On the theory of maximum likelihood estimation of structural relations. Part 1: One dimensional case

ON THE THEORY OF MAXIMUM LIKELIHOOD ESTIMATION OF STRUCTURAL RELATIONS

Part I: One dimensional case

by

J. Jansen and J.F. Barrett

Group Measurement and Control
Department of Electrical Engineering
Eindhoven University of Technology
Eindhoven, The Netherlands

ON THE THEORY OF MAXIMUM LIKELIHOOD ESTIMATION OF STRUCTURAL RELATIONS

Part I: One dimensional case.

by

J. Jansen and J.F. Barrett

TH-Report 78-E-78

September 1977

ISBN 90 6144 0785

Contents

INTRODUCTION

The most commonly used method for experimental determination of functional relationships is that of least squares because it provides a very convenient method for estimating parameters from experimental data. In its usual form, use of least squares is equivalent to the assumption that one of the variables (the dependent variable) has an observation error while the other one (the independent variable) is free from error. This assumption is frequently made in writing down the descriptive equations though often more for the convenience, in order to use least squares, than because it is an accurate, representation of reality: usually both variables, dependent and independent, will be subject to observation error. A similar assumption is also frequently made in system analysis where it is common practice to add noise (usually white noise) to the output while leaving the input free of noise. In linear systems it is, of course, possible to transfer input noise to output noise but if this is done the usual least square theory does not apply.

The problem of determining a functional relationship when both dependent and independent variables are subject to observation error is the problem of structural relationship which is the subject of the present report.

The problem of structural relationship has a fairly long history in the statistical literature going back to an early paper of Adcock (1877). Later, K. Pearson discussed in relation with the regression problem and a number of contributions were also made by other writers notably Van Uven (1930). The fullest account was given by Koopmans (1937) in a book entirely devoted to econometric applications. The more recent literature, beginning e.g. with the paper of Lindley (1947), has focussed attention on the difficulties associated with the maximum likelihood solution of the problem in the case when the errors are Gaussian. Other procedures, based on the idea of generalised least squares of Sprent (1963) - which is essentially the method of Van Uven and Koopmans - have also received attention.

A number of papers have also appeared fairly recently on the corresponding systems analysis problem of determining an input-output relation when both input and output have observation noise. Koopmans was the first to treat this problem in its econometric applications and more recent work beginning with Levin (1964) is strongly influenced by his treatment. The present state of the theory for systems applications is rather incomplete and unsatisfactory and this situation comes about largely because of many unclear points in the theory of the underlying statistical problem.

The present report has the double aim of giving a convenient readable account of basic existing theory and also of clarifying and extending some points of theory. Attention is restricted to the simplest linear relation between two real variables. It is intended that this report should be the basis for further work in extending the theory to relations between vectors and to input-output relations of systems analysis.

The first section describes the well known maximum likelihood solution, presenting it in convenient graphical form and giving attention to the solution of Dent which, though it has its theoretical limitations, is of practical importance. The second method is about the method of generalised least squares and its relation with the maximum likelihood solution. The third section shows how the maximum likelihood formulation may be decomposed into two simpler problems. This decomposition provides the basis for an improved theoretical treatment which automatically includes the generalised least squares principle. The material of this section has not, to the authors' knowledge, previously appeared in the literature. The report concludes with a reasonably complete bibliography.

1. THE PROBLEM OF ESTIMATION OF LINEAR STRUCTURAL RELATIONS WITH GAUSSIAN ERRORS

In this first section we will introduce the subject by describing the maximum likelihood solution of the problem of estimation of linear strutural relations with Gaussian errors in the form it is usually given in statistical texts, for example in the book of Kendall & Stewart (1958) and Graybill (1961). The original discussion along these lines goes back to Dent (1935) and Lindley (1949).

1.1. Statement of the problem

A strutural relation between two variables X and Y is just a functional relation

$$Y = f(X) \qquad (1.1.1.)$$

which requires to be determined by observation. Here we will restrict attention to linear relations

$$Y = aX + b \qquad (1.1.2.)$$

where in general X and Y could be vectors. Since the ideas are most conveniently described when X and Y are real variables, we shall assume this to be the case for the present.

Suppose that the observed values (x,y) of (X,Y) are

$$x = X + \varepsilon \qquad (1.1.3.)$$

$$y = Y + \eta \qquad (1.1.4.)$$

where ε,η are statistically independent Gaussian observation orrors with zero means and standard deviations σ_ε and σ_η respectively.

