0$$

for every such solution $\mathbf{g}$.

Let $\mathbf{h}^{(p)}$ be written as a linear combination

$$(14)\qquad \mathbf{h}^{(p)} = h^{(p\nu)} \mathbf{g}^{(\nu)}$$

of a set of K vectors of which the n-th one is a solution of

$$(15)\qquad (e_k m^{k\lambda} - l_n^{-1}(\mathbf{e})\, \delta_{k\lambda})\, g_\lambda^{(n)} = 0,$$

the $l_n^{-1}(\mathbf{e})$ being the roots of (9) in order of *decreasing* size. As was mentioned in section 2, it is always possible to choose the $\mathbf{g}^{(n)}$ such that

$$(16)\qquad \mathbf{g}^{(m)} \mathbf{e}^{-1} \mathbf{g}^{(n)} = \delta^{(mn)}.$$

Then, it follows from (12), with (14) inserted, that

$$(17)\qquad h^{(p+1,\, n)} = \varphi^{(p)} h^{(pn)} l_n^{-1}(\mathbf{e}),$$

with $\varphi^{(p)}$ such that

$$(18)\qquad \Sigma_\nu (h^{(p+1,\,\nu)})^2 = \Sigma_\nu (h^{(p\nu)})^2 > 0;$$

whence, as soon as (6) holds, for $p \to \infty$ only $h^{(p1)}$ remains

---

[1]) The use of this series of transformations is a generalization to the "skew" case of a principle used by Hotelling (**17**) for successive numerical computation of the characteristic vectors corresponding to the largest characteristic value, the next largest, and so on, of a moment matrix. This method can also be adapted to compute the orthogonal, the special, and even the general weighted regression. Then, the rapidity of convergence in the sequence, and with that the usefulness of this application of Hotelling's method, varies considerably from one case to the other.

finite, the remaining $h^{(pn)}$ vanishing in approximately geometric progression — unless, according to (13), $h^{(11)} = 0$, in which case $h^{(p1)} = 0$ and, if $h^{(12)} \neq 0$ and $l_2^{-1}(\mathbf{e})$ is single, only $h^{(p2)}$ remains finite for $p \to \infty$, and so on.

Thus, if (6) holds, $\mathbf{h}^{(p)}$ converges to a solution $\mathbf{g}^{(1)}$ of (8). If, however, $l_1^{-1}(\mathbf{e})$ is an n_1-fold multiple root of (9), only the $h^{(pn)}$ with $n \leq n_1$ may remain finite for $p \to \infty$, in which case $\mathbf{h}^{(p)}$ converges to a linear combination of the first n_1 vectors $\mathbf{g}^{(n)}$ which all satisfy (8).

Since, owing to (5), (10), (11), (12),

$$h_k^{(p)} > 0,$$

it follows that

$$g_k \geq 0, \quad \mathbf{g}\mathbf{e}^{-1}\mathbf{g} > 0, \tag{19}$$

or, from (7),

$$\boxed{m_{k\lambda} c_\lambda(\mathbf{e}) \geq 0.} \tag{20}$$

Moreover, since

$$c_k(\mathbf{e}) = m^{k\lambda} g_\lambda, \tag{21}$$

it follows from (5), as at least one of the g_k is *positive*,

$$c_k(\mathbf{e}) > 0. \tag{22}$$

According to (19) the vector $\mathbf{g}$ is limited to one of the 2^K space angles bounded by the coordinate hyperplanes, which will be called the positive coordinate space angle. The K vectors of unit length along the coordinate axes which form the edges of this space angle are transformed by the linear transformation (21) into the vectors in the columns of $||m^{kl}||$ which, but for normalization, indicate the K elementary regressions. Consequently, the conditions (20) mean that, *if* (4), (5) *and* (10) *hold, the special weighted sample regression vector is, whatever the* e_k, *confined to that space angle, constructed on the elementary regression vectors as edges, which contains the orthogonal regression vector; this angle is itself entirely contained within the positive coordinate space angle* — owing to (22). This angle will be called the *elementary regressions space angle*, or, short, *the angle B*.

In the above formulation, the restriction to values of e_k for which (6) holds has been omitted on purpose, because it can now be proved that (6) is a consequence of (5). Suppose that

$$l_1(\mathbf{e}) = l_2(\mathbf{e}) \tag{23}$$

for some set of positive values e_k. Then, there is a plane or hyperplane G of at least two dimensions through the origin O, consisting entirely of vectors $\mathbf{g}$ which are not affected by the transformations in (12), each of them being a solution of (8). As now for nearly every initial vector $\mathbf{h}^{(1)}$ the transformation series (12) must lead to a limiting vector $\mathbf{g}$ within G, at least one non vanishing vector in G must satisfy (19), that is, must lie within or on the border of the positive coordinate space angle. If G contains a vector within that space angle, it must also have two non vanishing vectors in common with the border, since a plane through the summit of a rectangular space angle cannot be entirely contained within it. So, at any rate, G contains at least one non vanishing vector $\mathbf{g}$ on the border, or satisfying

$$g_k \geq 0, \quad g_{k_0} = 0, \quad g_{k_1} > 0,$$

for some pair of values k_0, k_1 of k. Being a vector of G, $\mathbf{g}$ should be invariant for the transformation (12). On the other hand, it follows from (5) and (10) that

$$\varphi\, e_{k_0}\, m^{k_0 \lambda}\, g_\lambda > 0,$$

which shows that (5), (10) and (23) are incompatible.

These results are easily extended to the case where some of the e_k are zero. It follows from (8), in connection with (21) and (22), that the equality sign in (19) can hold only for values of k for which $e_k = 0$. Now if

$$\begin{cases} e_{k_1} > 0, & \text{for} \quad k_1 = 1, 2 \ldots . K_1, \\ e_{k_2} = 0, & \text{for} \quad k_2 = K_1 + 1 \ldots . K, \quad \text{say}, \end{cases} \tag{24}$$

the $K - K_1$ smallest roots of (9) coincide and are equal to zero, while the remaining ones are roots of the equation of degree K_1,

$$| e_{k_1}\, m^{k_1 l_1} - l^{-1}(\mathbf{e})\, \delta_{k_1 l_1} | = 0. \tag{25}$$

Since the above argument can now be repeated for the space of the first K_1 variables only, (19) holds also in the case (24) and can, then, be specialized to

$$g_{k_1} > 0, \quad g_{k_2} = 0, \tag{26}$$

showing that $\mathbf{c}(\mathbf{e})$ lies on the border of the angle B, being a linear combination (with positive coefficients) of the first K_1 elementary regression vectors only.

Owing to (6), only one normalized vector corresponds to a given set of non negative values e_k; on the other hand, it follows from (10.3), with (4) inserted, that

$$l(\mathbf{e})\, e_k = \frac{m_{k\lambda}\, c_\lambda(\mathbf{e})}{c_k(\mathbf{e})}, \tag{27}$$

by which relation, together with their normalization rule, the e_k are defined as single valued functions of the $c_k(\mathbf{e})$.

It appears from (27) and (22) that the e_k are non negative so long as the vector $\mathbf{c}(\mathbf{e})$ satisfies the conditions (20). Thus, any regression vector not outside the angle B represents a special weighted regression vector for some set of weights (proportional to $1/e_k$), and the conditions (20) may be called *the conditions for positive weights.*

The property, shown in section 10 for the mean values of the special weighted regression coefficients, appears to hold in every individual sample for which (5) is satisfied. It is in the form of this proposition that the uncertainty in the weighted regression coefficients, due to the *error of weighting*, occurs in the inductive reasoning from the sample to the parent distribution parameters.

There is, however, a new feature with regard to the error of weighting in the present inductive position which can be best understood by considering some extreme cases are to the spread in the elementary regressions.

Suppose first (*case I*) that, in standard units, all of the orthogonal regression coefficients $a_k^{(1)}$, defined by

$$(m_{k\lambda} - m_1\, \delta_{k\lambda})\, a_\lambda^{(1)} = 0, \tag{28}$$

m_1 being the smallest root of

$$|\mathbf{m} - m_1\, \boldsymbol{\delta}| = 0, \tag{29}$$

are about of equal magnitude, while

$$m_1 \ll \frac{m_2}{K}, \tag{30}$$

if m_2 is the next smallest root of (29). In this case, the spread in the elementary regression vectors is small. For, if the characteristic values of $\mathbf{m}$ are denoted by m_n, in order of increasing size,

$$m_1 < m_2 \leq m_3 \ldots . \leq m_K, \tag{31}$$

and if the corresponding characteristic vectors are denoted by $\mathbf{a}^{(n)}$ and normalized according to

$$a_\varkappa^{(m)} a_\varkappa^{(n)} = \delta^{(mn)}, \tag{32}$$

we have (see section 2)

$$m^{kl} = a_k^{(\nu)} m_\nu^{-1} a_l^{(\nu)} = a_k^{(1)} m_1^{-1} a_l^{(1)} + a_k^{(\nu')} m_{\nu'}^{-1} a_l^{(\nu')}. \tag{33}$$

Since the elements in the l-th column of $\| m^{kl} \|$ are proportional to the components $b_k^{(l)}$ of the l-th elementary regression vector, the latter, if normalized by

$$a_\varkappa^{(1)} b_\varkappa^{(l)} = 1, \tag{34}$$

are, according to (32), equal to

$$b_k^{(l)} = a_k^{(1)} + a_k^{(\nu')} \frac{m_1}{m_{\nu'}} \frac{a_l^{(\nu')}}{a_l^{(1)}}. \tag{35}$$

The $a_k^{(1)}$ being about equal, each of them is, according to (32), about $K^{-\frac{1}{2}}$. Thus, owing to (30), the first term in the right hand member of (35) is large compared with the remaining sum, whatever k, whence the elementary regression vectors cannot differ by much from the orthogonal regression vector $\mathbf{a}^{(1)}$.