The joint probability density function of ε and η is thus

$$p(\varepsilon,\eta) = \frac{1}{2\pi\sigma_\varepsilon\sigma_\eta} \exp\left[-\frac{1}{2}\left[\frac{\varepsilon^2}{\sigma_\varepsilon^2} + \frac{\eta^2}{\sigma_\eta^2}\right]\right] \qquad (1.1.5)$$

The problem is to estimate, from a sequence of statistically independent observations $(x_1, x_1),\ldots,(x_n, y_n)$, the parameters a and b defining the linear relation and also, if they are unknown, the standard deviations σ_ε and σ_η of the errors. The parameters $a,b,\sigma_\varepsilon,\sigma_\eta$ are called the <u>structural parameters</u> of the problem. Thus the structural parameters must be found. In order to do this, the usual method of solution also requires estimation of the true values $(X_1,Y_1),\ldots,(X_n,Y_n)$. These are termed the <u>incidental parameters</u> of the problem.

1.2. The Maximum Likelihood Solution

The likelihood function for a single observation is defined by

$$L\,\{(X,Y),a,b,\sigma_\varepsilon,\sigma_\eta;(x,y)\}$$

$$\propto p\{(x,y)\,|\,(X,Y),a,b,\sigma_\varepsilon,\sigma_\eta\} \qquad (1.2.1.)$$

the proportionality sign indicating that the likelihood function is usually left undetermined up to a multiplicative constant. The constant of proportionality will here be taken unity so that

$$L\{(X,Y),a,b,\sigma_\varepsilon,\sigma_\eta;(x,y)\} =$$

$$= \frac{1}{2\pi\sigma_\varepsilon\sigma_\eta} \exp\left[-\frac{1}{2}\left\{\frac{(x-X)^2}{\sigma_\varepsilon^2} + \frac{(y-aX-b)^2}{\sigma_\eta^2}\right\}\right] \qquad (1.2.2.)$$

The likelihood function for a sequence of n independent observations is

$$L\{(X_1,Y_1),\ldots,(X_n,Y_n),a,b,\sigma_\varepsilon,\sigma_\eta;(x_1,y_1),\ldots,(x_n,y_n)\} =$$

$$= \frac{1}{(2\pi)^n\sigma_\varepsilon^n\sigma_\eta^n} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{(x_1-X_i)^2}{\sigma_\varepsilon^2} + \frac{(y_1-aX_i-b)^2}{\sigma_\eta^2}\right\}\right] \qquad (1.2.3.)$$

The maximum likelihood estimates of the parameters are those values which maximise the likelihood L or, what is the same thing, its logarithm ln L which is

$$\ln L = -n \ln 2\pi - n \ln \sigma_\varepsilon - n \ln \sigma_\eta + \\ -\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{(x_1 - X_i)^2}{\sigma_\varepsilon^{\;2}} + \frac{(y_1 - aX_i - b)^2}{\sigma_\eta^{\;2}}\right\} \tag{1.2.4.}$$

The unknown parameters consist of the incidental parameters X_i, $i=1,\ldots\ldots,n$ the structural parameters a,b and possibly $\sigma_\varepsilon,\sigma_\eta$. So we have the conditions

$$\frac{\partial \ln L}{\partial X_i} = -\frac{(X_i - x_1)}{\sigma_\varepsilon^{\;2}} - \frac{a(aX_i + b - y_1)}{\sigma_\eta^{\;2}} = 0 \tag{1.2.5.}$$

$$\frac{\partial \ln L}{\partial a} = -\frac{1}{\sigma_\eta^{\;2}}\sum_{i=1}^{n} X_i(aX_i + b - y_1) = 0 \tag{1.2.6.}$$

$$\frac{\partial \ln L}{\partial b} = -\frac{1}{\sigma_\eta^{\;2}}\sum_{i=1}^{n} (aX_i + b - y_1) = 0 \tag{1.2.7.}$$

If the variances $\sigma_\varepsilon,\sigma_\eta$ are unknown, there will be two additional equations arising from the conditions: $\partial/\partial_{\sigma_\varepsilon} = 0$; $\partial/\partial_{\sigma_\eta} = 0$. These will be considered below.