Next we consider *case II* where only one, say $a_K^{(1)}$, of the orthogonal regression coefficients, is small compared with the others, which are again of about equal magnitude. If (30) holds, the above argument can be repeated for the first $K - 1$ elementary regression vectors. It breaks down, however, for the last one, because of the small factor $a_K^{(1)}$ in the denominators in the $K - 1$ last terms for $b_k^{(K)}$ in (35). These terms can now make for large discrepancies between $b_k^{(K)}$ and $a_k^{(1)}$. In particular, since

$$a_K^{(\nu')} m_{\nu'}^{-1} a_K^{(\nu')}$$

is a sum of positive terms, there will be a large *positive* difference

$$b_K^{(K)} - a_K^{(1)} \tag{36}$$

in the last components.

The situation is illustrated, for $K = 3$, in figure 13.1 on p. 111, which is drawn within the plane

$$a_\varkappa^{(1)} c_\varkappa = 0. \tag{37}$$

The large triangles represent the intersection of that plane with the coordinate planes (compare fig. 10.1), the small circles indicate the orthogonal regression vector. The edge points of the full-drawn inner triangles represent the elementary regression vectors.

Case II gives rise to a dilemma. One would expect the condition (30) (in standard units!) to be sufficient for an accurate estimation of the γ_k (if these are also measured in standard units), whatever their sizes — provided that $\mathbf{m}$ has been obtained from a sufficiently large number of observations. However, if the conditions for positive weights constituted the only relevant limitation to the error of weighting, the estimation of small regression coefficients would in general be subject to a large error of weighting.

Since the $c_k(\mathbf{e})$ are continuous functions of the e_k, the discrepancies

$$c_K(\mathbf{e}) - a_K^{(1)}$$

can be expected to be especially large if e_K considerably exceeds the remaining e_k, because then the weighted regression will come close to the K-th elementary regression.

On the other hand, the geometric picture presented on p. 62 suggests that ε_K will have the less influence on the sampling distribution of estimated regression coefficients the smaller γ_K compared with the remaining components γ_k. For, if γ_K is small, the last components $X_K^{(t)} - \xi_K^{(t)}$ of the erratic displacements $\mathbf{X}^{(t)} - \boldsymbol{\xi}^{(t)}$ will nearly fall within the "true" regression hyperplane Π, and little influence can be exerted by the variance $\sigma^2 \varepsilon_K$ of these components. This is confirmed by the fact that, in the expressions (9.47) approximating the

sampling variances of the weighted regression coefficients, the ε_k enter only by the factor

$$\gamma\epsilon\gamma$$

on which ε_K has only a small influence if γ_K is small.

At first sight one might, therefore, be astonished to find, by the above argument, a large error of weighting due to ε_K being unknown. The point can be cleared up by considering the expression (10.11) for the bias in the tentatively weighted regression coefficients, which shows that, if γ_K is small, the influence of e_K on the right hand member of (10.11) is small only so long as e_K itself is not large compared with the remaining e_k. Now, just this latter possibility is excluded by *a second limitation on the error of weighting, which is set by the nature of the problem rather than by the mathematics of the last sections.*

In making use of the parent distribution (7.1) (7.2), (7.4) for the interpretation of data, we do not assume only, and this not even in the first place, that the data are governed by a distribution of this mathematical form. We also assume that the variances of the erratic components are small compared with those of the systematic components. This has been recognized in (9.17) as a necessary condition for the adequacy of our approximations. But it is more than that. On p. 6, the erratic components in the determining variables were identified with the "technical" errors in the statistical source from which the data were obtained. No investigator would use a series of values of a variable in regression analysis if he did not expect the variance of this technical error to be a relatively small fraction of the variance of the series itself. Furthermore, the requirement that the set of determining variables is a complete set involves (see p. 6) that the erratic component in the dependent variable, which also contains the combined influence of neglected determining variables, should satisfy the same condition; and, the discussion in section 6 just led us to consider the completeness of the set of determining variables as an essential condition for the reaching of significant results by regression analysis of economic time series. *We are,* therefore, *justified in considering* (9.17) *as an intrinsic part of the specification of the parent distribution* given in section 7.

As we are in these sections engaged in inductive reasoning

from the sample, the condition (9.17) must be considered in terms of estimates instead of the corresponding parameters:

$$\frac{T}{T-K}\,\frac{l(\mathbf{e})\,e_k}{m_{kk}} \ll 1. \tag{38}$$

The left hand member still depends on the assumed values e_k, since, as was stated already in section 8, the variances

$$\sigma_k^2 \equiv \sigma^2\,\varepsilon_k \tag{39}$$

of the individual erratic components can only be estimated if values for their ratio's are assumed.

Now there are different possible cases. Firstly, if (38) is not satisfied for any set of non negative values e_k, the set of determining variables should, on statistical evidence, be considered as incomplete, and should for that reason be rejected, as was pointed out on p. 58. In the opposite case, where (38) is satisfied for every set of non negative values e_k, the angle B sets the only limitation on the error of weighting if nothing is known a priori about the ε_k.

Practically most important is the third, intermediate, case in which (38) is satisfied for some values of the e_k but not for others. This is just what happens in case II above. Now, if there is no reason outside the actual data to suspect the completeness of the set of determining variables, it is quite legitimate to use (38) as a limitation on possible values of e_k. This means that *we reject a posteriori such ratio's of the e_k which lead to estimates of the variance(s) of one or more of the erratic components exceeding reasonable limits.*

Before putting this rule into effect, it is interesting to study the range of variation of the left hand member in (38) for all non negative values of e_k in both cases I and II. For that purpose it is useful to take the $c_k(\mathbf{e})$ instead of the e_k as independent variables.

By (7) and (21), (27) can also be written

$$l(\mathbf{e})\,\mathbf{e}_k = \frac{g_k}{m^{k\lambda}\,g_\lambda}\left(= \frac{1}{m^{kk} + \sum\limits_{\lambda \neq k} m^{k\lambda}\,\dfrac{g_\lambda}{g_k}}\right) \text{ if } g_k > 0\ , \tag{40}$$

where the g_l are limited by (19). Consequently, owing to (5), $l(\mathbf{e})\, e_k$ is largest if

$$g_l = 0 \quad \text{for} \quad l \neq k,$$

that is, according to (40), if

$$e_l = 0 \quad \text{for} \quad l \neq k, \tag{41}$$

and then reaches the value

$$[l(\mathbf{e})\, e_k]_{max} = \frac{1}{m^{kk}}, \tag{42}$$

in accordance with (3.47).

In virtue of (33) this upper limit satisfies

$$[l(\mathbf{e})\, e_k]_{max} \leq \frac{m_1}{(a_k^{(1)})^2}. \tag{43}$$

On the other hand, m_{kk} satisfies

$$m_{kk} = a_k^{(\nu)}\, m_\nu\, a_k^{(\nu)} \geq a_k^{(\nu')}\, m_2\, a_k^{(\nu')} = m_2(1 - a_k^{(1)^2}) \tag{44}$$

on account of (31) and

$$a_k^{(\nu)}\, a_l^{(\nu)} = \delta_{kl}, \tag{45}$$

which is a consequence of (32). In case I, therefore, $l(\mathbf{e})\, e_k$ is, owing to (30), small compared with m_{kk} and (38) is satisfied for every value of k, whatever the e_k, if T exceeds K sufficiently. In case II, however, the left hand member in

$$[l(\mathbf{e})\, e_K]_{max} \geq \frac{1}{\dfrac{(a_K^{(1)})^2}{m_1} + \dfrac{1}{m_2}} = \frac{m_2}{1 + (a_K^{(1)})^2 \dfrac{m_2}{m_1}} \tag{46}$$

need not be small compared with that in

$$m_{KK} \leq m_K, \tag{47}$$

as appears from the right hand members in these inequalities.