A more symmetrical solution comes about if Lagrange multipliers are used. In this case we look for an extreme value of

$$F = \text{const} - n \ln \sigma_\varepsilon - n \ln \sigma_\eta + \\ -\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{(x_i - X_i)^2}{\sigma_\varepsilon^{\;2}} + \frac{(y_i - Y_i)^2}{\sigma_\eta^{\;2}}\right\} - \sum_{i=1}^{n} \lambda_i (Y_i - aX_i - b) \tag{1.2.8.}$$

We then have the conditions

$$\frac{\partial F}{\partial X_i} = -\frac{1}{\sigma_\varepsilon^2}(X_i - x_i) + a\lambda_i = 0 \qquad i=1,\dots,n \tag{1.2.9.}$$

$$\frac{\partial F}{\partial Y_i} = -\frac{1}{\sigma_\eta^2}(Y_i - y_i) - \lambda_i = 0 \qquad i=1,\dots,n \tag{1.2.10.}$$

$$\frac{\partial F}{\partial a} = \sum_{i=1}^{n} \lambda_i X_i = 0 \tag{1.2.11.}$$

$$\frac{\partial F}{\partial b} = \sum_{i=1}^{n} \lambda_i = 0 \tag{1.2.12.}$$

The equations of constraint are

$$\frac{\partial F}{\partial \lambda_i} = -Y_i + aX_i + b = 0 \qquad i=1,\dots,n \tag{1.2.13.}$$

These equations are equivalent to the previous ones but have a more symmetrical form.

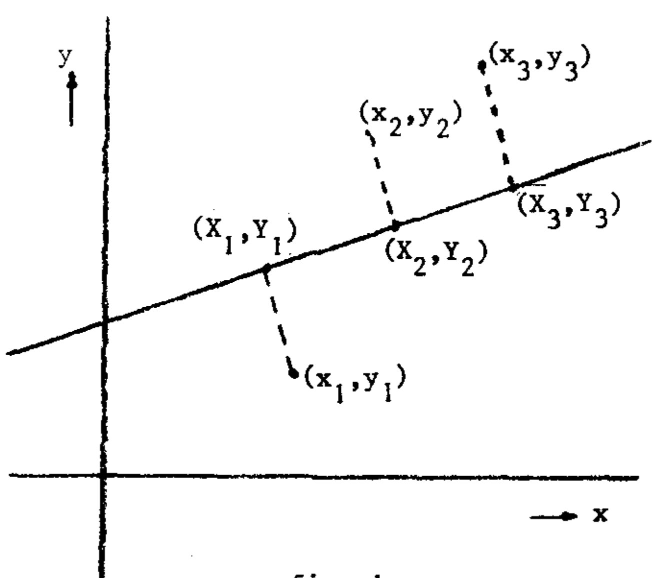

fig. 1.

From these equations we obtain

$$X_i - x_i = a\lambda_i\sigma_\varepsilon^2 \qquad (1.2.14.)$$

$$Y_i - y_i = -\lambda_i\sigma_\eta^2 \qquad (1.2.15.)$$

From which it follows, by summation, that

$$\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} x_i = a\sum_{i=1}^{n} \lambda_i\sigma_\varepsilon^2 = 0 \qquad (1.2.16.)$$

Thus

$$\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} x_i \qquad (1.2.17.)$$

i.e.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \qquad (1.2.18.)$$

Similarly

$$\bar{Y} = \bar{y} \qquad (1.2.19.)$$

We also deduce that

$$(X_i - x_i) : (Y_i - y_i) = - a\sigma_\varepsilon^2\ \sigma_\eta^2 \qquad (1.1.20.)$$

which shows that all the vectors

$$(X_i - x_i,\ Y_i - y_i) \qquad (1.2.21)$$

which project observed points on to the line are parallel (fig.1)

Further, since

$$X_i = x_i + \lambda_i a\sigma_\varepsilon^2 \qquad (1.2.22.)$$

$$Y_i = y_i - \lambda_i\sigma_\eta^2 \qquad (1.2.23.)$$

we have

$$(Y_i - aX_i - b) = (y_i - ax_i - b) - \lambda_i(\sigma_\eta^2 + a^2\sigma_\varepsilon^2) \tag{1.2.24.}$$

The left hand side vanishes, consequently

$$\lambda_i = \frac{y_i - ax_i - b}{\sigma_\eta^2 + a^2\sigma_\varepsilon^2} \tag{1.2.15.}$$

By summation using (1.2.12.) we see that

$$\sum_{i=1}^{n} (y_i - ax_i - b) = 0 \tag{1.2.26.}$$

or

$$b = \bar{y} - a\bar{x} \tag{1.2.27.}$$

The estimated line may therefore be written

$$(y - \bar{y}) = a(x - \bar{x}) \tag{1.2.29.}$$

and so passes through the common centroid of the observations and of the estimated values (X_i, Y_i), $i = 1,\dots,n$

From (1.2.14.), (1.2.18.) and (1.2.25.) we deduce that

$$\begin{aligned} X_i - \bar{X} &= x_i - \bar{x} + \frac{\{(y_i - \bar{y}) - a(x_i - \bar{x})\}}{\sigma_\eta^2 + a^2\sigma_\varepsilon^2} \cdot a\sigma_\varepsilon^2 \\ &= \frac{\sigma_\eta^2(x_i - \bar{x}) + a\sigma_\varepsilon^2(y_i - \bar{y})}{\sigma_\eta^2 + a^2\sigma_\varepsilon^2} \end{aligned} \tag{1.2.30.}$$

Now from (1.2.11.) and (1.2.12.)

$$\sum_{i=1}^{n} \lambda_i(X_i - \bar{X}) = 0 \tag{1.2.31.}$$

and by substitution from 1.2.25.), (1.2.27.) and (1.2.30.)

$$\sum_{i=1}^{n} \{(y_i - \bar{y}) - a(x_i - \bar{x})\}\{\sigma_\eta^2 (x_i - \bar{x}) + a\sigma_\varepsilon^2 (y_i - \bar{y})\} = 0 \qquad (1.2.32.)$$

which is

$$a^2\sigma_\varepsilon^2 s_{xy} - a(\sigma_\varepsilon^2 s_{yy} - \sigma_\eta^2 s_{xx}) - \sigma_\eta^2 s_{xy} = 0 \qquad (1.2.33.)$$

where

$$s_{xx} = \overline{(x - \bar{x})^2} \qquad (1.2.34.)$$

$$s_{xy} = \overline{(x - \bar{x})(y - \bar{y})} \qquad (1.2.35.)$$

$$s_{yy} = \overline{(y - \bar{y})^2} \qquad (1.2.37.)$$

The solution for a is

$$a = \frac{s_{yy} - k^2 s_{xx} + \sqrt{\{(s_{yy} - k^2 s_{xx})^2 + 4k^2 s_{xy}^2\}}}{2s_{xy}} \qquad (1.2.38.)$$

and thus depends on the ratio of the variances

$$k = \frac{\sigma_\eta}{\sigma_\varepsilon} \qquad (1.2.39.)$$

It is not difficult to show (see appendix 1) that this estimate of a changes monotonically from one regression line to the other as k increases from 0 to ∞.

1.3. Scale-free Form for the Estimate of Slope

On introducing the parameter

$$\Theta = \frac{\dfrac{s_{yy}}{\sigma_\eta^2} - \dfrac{s_{xx}}{\sigma_\eta^2}}{\dfrac{2s_{xy}}{\sigma_\varepsilon \sigma_\eta}} \qquad (1.3.1.)$$

the equation for slope assumes the simpler form

$$(\frac{a}{k})^2 - 2\Theta(\frac{a}{k}) - 1 = 0 \qquad (1.3.2.)$$

with solution

$$\frac{a}{k} = \Theta + \sqrt{\Theta^2 + 1} \qquad (1.3.3.)$$

The ratio a/k may be determined conveniently by the following trigonometrical method. Find θ such that

$$\cot 2\theta = \Theta \qquad (1.3.4.)$$

and then, in view of the identity

$$\cot \theta = \cot 2\theta + \sqrt{\cot^2 2\theta + 1} \qquad (1.3.5.)$$

we get

$$\frac{a}{k} = \cot \theta \qquad (1.3.6.)$$

so a is determined.