In cases where the situation II is approached it may, therefore, be useful to state a set of limits that should not be exceeded by the estimated variances of the erratic components; or, *to impose a system of inequalities,*

$$\boxed{l(\mathbf{e})\, e_k \leq \rho_k\, m_{kk} \quad \text{say,}} \tag{48}$$

on the e_k. For the determining variables, the $\rho_{k'}$ must be chosen in accordance with the expected reliability of the sources from which the variables are obtained. If one or more of the variables are index numbers, some information may also be derived from the degree of agreement between the various series of which the index number constitutes an average. For the choice of ρ_1 there can also be used such information regarding the importance of the influence of neglected determining variables as is furnished by the data.

In order to show the implications of (48) it is useful to choose new units of measurement which make the limits in (48) numerically equal:

$$l(\mathbf{e})\, e_k \leq \rho \text{ say.} \tag{49}$$

Such units will be called *adapted units*. In these units, of course, characteristic values and vectors of $\mathbf{m}$ are different from those computed in standard units. In fact, standard units have been adopted only provisionally, in order to give a well defined meaning to the characteristic values m_n. Henceforth we shall consider all of the formulae in this section as being written in adapted units. Then, in particular, the definitions of cases I and II may have a somewhat altered meaning; the less, however, the closer lie together the factors ρ_k in (48), which do not depend on the units of measurement.

According to (27) and (22), (49) is equivalent to

$$n_{k\lambda}\, c_\lambda(\mathbf{e}) \leq 0, \tag{50}$$

if

$$\mathbf{n} \equiv \mathbf{m} - \rho\boldsymbol{\delta}. \tag{51}$$

From (50) it can be seen what condition the data impose on the choice of ρ_1, if that of the $\rho_{k'}$ has already been made. If ρ_1 is chosen so small that

$$\rho < m_1,$$

$\mathbf{n}$ is positive definite and non singular, all of its characteristic values

$$n_k = m_k - \rho \tag{52}$$

still being positive. Since in that case

$$\mathbf{c}(\mathbf{e})\, \mathbf{n}\, \mathbf{c}(\mathbf{e}) > 0, \tag{53}$$

whatever the non-vanishing vector $\mathbf{c}(\mathbf{e})$, (50) is excluded by (22). Thus, the limits to the variances of erratic components have been set so low that no tentatively fitted regression plane leads to estimates of these variances which are inside the limits. If further

$$\rho = m_1, \tag{54}$$

the only vector for which (53) does not hold is the orthogonal regression vector $\mathbf{a}^{(1)}$, leading to estimated variances which just equal their prescribed limits. Evidently, if the $\rho_{k'}$ have been rightly chosen, ρ_1 must be taken so large that

$$\rho > m_1 \tag{55}$$

in order to allow for the influence of neglected determining variables. In general, it will be safe to choose ρ_1 considerably above the value which leads to (54). Another point of support is supplied by the value $(m^{11})^{-1}$, assumed by $l(\mathbf{e})\, e_1$, according to (42), if the determining variables are given without any errors.

The effect of (50) will be illustrated by again considering the extreme cases I and II. In that, we shall assume m_2 to be so large that, besides (30), also

$$m_1 < \rho \ll m_2. \tag{56}$$

$\mathbf{n}$ then being non singular, (50) is transformed by the linear transformation

$$c_l(\mathbf{e}) \equiv n^{l\varkappa} h_\varkappa, \quad \text{so} \quad h_k \equiv n_{k\lambda}\, c_\lambda(\mathbf{e}), \tag{57}$$

into the inequalities defining the *negative* coordinate space angle

$$h_k \leq 0. \tag{58}$$

Since, owing to (51),

$$n_{k\lambda} a_\lambda^{(1)} = -(\rho - m_1) a_k^{(1)} < 0,$$

$\mathbf{a}^{(1)}$ is one of the vectors satisfying both (20) and (50). Therefore, *as a consequence of the imposition of the limits* (48), *the special weighted sample regression vector is confined to that space angle, constructed on the* K *vectors with components* $-n^{kl}$, $l = 1, 2 \ldots . K$, *as edges, which contains the orthogonal regression vector.*

According to (51),

$$(59) \qquad -n^{kl} = a_k^{(1)}\,(\rho - m_1)\,a_l^{(1)} - a_k^{(\nu')}\,(m_{\nu'} - \rho)\,a_l^{(\nu')}.$$

The vector $\mathbf{j}^{(l)}$ proportional to that in the l-th column of $-n^{kl}$ but normalized analogously to (34) has, therefore, the components

$$(60) \qquad j_k^{(l)} = a_k^{(1)} - a_k^{(\nu')}\,\frac{\rho - m_1}{m_{\nu'} - \rho}\,\frac{a_l^{(\nu')}}{a_l^{(1)}}.$$

Owing to their normalization, the end points of these vectors lie in the plane (37) in which fig. 13.1 has been drawn. Their positions are, according to (56), only slightly affected if in the denominators in the last $K-1$ terms in (60) $m_{n'} - \rho$ is simply replaced by $m_{n'}$. Then, a comparison with (35) shows, that the points (60) are approximately obtained from the points (35) by a multiplication out of the centre $\mathbf{a}^{(1)}$ with a factor

$$(61) \qquad -\frac{\rho - m_1}{m_1}.$$

In figure 13.1, the approximated points (60) are the edge points of the dotted triangles, drawn in accordance with $\rho = 3m_1$.

It is seen that, in case I, the limitations (48) are unimportant. If ρ is sufficiently larger than m_1 they are even, as in fig.

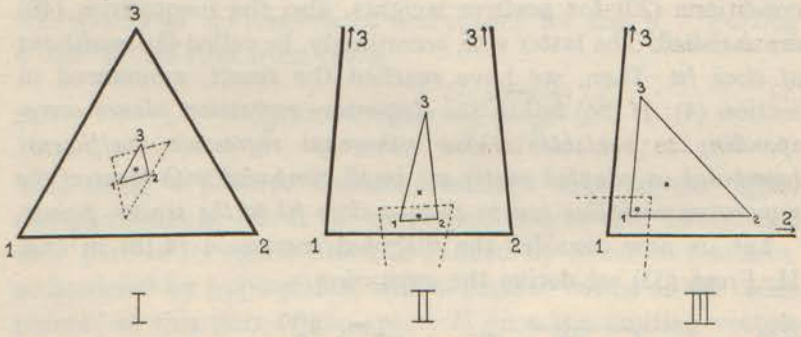

FIGURE 13.1. Limits imposed on the weighted regression by the conditions for positive weights (full drawn triangles) and the conditions of close fit (dotted triangles), in three extreme cases. The small rounds indicate the orthogonal regression, the crosses the diagonal regression.

13.1 *ineffective*. In case II, however, the inequality (48) for $k = 3$ is significantly *effective*, excluding ratio's of the e_k

in which e_K is large compared with the remaining e_k. That this must occur can also be seen in connection with the remarks made on p. 105. Suppose that, in adapted units, all of the e_k are approximately equal. If the ratio's of the e_k happen to be chosen close to the corresponding ratio's of the ε_k, they will, under the conditions of section 9, yield good estimates of $\boldsymbol{\gamma}$ and of the variances σ_k^2 in (39). And, if γ_K is small, the estimation of $\boldsymbol{\gamma}$ and of the first $K-1$ variances σ_k^2 will not be appreciably affected, if only the chosen value of e_K, say, considerably falls short of the corresponding ε_K, according to our remark on p. 106. Just because of this small influence of e_K, however, the estimation of $\boldsymbol{\gamma}$ and of the σ_k^2 is spoiled if the e_k are chosen such that e_K considerably exceeds the remaining e_k. For, in that case, the estimate of $l(\mathbf{e})$ is based on adjusting displacements which are nearly parallel to the X_K axis, and hence, as γ_K is small, also nearly parallel to the true regression plane Π. Thus the errors in the data are forced on to that variable which is least capable of absorbing them, the result being a far too large estimate of σ_K^2, which may even come close to m_{KK}, and a biased estimation of $\boldsymbol{\gamma}$, particularly of γ_K.

A tentatively weighted regression plane will be said to give a *close fit* to the scatter points $\mathbf{X}^{(t)}$ if and only if, besides the conditions (20) for positive weights, also the inequalities (48) are satisfied. The latter will, accordingly, be called *the conditions of close fit*. Then, we have reached the result, announced in section (4): If (56) holds, *the elementary regression planes corresponding to variables whose orthogonal regression coefficients (computed in adapted units) are small compared with those of the remaining variables fail to show a close fit to the scatter points.*

Let us now consider the diagonal regression (4.18) in case II. From (33) we derive the expansion

$$\sqrt{m_1 m^{kk}} = a_k^{(1)} + \tfrac{1}{2} a_k^{(\nu)} \frac{m_1}{m_\nu} \frac{a_k^{(\nu)}}{a_k^{(1)}} - \dots . \tag{62}$$

For $k = K$, the convergence of this series is not certain. If, however, $a_K^{(1)}$ is still large enough to make it convergent, the first term suggests that the *diagonal regression vector* $\mathbf{d}$ need not satisfy (48), but *can show a considerable positive bias* in its last component d_K, or, in general, *in its components corres-*

ponding to small values of γ_k. I found this confirmed in a number of practical examples, with one of which this section is concluded.