Note that the parameters a/k and Θ are independent of scaling along the x- and y- axes. The parameter k, on which Θ depends, is not independent of scale. However, in place of k we may take

$$k^1 = \frac{\sigma_\eta}{\sigma_\varepsilon} \cdot \sqrt{\frac{s_{yy}}{s_{xx}}} \qquad (1.3.7.)$$

or its inverse, as a scale-free parameter equivalent to k. Θ may be expressed in terms of k as follows.

$$\Theta = \frac{\dfrac{s_{yy}}{\sigma_\eta^2} - \dfrac{s_{xx}}{\sigma_\varepsilon^2}}{\dfrac{2s_{xy}}{\sigma_\varepsilon \sigma_\eta}}$$

$$= \frac{\sigma_\varepsilon}{\sigma_\eta} \cdot \frac{s_{xx}}{s_{yy}} - \frac{\sigma_\eta}{\sigma_\varepsilon} \cdot \frac{s_{xx}}{s_{yy}}$$

$$= \frac{\sqrt{s_{xx} s_{yy}}}{s_{xy}} \left(\frac{\sigma_\varepsilon}{\sigma_\eta} \sqrt{\frac{s_{yy}}{s_{xx}}} - \frac{\sigma_\eta}{\sigma_\varepsilon} \sqrt{\frac{s_{xx}}{s_{yy}}} \right) \qquad (1.3.8.)$$

Here

$$r = \frac{s_{xy}}{\sqrt{s_{xx} s_{yy}}} \qquad (1.3.9.)$$

is the empirical correlation coefficient which is also scale-free. Now the equation may be put in symmetrical form as

$$\frac{a\sigma_\varepsilon}{\sigma_\eta} - \frac{\sigma_\eta}{a\sigma_\varepsilon} = \frac{\sqrt{s_{xx} s_{yy}}}{s_{xy}} \left(\frac{\sigma_\varepsilon}{\sigma_\eta} \ \frac{s_{yy}}{s_{xx}} - \frac{\sigma_\eta}{\sigma_\varepsilon} \ \frac{s_{xx}}{s_{yy}} \right) \qquad (1.3.10.)$$

and, defining (scale-free) angles θ_1, θ_2 in the range $(0, \frac{\pi}{2})$ by

$$\cot \theta_1 = \frac{a\sigma_\varepsilon}{\sigma_\eta} \qquad (1.3.11.)$$

$$\cot \theta_2 = \frac{\sigma_\varepsilon}{\sigma_\eta} \cdot \frac{s_{xx}}{s_{yy}} \qquad (1.3.12.)$$

the equation takes the form

$$\cot 2\theta_1 = \frac{1}{r} \cot 2\theta_2 \qquad (1.3.13.)$$

If now θ_2 is plotted against θ_1, the result is the symmetrical graph shown in fig. 2 which is also scale-free.

From this graph we see that if the errors are small and so the observations are well correlated and $r \simeq 1$, then to reasonable practical approximation, $\theta_2 \simeq \theta_1$ giving

$$\frac{a\sigma_\varepsilon}{\sigma_\eta} = \cot\theta_1 \simeq \cot\theta_2 = \frac{\sigma_\varepsilon}{\sigma_\eta}\sqrt{\frac{s_{yy}}{s_{xx}}} \qquad (1.3.14.)$$

Thus, independently of the ratio $k = \sigma_\eta/\sigma_\varepsilon$

$$a \simeq \sqrt{\frac{s_{yy}}{s_{xx}}} \qquad (1.3.15.)$$

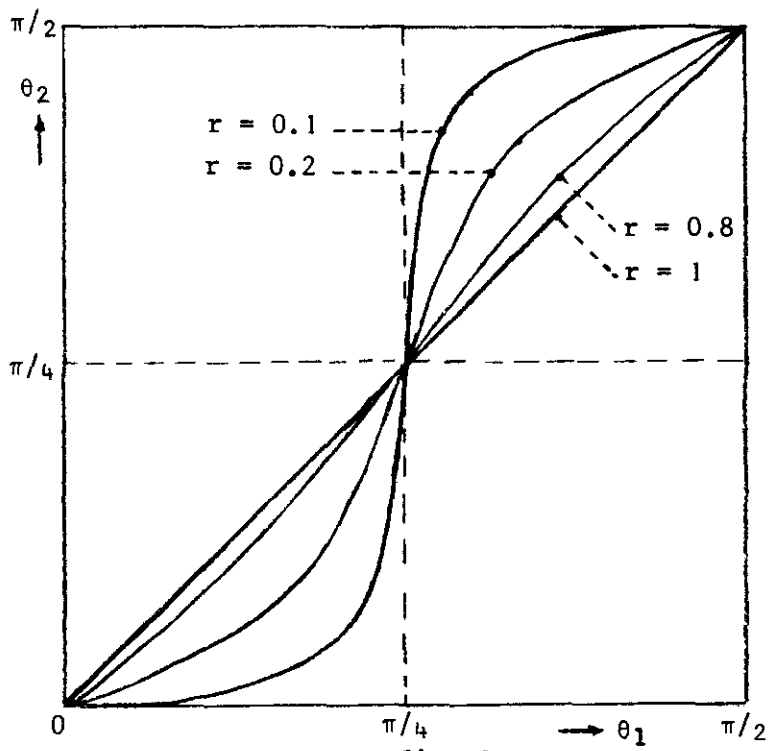

fig. 2

1.4. The Case of Unknown Variances

If the variances are unknown, they can be estimated by using the two further conditions

$$\frac{\partial F}{\partial \sigma_\varepsilon} = -\frac{n}{\sigma_\varepsilon} + \frac{1}{\sigma_\varepsilon^3} \sum_{i=1}^{n} (x_i - X_i)^2 = 0 \qquad (1.4.1.)$$

$$\frac{\partial F}{\partial \sigma_\eta} = -\frac{n}{\sigma_\eta} + \frac{1}{\sigma_\eta^3} \sum_{i=1}^{n} (y_i - Y_i)^2 = 0 \qquad (1.4.2.)$$

which give

$$\sigma_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - X_i)^2 \qquad (1.4.3.)$$

$$\sigma_\eta^2 = \frac{1}{n} \sum_{i=1}^{n} (y_i - Y_i)^2 \qquad (1.4.4.)$$

Substituting the ratio $(x_i - X_i):(y_i - Y_i)$ from (1.1.20) it follows that

$$\frac{\sigma_\varepsilon^2}{\sigma_\eta^2} = \frac{\sum_{i=1}^{n} (x_i - X_i)^2}{\sum_{i=1}^{n} (y_i - Y_i)^2} = \frac{a^2\sigma_\varepsilon^4}{\sigma_\eta^4} \qquad (1.4.5.)$$

Hence

$$\frac{a^2\sigma_\varepsilon^2}{\sigma_\eta^2} = 1 \qquad (1.4.6.)$$

giving

$$\frac{a\sigma_\varepsilon}{\sigma_\eta} = \pm 1 \qquad (1.4.7.)$$

A more detailed analysis (Solari, 1963) shows that the positive sign gives the larger value of ln L. Thus

$$\frac{a\sigma_\varepsilon}{\sigma_\eta} = 1 \qquad (1.4.8.)$$

From equations (1.4.3) it then follows that also

$$\frac{\sigma_\varepsilon}{\sigma_\eta} \frac{s_{yy}}{s_{xx}} = \frac{\sigma_\eta}{\sigma_\varepsilon} \frac{s_{xx}}{s_{yy}} \tag{1.4.9.}$$

or

$$\frac{\sigma_\varepsilon}{\sigma_\eta} \frac{s_{yy}}{s_{xx}} = \pm 1 \tag{1.4.10.}$$

The positive sign is chosen because all quantities on the left are positive. Thus from (1.4.8.) and (1.4.10.)

$$a = \frac{\sigma_\eta}{\sigma_\varepsilon} = \sqrt{\frac{s_{yy}}{s_{xx}}} \tag{1.4.11.}$$

Thus, both the gradient of the estimated straight line and the ratio of variances are estimated by the quantity $\sqrt{s_{yy}/s_{xx}}$. The estimated line is

$$(y - \bar{y}) = \sqrt{\frac{s_{yy}}{s_{xx}}} \quad (x - \bar{x}) \tag{1.4.12.}$$

which has a gradient which is the geometric mean of the gradients of the two regression lines i.e. s_{yy}/s_{xy} and s_{xy}/s_{xx} This solution is due to Dent (1935). We see that it agrees with the result suggested by the graph in fig. 2.