Similarly, in case II, the "significance factor" f_{1K} in (4.21) overestimates the real error of weighting, since it allows

$$c_{1K}(\mathbf{e}) \equiv - \frac{c_K(\mathbf{e})}{c_1(\mathbf{e})}$$

to assume any value between $b_{1K}^{(1)}$ and $b_{1K}^{(K)}$, whereas the latter intercoefficient does not permit a close fit to the scatter.

The third part of fig. 13.1 shows the case III in which both γ_K and γ_{K-1} are small. If $K = 3$ and if (56) holds, it cannot occur in standard units, but may occur in adapted units in the common case in which the $\rho_{k'}$ are chosen considerably smaller than ρ_1, according to a situation where the influence of neglected determining variables considerably outweighs in importance the errors in the determining variables. Then, two of the conditions (50) of close fit become effective, together confining the tentatively weighted regression vector to the neighbourhood of the first elementary regression vector.

A few words may be devoted to the remaining problem of the uncertainty in the estimated sampling standard deviations (3) of the weighted regression coefficients, which determine the widths of the regions of acceptance for the γ_k. Putting $\mathbf{e}$ for $\boldsymbol{\epsilon}$, we find from (10.3)

$$s_{c_{k'}}^2(\mathbf{e}) = m_{(11)}^{k'k'} \cdot \frac{\mathbf{c}(\mathbf{e})\,\mathbf{m}\,\mathbf{c}(\mathbf{e})}{T - K} \,. \tag{63}$$

This being a positive definite quadratic form in the $c_k(\mathbf{e})$, the maximum of (63) if the normalized vector $\mathbf{c}(\mathbf{e})$ is confined to a part of its space which is defined by linear inequalities, or bordered by hyperplanes, will be reached in one of the "edge points" of this part of the space. If once the limiting vectors (35) and (60) — in so far as they are relevant — have been found, it will not be difficult (see p. 85) to ascertain the range of variation of (63) for vectors $\mathbf{c}(\mathbf{e})$ satisfying both (20) and (48) by computing the values assumed by $\mathbf{c}(\mathbf{e})\mathbf{m}\,\mathbf{c}(\mathbf{e})$ in some of these "edge points". Thus we may correct for the first one of the causes, mentioned in section 9, by which the errors in the determining variables unfavourable affect the reliability

of the classical formula (3.61) in judging the precision of estimated regression coefficients.

In practical applications, it will not always be possible to have full confidence in the validity of the assumption of uncorrelated errors in the variables. If there is an assignable cause of correlation between the errors in different variables — as may be the case where different variables are quotients with the same index number in the denominators — it may perhaps be possible to replace the set of variables X_k by K independent linear combinations for which the assumption of uncorrelated errors is more likely to hold. If, however, correlation between the errors is only suspected but cannot be quantitatively predicted, there must be left to the weighted regression vector a somewhat larger range of variation than is given by the two limitations (20) and (48). Nevertheless, the inspection of the spread in the elementary regression vectors and in the vectors (60) will in such cases give a point of support in guessing that range of variation. The discussion in this section is to be considered as an example showing how any information or conjecture about the matrix ϵ can be "translated" in terms of limitations to the weighted regression vector. In cases where the problem imposes less rigorous conditions on ϵ than those considered here, it will not be difficult to find, along similar lines, the adequate expression of these requirements in terms of the weighted regression coefficients.

It may be expected that the second restricting assumption — the condition (5) of sign compatibility in the elementary regressions — also does not constitute a serious limitation to useful application of the present results. If (30) holds (in adapted units), the condition (5) can, nevertheless, fail to be satisfied, and that in two different ways.

In the first place, the border of the positive coordinate space angle might be passed by one or more of the elementary regression vectors of which none show a close fit to the scatter. (This would be the case if, in fig. 13.1, II or III, the third elementary regression passed across the coordinate plane 1—3). It appears that, in cases of this type, the conclusions remain the same, since the deviating elementary regressions are far and away excluded by the conditions of close fit.

Secondly, one or more of the elementary regressions of close

fit may have signs deviating from, say, those of the orthogonal regression. (For instance, in fig. 13.1, II, the elementary regression vector $\mathbf{b}^{(2)}$ may pass through the coordinate plane 1—2). In that case, the part of the space filled by regression vectors $\mathbf{c}(\mathbf{e})$ satisfying both the conditions for positive weights and the conditions of close fit extends to the border of the positive coordinate space angle. The sampling error being superposed on the error of weighting, there is in such cases no significant indication that the true regression coefficients γ_k of variables X_k corresponding to the deviating elementary regressions differ from zero. By leaving out the variables concerned, there can, then, be produced a situation where the condition (5) holds, or where only deviations of the first type occur.

14. *An application to the ship freight market.*

We shall illustrate the preceding results by means of the figures contained in a study made by T i n b e r g e n (**30**) of the world ship freight market in the period 1880–1911.

The dependent variable X_1, the freight rate, is statistically expressed by the *"Fairplay" index of homeward freights* (**41**), for the years previous to 1885 completed by means of figures given by H o b s o n (**43**).

According to market theory, the relation between X_1 and its determining variables is obtained by equating two expressions for the *actual transport* X_2, one, the *demand equation*

$$X_2 = fX_1 + D, \tag{1}$$

expressing X_2 in terms of the freight rate X_1 and demand factors, here summarized by the symbol D, and another, the *supply equation*

$$X_2 = gX_1 + S, \tag{2}$$

expressing X_2 in terms of X_1 and the summarized supply factors S. The result is

$$X_1 = \frac{D - S}{g - f}. \tag{3}$$

A further analysis suggests that D depends, in the simplest case by a linear relation, on the differences in the prices — of the

goods to be transported — between the countries of origin and those of destination. Owing to the large number of goods and of countries, however, it is difficult to find an adequate statistical expression for D. For that reason, T i n b e r g e n argued as follows. The freight rate being of much more importance to the persons on the supply side of the market than to those on the demand side, f can be expected to be small compared with g. Putting $f = 0$, D becomes, according to (1), equal to X_2, and (3) takes the form

$$X_1 = \frac{1}{g}(X_2 - S). \tag{4}$$

Mathematically, this equation is identical with the supply equation (2). It is, however, considered as an approximation to equation (3), which has a logically different meaning in that it expresses X_1 in terms of its determining variables. Thus X_2 is in (4) to be taken as an index of demand factors, replacing the logically more correct but statistically inaccessible index D. Then, in particular, the ratio of the coefficients of X_2 and S in (4) is the same as that of the coefficiënts of D and S in (3).

Statistically, X_2 is obtained by adding the transport, measured in milliards of tons × miles, of the most important goods, namely grains, coal, oil, Chilean nitrates, and wood, between the principal countries involved [1]).

For S two main supply factors are introduced, the *total tonnage* X_3 and the *coal price* X_4. A series representative of X_3 is found by adding the tonnage of the trade fleets of England, the United States, Germany and Norway, counting a ton of a steamer as 3 times a ton of a sailing ship. For X_4 is taken the export price of coal from England.

T i n b e r g e n's computations have been performed by means of three years moving averages of percentual augmentations over three years, a method which eliminates a great deal of the influence of rapidly fluctuating determining variables not included in the above set, which influence appears to be considerable in the data. This procedure does not admit an application of the present theory, because it introduces considerable serial correlation in the errors in the variables. Therefore, *regression coefficients*

---

[1]) For particulars, see T i n b e r g e n (**30**, 1934).

have been recomputed for percentual deviations from trend. For X_4 a linear trend, fitted by the "least squares" principle, appeared to be sufficient; for the remaining series, which show cyclical oscillation superposed on a somewhat irregular trend, great care has been given to the fitting of smooth trends by moving averages of a variable number of years. Such a procedure cannot fail to introduce some serial correlation in the errors in the trend deviations, which might affect the applicability of the formulae derived in this study on the assumption of independent errors in the successive years. The following reasoning suggests, however, that it may be possible to allow for this circumstance by a simple correction.

It has been proved by Frisch and Waugh (**42**) that if, in a set of variables, an elementary regression equation is fitted to the *absolute* deviations from linear trends fitted by the "least squares" procedure, the regression coefficients are the same as those of the corresponding elementary regression in another set of variables combining the original variables with the time t. If in that larger set of variables the conditions of Fisher's specification apply, the sampling variances of these regression coefficients may be unbiasedly estimated by means of formula (3.61) *computed from moments of trend deviations,* if only $T - K$ is replaced by $T - K - 1$.

These results are easily extended to trends of the same degree P in t for each of the variables and to the weighted regression if the additional variables, the first P powers of t, are taken as variables without errors. If in the larger set of variables the conditions of the present specification are satisfied, with

$$\varepsilon_{K+p} = 0, \qquad p = 1, 2 \ldots . P, \tag{5}$$

the errors of sampling can be studied by means of the formulae derived in this investigation, computed from trend deviations, if only P is substracted from the number $T - K$ of degrees of freedom entering into the estimation of σ^2.