In general, the gradient of this estimate will have a bias and this remains true even if the number of observations tends to infinity i.e. the estimate is "inconsistent". For, asuming that the observations (X_i, Y_i) possess a finite mean and finite variances σ_X^2, σ_Y^2 as $n \to \infty$ we shall have asymptotically,

$$\frac{s_{yy}}{s_{xx}} \to \frac{\sigma_Y^2 + \sigma_\eta^2}{\sigma_X^2 + \sigma_\varepsilon^2} \tag{1.4.13.}$$

and since

$$\frac{\sigma_Y^2}{\sigma_X^2} = a^2 \lessgtr \frac{\sigma_Y^2 + \sigma_\eta^2}{\sigma_X^2 + \sigma_\varepsilon^2} \lessgtr \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$$

with equality only if $\sigma_\eta/\sigma_\varepsilon = a$, we see that, unless $\sigma_\eta/\sigma_\varepsilon = a$ the estimate $\sqrt{s_{yy}/s_{xx}}$ will lie between the true values of a and $\sigma_\eta/\sigma_\varepsilon$ overestimating the one and underestimating the other.

2. GENERALISED LEAST SQUARES

This section shows how the maximum likelihood solution of the last section can be given a geometrical interpretation which is a generalisation of that used by K. Pearson (1901) and other early writers. This approach leads to the method of generalised least squares of Sprent (1970).

2.1. The Use of Homogeneous Line Coordinates

In the further discussion it will be convenient to use homogeneous line coordinates α, β, ν for the undetermined linear relation writing it as

$$\alpha X + \beta Y = \nu \tag{2.1.1.}$$

The log-likelihood function for n independent observations is then, as before,

$$\ln L\{(x_i, y_i), i = 1,\dots,n, (X_i, Y_i), i = 1,\dots,n; \sigma_\varepsilon, \sigma_\eta\} =$$

$$= -\ln(2\pi\sigma_\varepsilon\sigma_\eta) - \tfrac{1}{2}\sum_{i=1}^{n}\left\{\frac{(x_i - X_i)^2}{\sigma_\varepsilon^2} + \frac{(y_i - Y_i)^2}{\sigma_\eta^2}\right\} \tag{2.1.2.}$$

It does not explicitly depend on the parameters of the line. It must be maximised subject to the constraints

$$\alpha X_i + \beta Y_i = \nu \qquad i=1,\dots,n \tag{2.1.3.}$$

and so we introduce the function

$$F\{(x, y)\ i=1,\dots,n; (X_i, Y_i)\ i=1,\dots,n; \alpha, \beta, \nu, \sigma_\varepsilon, \sigma_\eta\} =$$

$$= \ln L - \sum_{i=1}^{n} \lambda_i(\alpha X_i + \beta Y_i - \nu) \tag{2.1.4.}$$

which does depend on the line parameters.

The conditions for vanishing first derivatives of F then leads to the following equations which are equivalent to those previously given:

$$\frac{\partial F}{\partial X_i} = -\frac{(X_i - x_i)}{\sigma_\varepsilon^2} - \alpha\lambda_i = 0 \qquad (2.1.5.)$$

$$\frac{\partial F}{\partial Y_i} = -\frac{(Y_i - y_i)}{\sigma_\eta^2} - \beta\lambda_i = 0 \qquad (2.1.6.)$$

$$\frac{\partial F}{\partial \alpha} = -\sum_{i=1}^{n} \lambda_i X_i = 0 \qquad (2.1.7.)$$

$$\frac{\partial F}{\partial \beta} = -\sum_{i=1}^{n} \lambda_i Y_i = 0 \qquad (2.1.8.)$$

$$\frac{\partial F}{\partial \gamma} = \sum_{i=1}^{n} \lambda_i = 0 \qquad (2.1.9.)$$

$$\frac{\partial F}{\partial \sigma_\varepsilon} = -\frac{n}{\sigma_\varepsilon} + \frac{1}{\sigma_\varepsilon^3} \sum_{i=1}^{n} (x_i - X_i)^2 = 0 \qquad (2.1.10.)$$

$$\frac{\partial F}{\partial \sigma_\eta} = -\frac{n}{\sigma_\eta} + \frac{1}{\alpha_\eta^3} \sum_{i=1}^{n} (y_i - Y_i)^2 = 0 \qquad (2.1.11.)$$

The solution of these equations and the derivation of the equation for the estimate of the ratio $\beta:\alpha$ (which now takes the place of the parameter a) follows the same procedure as in the last section.

We shall here note the principal formula which will be needed in what follows.