In the present problem, the situation is different in two respects. Firstly, a linear relation is fitted to *relative* instead of absolute trend deviations; then, a trend of moving averages has been adopted for three out of the four variables. Nevertheless, it is considered as a good approximation to the correction for serial correlation due to trend fitting if the present formulae are

applied with 5 substracted from the number of observations, the trends of moving averages having the general aspect of about fifth degree polynomial trends.

In table 14.1 the original variables are listed. Table 14.2 indicates the *standard units* in which the computations have been effected. But for their signs, which are chosen so as to make positive all coefficients of the first elementary regression, they

Table 14.1

Year	Freight index 100 in 1900	Transport milliards of tons × miles	Tonnage millions of tons	Coal price shillings per ton
1880	154	101	16.59	8.9
1881	138	100	17.25	9.0
1882	135	109	18.16	9.1
1883	133	114	19.39	9.3
1884	120	118	20.39	9.3
1885	106	120	20.80	8.9
1886	97	114	20.59	8.4
1887	95	120	20.55	8.3
1888	106	136	20.91	8.4
1889	125	143	22.15	10.2
1890	103	142	23.47	12.6
1891	104	159	24.78	12.2
1892	85	140	25.79	11.0
1893	86	154	26.49	9.9
1894	82	173	27.12	10.5
1895	80	173	27.59	9.3
1896	76	178	27.97	8.8
1897	78	191	28.27	8.8
1898	94	195	28.90	9.8
1899	85	211	30.11	10.5
1900	100	223	31.29	16.5
1901	74	210	32.92	13.7
1902	66	219	34.63	12.2
1903	71	231	36.20	11.6
1904	72	253	37.44	11.0
1905	69	264	38.77	10.5
1906	68	277	40.54	10.8
1907	70	286	42.51	12.6
1908	58	272	43.77	12.6
1909	64	299	44.17	11.2
1910	65	316	44.46	11.6
1911	75	314	45.47	11.3

Table 14.2 $T = 32$

k	1	2	3	4
Standard units, in percents of trend	—53.6	9.87	—12.27	76.1

Table 14.3 $M = 0.341$

m_{kl}	$l = 1$	2	3	4
$k = 1$	1.000	—0.666	—0.167	—0.451
2		1.000	0.018	0.150
3			1.000	—0.223
4				1.000

Table 14.4 $r_{1.234} = 0.795$

M_{kl}	$l=1$	2	3	4
$k=1$	0.927	0.559	0.231	0.392
2		0.685	0.119	0.174
3			0.422	0.180
4				0.532

Table 14.5

k	2	3	4
$M_{11.kk}$	0.950	0.978	1.000

Table 14. 7 $j_1 = 1.00$

k	2	3	4
$j_k^{(1)}$	0.66	0.29	0.46
$j_k^{(2)}$	0.50	0.30	0.49
$j_k^{(3)}$	0.68	0.10	0.36
$j_k^{(4)}$	0.70	0.22	0.21

Table 14.6 $b_1^{(1)} = 1.00$ etc.

k	2	3	4	row
$b_k^{(1)}$	0.60 0.13	0.25 0.13	0.42 0.13	(1)
$b_k^{(2)}$	1.23	0.21	0.31	(2)
$b_k^{(3)}$	0.52	1.83	0.78	(3)
$b_k^{(4)}$	0.44	0.46	1.36	(4)
d_k	0.86	0.68	0.76	(5)
$*b_k^{(1)}$	0.76	0.31	0.53	(6)
$\Delta b_k^{(1)}$	0.31	0.37	0.57	(7)
$^{I}b_k^{(1)}$	0.52	0.20	0.45	(8)
$^{II}b_k^{(1)}$	0.72	0.43	0.43	(9)

equal the square roots of the square moments of the percentual deviations from trend of the variables given in table 14.1. The moment matrix $\mathbf{m}$, in these units equal [1]) to the correlation matrix $\mathbf{r}$, is given in table 14.3, together with its determinantal value M. Its adjoint $||M_{kl}||$ is shown in table 14.4, together with the total correlation coefficient

$$r_{1.234} \equiv \left(1 - \frac{M}{m_{11}M_{11}}\right)^{\frac{1}{2}}. \tag{6}$$

It appears that the signs in the rows of $||M_{kl}||$ are compatible. Furthermore, M_{11} is only slightly below its maximum $m_{22}m_{33}m_{44} = 1$, whence there is no possibility of important linear dependence between the determining variables diminishing the rapidity of convergence of the expansions given in sections 9–12. Table 14.5 supplies the principal minors in $|\mathbf{m}|$ of order 2 not containing the element m_{11}.

Table 14.6 contains the coefficients of a number of different regressions which are discussed in what follows. All of them are normalized by equating to unity the first coefficient. The first four are the elementary regressions, having coefficients proportional to the elements in the rows of table 14.4. There exists a considerable spread in the four values for each of the three coefficients $b_{k'}$, and that the more, the smaller the coefficient $b_k^{(1)}$. The *conditions of positive weights* leave, in fact, a range of variation to the weighted regression coefficients which is large compared with the classical "standard errors" in the first elementary regression coefficients, attached in row (1) of table 14.6 to the corresponding coefficients and computed from (3.61) with 5 substracted from the number of observations. It is, therefore, of primary importance to see whether the error of weighting is further limited by the *conditions of close fit.*

If only the k-th variable is supposed to be subject to error the estimated variance of that error bears to the actual variance of the variable in which it occurs the ratio

$$\frac{T}{T-K-P}(m_{kk}m^{kk})^{-1}, \text{ with } P = 5, \tag{7}$$

[1]) It seems superfluous to break the continuity in the formulae by writing $\mathbf{r}$ for $\mathbf{m}$.

in consequence of (3.47). Owing to (13.42) this is at once the maximum of that ratio for all positive weights e_k^{-1}. The values, assumed by $(m_{kk}m^{kk})^{-1}$ for $k = 1, 2, 3, 4$ are

(8) 0.37, 0.50, 0.83, 0.64.

There is no doubt that the ratio's (7) to which they lead for $k = 2$, 3, or 4 must be excluded by the conditions of close fit. Accepting the second elementary regression as a possibility would mean to admit in the series for the actual transport an error whose standard deviation exceeds 0.8 of that of the series itself; and still more preposterous assumptions would be implied in the acceptance of the third and fourth elementary regression.

To make a good guess at the maximum possible error variances in the series for the determining variables would require more knowledge of the statistics concerned than the present author disposes of. To be cautious, we shall take

(9) $$\rho_2 = \rho_3 = \rho_4 = 0.10$$

in the three last of the conditions (13.48) of close fit, that is, we shall assume the estimated standard deviations of the errors in the determining variables not to exceed $\sqrt{(32/23) \times 0.10} = 0.37$ of the actual standard deviations of these variables in the period concerned. For X_3 and X_4 these limits are doubtless taken generously large, because X_3 is rather accurately known and X_4, if measured in percentual trend deviations, has a relatively large variance (see table 14.2), of which more than one third is due to the single year 1900. Yet, the limits (9) will appear sufficient for an enormous reduction of the range of variation permitted to the special weighted regression.

The value 0.37 for $(m_{11}m^{11})^{-1}$, though the smallest of the four values in (8), suggests that the freight index X_1 is subject to considerable influence from other determining variables than those already considered. In fact, the first elementary regression "explains" only $1-(32/23) \times 0.37 = 0.49$ of its variance by the influence of X_2, X_3 and X_4, and the set of determining variables cannot be considered as complete in the sense that it leaves only small "unexplained" residuals in the dependent variable. Therefore, the estimated regression coefficients and the valuation of their reliability to be presented here can only be trusted in so far as the hypothesis is justified that the remaining determining

variables are so numerous and so erratic in nature that their combined influence, though being large, still has the character assumed in the present specification for the erratic components, viz., independence, from the variables, from the erratic components in other variables, and from each other in successive observations; and an approximately normal distribution (though this latter condition is probably less imperative).

Without venturing a definite judgement about this hypothesis, we shall for purposes of illustration assume that it holds. Then, it will be safe not to impose an effective limitation on the estimated variance of the "error" in X_1. To visualize the effect of the remaining conditions in (13.48) we shall, however, complete them by a condition for $k = 1$ with

$$\rho_1 = 0.40, \tag{10}$$

which is ineffective since ρ_1 exceeds the maximum value 0.37 of $l(\mathbf{e})e_1/m_{11}$. The standard units are, then, at once *adapted units* only for the last three variables, and the matrix (13.51) now has the elements

$$n_{kl} = m_{kl} - \rho_k \delta_{kl}. \tag{11}$$

The vectors $\mathbf{j}^{(l)}$ with components

$$j_k^{(l)} = \frac{n^{kl}}{n^{k1}}, \quad \text{so} \quad j_1^{(l)} = 1, \tag{12}$$

are given in table 14.7. They constitute the edges of the space angle to which, according to the conditions of close fit, the weighted regression must be confined. This angle and the elementary regressions space angle each intersect the three-dimensional hyperplane

$$c_1 = 1 \tag{13}$$

in a tetrahedron, which have both been drawn in fig. 14.1 in three orthogonal projections on the three coordinate planes.

It appears most strikingly from this figure that the conditions of close fit exclude nearly the whole of the region permitted by the conditions of positive weights, admitting only regression vectors which lie in a small part of that region close to the first elementary regression. In fact, the error of weighting for each of the three coefficients is reduced by the conditions of close fit to a small

fraction of the standard deviation of the classical error of sampling, shown in row (1) of table 14.6.

The diagonal regression **d**, computed from the diagonal elements in table 14.4 and shown in the fifth row of table 14.6, is indicated in fig. 14.1 by the point D. It turns out to contain a positive bias considerably exceeding the classical standard

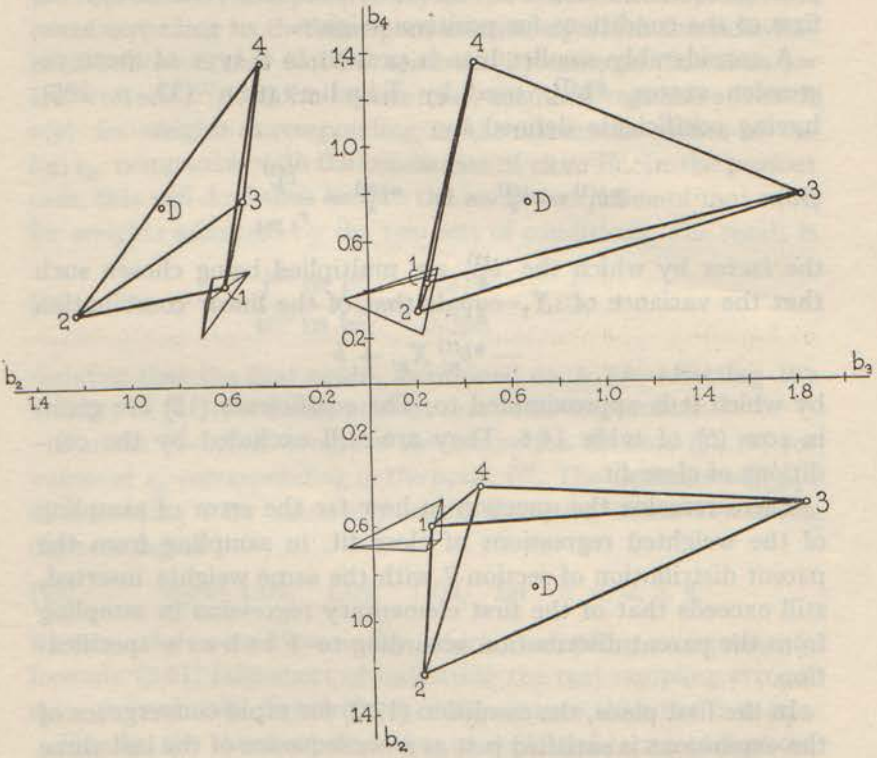

Figure 14.1. Limits imposed on the weighted regression coefficients c_k ($c_1 = 1$) of the freight index X_1 on the actual transport X_2, the total tonnage X_3, and the coal price X_4 (percentual trend deviations, *measured in standard units*, all signs being made positive). The point (c_2, c_3, c_4) satisfies the *conditions of positive weights* by lying within the large tetrahedron (drawn in three orthogonal projections) and is in addition by the *conditions of close fit* confined to the small tetrahedron. D indicates the diagonal regression.

error, especially in the coefficient d_3 corresponding to the determining variable X_3 of smallest influence. It would lead to estimates of the variances of the erratic components which

bear to the actual variances of the variables in which they occur the ratio's

(14) —0.04, 0.52, 0.73, 0.52,

for $k = 1, 2, 3, 4$ respectively, all of which are unacceptable. The negative sign of the first of these ratio's shows that, in the present case, the diagonal regression does not even satisfy the first of the conditions for positive weights.

A considerably smaller bias is present in a type of mean regression vector $*\mathbf{b}^{(1)}$, used by Tinbergen (**32**, p. 385), having coefficients defined by

$$(15) \qquad *b_1^{(1)} = b_1^{(1)} = 1, \quad *b_{k'}^{(1)} = \frac{b_{k'}^{(1)}}{r_{1.234}},$$

the factor by which the $b_{k'}^{(1)}$ are multiplied being chosen such that the variance of X_1 equals that of the linear combination

$$- *b_{\varkappa'}^{(1)} X_{\varkappa'} + b$$

by which it is approximated to. The coefficients (15) are given in row (6) of table 14.6. They are still excluded by the conditions of close fit.

There remains the question in how far the error of sampling of the weighted regressions of close fit, in sampling from the parent distribution of section 7 with the same weights inserted, still exceeds that of the first elementary regression in sampling from the parent distribution according to Fisher's specification.

In the first place, the condition (12.2) for rapid convergence of the expansions is satisfied just as a consequence of the last three of the conditions (13.48) of close fit, which, owing to (9), can now be written

$$(16) \qquad \frac{T}{T-K-P} \frac{le_{k'}}{m_{k'k'}} \leq 0.10,$$

while the factors $(m_{k'k'}\, m_{(11)}^{k'k'})^2$ are only

(17) 1.05, 1.12, 1.16,

for $k' = 2, 3, 4$ respectively, owing to the complete absence of linear dependence between determining variables. This is, of course, a particular case of a general feature of the present meth-

od: *In so far as the rapidity of convergence in the developments is not unfavourably affected by intercorrelation of determining variables, it is ensured by the conditions of close fit in all cases where the errors in the statistical series are known to be not particularly large.*

Thus, we may study the question at issue by a comparison of the expressions (13.63) and (3.61) for the estimated sampling variances according to the two specifications, of which the ratio has been indicated in (9.63). We shall simply compute this ratio for the vector $\mathbf{j}^{(1)}$, which equals the weighted regression vector $\mathbf{c}(\mathbf{e})$ for weights corresponding to the maximum values of the $l(\mathbf{e})\, e_{k'}$ compatible with the conditions of close fit. In the present case, this will doubtless lead to the maximum value of that ratio for weights admitted by the two sets of conditions. The result is

$$\frac{\mathbf{j}^{(1)}\,\mathbf{m}\,\mathbf{j}^{(1)}}{\mathbf{b}^{(1)}\,\mathbf{m}\,\mathbf{b}^{(1)}} = \frac{0.379}{0.368} = 1.03, \tag{18}$$

showing that the first cause, mentioned on p. 84, affecting the validity of (3.61) is in the present case of no importance.

Further, we shall compute the correction divisors (12.17) for values of ε_k corresponding to the point $\mathbf{j}^{(1)}$. The elements $m^{\nu\nu'}_{(11.k'k')}$ all exceeding 1 by relatively small amounts, these correction divisors equal

$$q_{k'} = 1.09, \quad 1.09, \quad 1.08, \quad \text{for} \quad k' = 2, 3, 4, \tag{19}$$

whence, the second cause by which, according to section 9, formula (3.61) falls short of indicating the real sampling error if $\varepsilon_{k'} > 0$, is of only slightly greater importance than the first one.

Finally, the standard deviation of neglected terms of first degree in the development (12.9) of $m^{k'k'}_{(11)}$, of which the square is given in (12.14), is in the point $\mathbf{j}^{(1)}$ estimated by a quantity which bears to $m^{k'k'}_{(11)}$ a ratio, up to the second decimal place equal to

$$0.14, \tag{20}$$

for $k' = 2, 3, 4$ respectively. Thus, even for errors so small as those admitted here in nearly not intercorrelated determining variables, the width of the region of acceptance (13.2), though corrected by means of (12.17) for the bias due to second degree terms in the expansion (12.9) of $m^{k'k'}_{(11)}$, falls in repeated samples irregularly above and below the right value with considerable

instability in consequence of the neglect of first degree terms in that expansion.

Apart from this uncertainty, the approximations in the preceding sections lead to a correction of the regions of acceptance (3.54), derived by means of Fisher's specification, owing to which the width of this region is multiplied by a factor, at most equal to 1.12, the product of the factors in (18) and (19), in the point $\mathbf{j}^{(1)}$ of largest admitted departure from the first elementary regression. In addition, allowance must be made for the error of weighting by admitting the mid-points of the regions of acceptance to assume any values corresponding to points which are in fig. 14.1 inside both of the tetrahedra.

Table 14.8 $T - K - P = 23$

Lower and upper limits of the regions of acceptance for	c_2	c_3	c_4
I. according to Fisher's specification, $\theta = 0.05$	0.33	—0.02	0.15
	0.87	0.52	0.69
II. according to its generalization given in section 7, $\theta = 0.05$	0.33	—0.02	0.15
	0.96	0.59	0.76
III. Frisch's limits	0.60	0.25	0.42
	1.23	1.83	1.36

Table 14.8 shows the regions of acceptance (3.54) compared with those obtained by the above corrections, both for the significance level $\theta = 0.05$, and with those corresponding to Frisch's limits (4.17).

The coefficient c_3 for the influence of total tonnage is just in a position of doubtful significance on purely statistical evidence. Since the economic significance is beyond suspicion, our conclusion must be that the present data admit only the indication of a wide range within which c_3 is likely to be included.

It appears that the present example is particularly suited to illustrate the criticism of some elements in Frisch's approach to the regression problem, to which this study gave rise. On the other hand, it can hardly serve as an illustration of the primary importance of Frisch's criticism of the arbitrary use

of Fisher's specification, a criticism which is arduously supported in these pages.

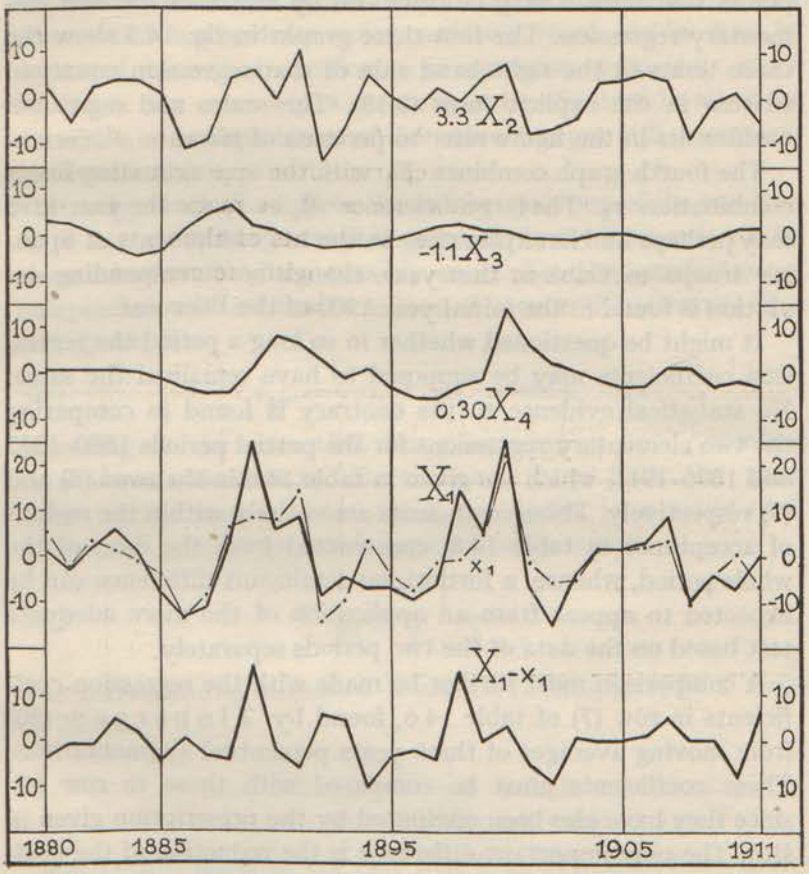

Figure 14.2. The freight index X_1 approximated to by the first elementary regression $x_1 = 3.3\,X_2 - 1.1\,X_3 + 0.30\,X_4 + b$ on the actual transport X_2, the total tonnage X_3 and the coal price X_4. *The units are percents of trend.*

This situation is due to two features of our present example: the complete absence of approximate linear dependence between determining variables, and the large influence exerted on X_1 by determining variables other than X_2, X_3, X_4. There is, however, no doubt that in other cases the possible dangerous effects of an abuse of Fisher's specification would become manifest by application of the results of this investigation.

The results derived by Fisher's specification being rehabilitated for the present example as a good approximation, some points that remain may be discussed by means of the first elementary regression. The first three graphs in fig. 14.2 show the three terms in the right hand side of that regression equation, written in the explicit form (3.18). The scales and regression coefficients in the figure refer to percents of trend.

The fourth graph combines X_1 with the approximating linear combination x_1. The large difference $X_1 - x_1$ for the year 1898 may perhaps find its explanation in the large shipments of Spanish troops to Cuba in that year, though no corresponding deviation is found in the initial year 1900 of the Boer war.

It might be questioned whether in so long a period the regression coefficients may be supposed to have remained the same. No statistical evidence to the contrary is found in comparing the two elementary regressions for the partial periods 1880–1895 and 1896–1911, which are given in table 14.6 in the rows (8) and (9) respectively. These coefficients are entirely within the regions of acceptance in table 14.8, constructed from the data of the whole period, whence, a fortiori, no significant difference can be expected to appear from an application of the more adequate test based on the data of the two periods separately.

A comparison must further be made with the regression coefficients in row (7) of table 14.6, found by Tinbergen (**30**) from moving averages of three years percentual augmentations. These coefficients must be compared with those in row (6), since they have also been computed by the prescription given in (15). The only important difference is the reduction of the coefficient estimating the influence of X_2, a variable which is much more subject to rapid and irregular fluctuations than X_3 and X_4 Without further inquiry, it is difficult to decide whether this discrepancy is due to a bias, introduced in the estimation of regression coefficients corresponding to variables of irregular short term variation by the averaging procedure in cases where the present (or Fisher's) specification applies — or whether it is an indication that the hypothesis of a simple linear relation between simultaneous values of the variables constitutes an oversimplification, where perhaps in reality either less or more influence is exerted by a value of a determining variable that has persisted during some years than by a value which itself differs considerably

from preceding values. To answer questions of this kind, there is need for a test capable of discerning the more complicated hypotheses from the simple "linear" hypothesis. If hardly any errors occur in the determining variables, much can be done by means of the test of significance of regression coefficients, given by (3.54). For cases where errors must be admitted in all of the variables, several problems of the indicated type still await their solution.

Finally, we shall compute the serial correlation of the residuals $X_1 - x_1$, exhibited in the last graph in fig. 14.2, in order to inquire whether the data possibly contradict the assumption of independence in successive values of erratic components. This correlation is

$$r_{ser} = -0.17 \quad \text{or} \quad -0.14, \tag{21}$$

according to the year 1898 being included or not. According to a formula derived by Bartlett (**2**, p. 537), the mean value of the square of the serial correlation in a series of T spherically normal variables approximately equals

$$E\, r_{ser}^2 = \frac{1}{T-1}. \tag{22}$$

The corresponding standard deviation is, therefore, in a series of 32,

$$(E\, r_{ser}^2)^{\frac{1}{2}} = \frac{1}{\sqrt{31}} = 0.18.$$

Thus, even if ignoring the fact that a small negative serial correlation must have been introduced in the residuals by the trend fitting procedure, no significant deviation from independence is found by this test.

REFERENCES

1 M. S. Bartlett, On the theory of statistical regression, Proc. Roy. Soc. Edinb., **53**, 1933, p. 260.

2 M. S. Bartlett, Some aspects of the time correlation problem, Journ. Roy. Stat. Soc., **98**, 1935, p. 536.

3 G. Darmois, Analyse et comparaison des séries statistiques qui se développent dans le temps, Metron, **8**, 1—2, 1929, p. 211.

4 J. B. D. Derksen, Inleiding tot de correlatierekening, Delft 1935.

5 R. A. Fisher, On the mathematical foundations of theoretical statistics, Phil. Trans. Roy. Soc. London, **A 222**, 1922, p. 309.

6 R. A. Fisher, The goodness of fit of regression formulae and the distribution of regression coefficients, Journ. Roy. Stat. Soc., **85**, 1922, p. 597.

7 R. A. Fisher, Theory of statistical estimation, Proc. Cambr. Phil. Soc., **22**, 1925, p. 700.

8 R. A. Fisher, Applications of "Student's" distribution, Metron, **5**, 3, 1926, p. 3.

9 R. A. Fisher, Inverse probability...., Proc. Cambr. Phil. Soc., **26**, 1930, p. 528; **28**, 1932, p. 257.
R. A. Fisher, The concepts of inverse probability...., Proc. Roy. Soc. London, **A 139**, 1933, p. 343.
R. A. Fisher, Probability, likelihood...., Proc. Roy. Soc. London, **A 146**, 1934. p. 1.

10 R. A. Fisher, Statistical methods for research workers 5th edition, Edinburgh 1934.

11 R. A. Fisher, The logic of inductive inference, Journ. Roy. Stat. Soc., **98**, 1935, p. 39.

12 R. A. Fisher, The mathematical distributions used in the common tests of significance, Econometrica, **3**, 1935, p. 353.

13 R. Frisch, Correlation and scatter in statistical variables, Nord. Stat. Tidskr., **8**, 1928, p. 36.

14 R. Frisch and B. D. Mudgett, Statistical correlation and the theory of cluster types, Journ. Am. Stat. Ass., **26**, 1931, p. 375.

15 R. Frisch, Propagation problems and impulse problems in dynamic economics, Economic essays in honour of Gustav Cassel, London 1933; reprinted as Publ. 3, Univ. Øk. Inst., Oslo, 1933.

16 R. Frisch, Statistical confluence analysis by means of

complete regression systems, Publ. 5, Univ. Øk. Inst., Oslo 1934.

17 H. Hotelling, Analysis of a complex of statistical variables into principal components, Journ. Ed. Psych., Sept. & Oct. 1933; reprinted: Baltimore 1933.

18 J. Neyman, On two different aspects of the representative method, Journ. Roy. Stat. Soc., **97**, 1934, p. 558.

19 J. Neyman and E. S. Pearson, On the use and interpretation of certain test criteria for purposes of statistical inference, Biometrika, **20 A**, 1928/29, p. 175, p. 263.
E. S. Pearson and N. K. Adyanthāya, The distribution of frequency constants in small samples from non-normal symmetrical and skew populations, Biometrika, **21**, 1929, p. 164.

20 E. S. Pearson, The test of significance for the correlation coefficient, Journ. Am. Stat. Ass., **26**, 1931, p. 128; **27**, 1932, p. 424.

21 K. Pearson, On the criterion that...., Phil. Mag., **V**, **50**, 1900, p. 157.
K. Pearson, On a brief proof... , Phil. Mag., **VI**, **31**, 1916, p. 369

22 K. Pearson, On lines and planes...., Phil. Mag., **VI**, **2**, 1901, p. 559.

23 E. C. Rhodes, On lines and planes of closest fit, Phil. Mag., **VII**, **3**, 1927, p. 357.

24 H. I. Richards, Analysis of the spurious effect of high intercorrelation of independent variables on regression and correlation coefficients, Journ. Am. Stat. Ass., **26**, 1931, p. 21.

25 P. R. Rider, Developments in the analysis of multivariate data, Econometrica, **4**, 1936, p. 264.

26 H. Schultz, Interrelations of demand, Journ. Pol. Ec., **41**, 1933, p. 468.

27 H. Schultz, Interrelations of demand, price, and income, Journ. Pol. Ec., **43**, 1935, p. 433.

28 Student, The probable error of a mean, Biometrika, **6**, 1908, p. 1.

29 D. H. Thomas, Social aspects of the business cycle, London, 1925.

30 J. Tinbergen, Scheepsbouw en Conjunctuurverloop, De Ned. Conjunctuur, Maart 1931, p. 14.
J. Tinbergen, Scheepsruimte en vrachten, De Ned. Conjunctuur, Maart 1934, p. 23.

31 J. Tinbergen, Der Einflusz der Kaufkraftregulierung auf den Konjunkturverlauf, Zs. für Nat. Oek., **5**, 1934, p. 289.

32 J. Tinbergen, Quantitative Fragen der Konjunkturpolitik, Weltwirtsch. Archiv, **42**, 1935, p. 366.

33 J. Tinbergen, Winsten, koersen en investeeringen, De Ned. Conjunctuur, Aug. 1935, p. 14.

34 J. Tinbergen, Grondproblemen der theoretische statistiek, Haarlem 1936.

35 J. Tinbergen, Fondements mathématiques d'une politique économique, à paraître, Paris 1936.

36 M. J. van Uven, Adjustment of N points...., Proc. Kon. Akad. van Wetensch. Amst., **33**, 1930, p. 143; **33**, 1930, p. 307.

37 E. J. Working, Demand studies during times of rapid economic change, Econometrica, **2**, 1934, p. 140.

38 H. Working and H. Hotelling, The application of the theory of error to the interpretation of trends, Proc. Am. Stat. Ass., **24**, 1929, p. 73.

39 G. U. Yule, An introduction to the theory of statistics, 10th ed. London 1932.

40 G. U. Yule, On the time-correlation problem, Journ. Roy. Stat. Soc., **84**, 1921, p. 497.

41 "Fairplay", 1916, supplement.

42 R. Frisch and F. V. Waugh, Partial time regressions as compared with individual trends, Econometrica, **1**, 1933, p. 387.

43 C. K. Hobson, The export of capital, London 1914.

STELLINGEN

I

Indien van twee reële hyperellipsoïden in de n-dimensionale Euclidische ruimte (twee reële $(n-1)$-dimensionale quadratische variëteiten die de oneigenlijke ruimte volgens een imaginaire figuur snijden) met gemeenschappelijk middelpunt de eerste (A) binnen (resp. niet buiten) de tweede (B) ligt, dan geldt, als $a_1, a_2 \ldots . a_n$ en $b_1, b_2 \ldots . b_n$ de lengten der assen van A en B voorstellen in volgorde van niet afnemende grootte,

$$a_k < b_k \quad (\text{resp. } a_k \leq b_k), \quad k = 1, 2 \ldots . n.$$

II

Zijn **A** en **B** twee reële symmetrische matrices, en zijn de wortels der karakteristieke vergelijking van **B** alle positief (resp. niet negatief), dan zijn de wortels der karakteristieke vergelijking $\mathbf{A} + \lambda\mathbf{B}$ monotoon stijgende (resp. niet dalende) functies van de reële parameter λ.

III

Er is geen grond om bij de studie van de atoombouw aan een nauwkeurige numerieke oplossing van de integro-differentiaalvergelijkingen van Fock groot gewicht toe te kennen.

IV

De wijze waarop de waarschijnlijkheidsrekening in de quantummechanica wordt gebruikt bij de interpretatie van waarnemingen vertoont overeenkomst met het statistisch vraagstuk om conclusies te trekken aangaande een niet of onvolledig bekende waarschijnlijkheidsverdeling als uit deze verdeling in principe slechts één steekproef kan worden verkregen.

V

De interpretatie, door Deming en Birge gegeven van de schattingsmethode volgens „maximum likelihood", is onjuist.

W. E. Deming and R. T. Birge, On the statistical theory of errors, Rev. of Mod. Physics, Vol. 6, 1934, p. 145.

VI

Het bezwaar, door Drion aangevoerd tegen de toepassing van „Student's" *t*-test bij de studie van biologische problemen, is niet ernstig.

E. F. Drion, On the interpretation of frequency curves in biology, Rec. des Trav. Botan. Néerl., Vol. 33, 1936, p. 101.

VII

Tegen de berekeningswijze der regressievergelijkingen, door Donner bepaald ter verklaring van de loop der aandelenkoersen in Duitschland van 1870—1913, zijn bezwaren aan te voeren die de juistheid van zijn uitkomsten twijfelachtig doen zijn.

O. Donner, Die Kursbildung am Aktienmarkt, Sonderh. 36, Vierteljahrsh. zur Konjunkturf., 1934, p. 21.

VIII

Het zou in vele opzichten een besparing betekenen indien de getallen in statistische publicaties slechts in zoveel cijfers werden aangegeven als door de nauwkeurigheid van de meetmethode der betreffende grootheid gewaarborgd zijn.

IX

Onder de verschillende methoden tot beïnvloeding der inkomensverdeling onderscheidt een heffing op de bedrijfswinsten zich hierdoor, dat met de toepassing van dit middel geen onmiddellijke terugwerking op de beslissingen der ondernemers ten aanzien van de omvang van productie en werkgelegenheid verbonden is.

X

Het derde postulaat, door Bowley gesteld aan een „utility function", is onverenigbaar met de eveneens door hem gestelde onmeetbaarheid der bevrediging.

A. L. Bowley, The mathematical groundwork of economics, p. 2.

XI

De bewering van Moore dat, als $\varphi(x)$ voorstelt de kosten van een bedrijf als functie der geproduceerde hoeveelheid x, een bedrijf, dat verkeert in een der drie gevallen

$$\varkappa = \frac{x}{\varphi}\frac{d\varphi}{dx} \gtreqless 1,$$

door hem genoemd resp. *diminishing, constant, increasing relative return*, dientengevolge ook verkeert in het overeenkomstige der drie gevallen

$$\varphi'' = \frac{d^2\varphi}{dx^2} \gtreqless 0,$$

door hem genoemd *diminishing, constant, increasing return*, is onjuist.

H. L. Moore, Synthetic economics, p. 80.

XII

De voorstelling van Moore dat wel door het vervuld zijn van een der relaties $\varkappa \gtreqless 1$ zou zijn vastgelegd, welke der relaties $\varphi'' \gtreqless 0$ geldt, maar niet omgekeerd het teken van φ'' dat van $\varkappa - 1$ zou bepalen, is logisch foutief.

H. L. Moore, l.c. p. 82.

XIII

Het werkt verwarrend dat in de economische litteratuur voor beide onderscheidingen

$$\varkappa \gtrless 1, \qquad \varphi'' \gtrless 0,$$

de benamingen verminderende resp. vermeerderende meeropbrengst worden gebruikt.