We have immediately

$$X_i = x_i - \lambda_i \alpha\, \sigma_\varepsilon^2 \qquad (2.1.12.)$$

$$Y_i = y_i - \lambda_i \beta\alpha_\eta^2 \qquad (2.1.13.)$$

from which

$$(\alpha x_i + \beta y_i - \nu) - \lambda_i(\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2) = 0 \qquad (2.1.14.)$$

and so

$$\lambda_i = \frac{\alpha x_i + \beta y_i - \nu}{\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2} \qquad (2.1.15.)$$

By summation and condition (2.1.9.)

$$\nu = \alpha\bar{x} + \beta\bar{y} \qquad (2.1.16.)$$

so that

$$\lambda_i = \frac{\alpha(x_i - \bar{x}) + \beta(y_i - \bar{y})}{(\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2)} \qquad (2.1.17.)$$

The equation corresponding to (1.2.33.) is

$$\alpha^2\sigma_\varepsilon^2 s_{xy} - \alpha\beta(\sigma_\varepsilon^2 s_{yy} - \sigma_\eta^2 s_{xx}) - \beta^2\sigma_\eta^2 s_{xy} \qquad (2.1.18.)$$

2.2. Geometrical Interpretation: Generalised Least Squares

Suppose we consider the case of equal variances

$$\sigma_\varepsilon = \sigma_\eta = \sigma \qquad (2.2.1.)$$

The log-likelihood function of n observations is then

$$\ln L = - n\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\{(x_i - X_i)^2 + (y_i - Y_i)^2\} \qquad (2.2.2.)$$

The problem is to maximise this when (X_i, Y_i) $i=1,\ldots,n$ lie on the line

$$\alpha X + \beta Y = \nu \qquad (2.2.3.)$$

As regards the choice of the (X_i, Y_i) we must <u>minimise</u> the sum of the squared distances

$$\sum_{i=1}^{n} \{(x_i - X_i)^2 + (y_i - Y_i)^2\} \tag{2.2.4.}$$

from the observed points (x_i, y_i) to the actual points (X_i, Y_i) lying on the given line. This means that the squared distance from each (X_i, Y_i) to its corresponding (x_i, y_i) must be minimised. Now the expression

$$(x - X)^2 + (y - Y)^2 \tag{2.2.5.}$$

is minimised when (X, Y) is the foot of the perpendicular from (x, y) on to the line, i.e. the points where a circle with centre (x, y) just touches the line. Thus each $(\tilde{X}_i, \tilde{Y}_i)$ is obtained by perpendicular projection of the observed point (x_i, y_i) on to the line.

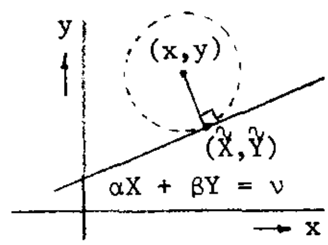

fig. 3.

Having minimised ln L with respect to the points (X_i, Y_i) it is then necessary to minimise it with respect to the line parameters and the error variance (if this is unknown). Minimisation with respect to the line parameters is just the problem of finding a line of closest fit to the observed points in the sense of minimisation of the sum of squared perpendicular distances from these points to the line.
In this way we have arrived at a generalisation of the principle of least squares.

Now let us consider the case when the error variances are unequal in which case the log-likelihood function is given by (2.1.2.) Suppose we take as distance function between points (x, y) and (X, Y) the value

$$d^2 = \frac{(x - X)^2}{\sigma_\varepsilon^2} + \frac{(y - Y)^2}{\sigma_\eta^2} \tag{2.2.6.}$$

which is just the distance between these points of the axes are re-scaled so that the error variances are both unity. In order to maximise ln L it is then again necessary to choose (X_i, Y_i) to minimise the sum of squared distances from the observed points. This means that each point (X_i, Y_i)

must be chosen at minimum distance from the corresponding point (x_i, y_i).

So we are lead to the following geometrical construction. Suppose that with (x, y) as centre ellipses

$$\frac{(x - X)^2}{\sigma_\varepsilon^2} + \frac{(y - Y)^2}{\sigma_\eta^2} = \text{const.} \qquad (2.2.7.)$$

are constructed giving the locus of point at equal distance from (c, y). There will be one ellipse which just touches the line, say at the point $(\tilde{X}, \tilde{Y})$ (see diagram). It is clear that this point is the point on the line at the least distance from (x, y). The line from the centre (x, y) *to the point of contact* $(\tilde{X}, \tilde{Y})$ is no longer perpendicular but is the conjugate direction to the line with respect to the ellipse. It is easy to see that a general point (X', Y') on this line satisfies

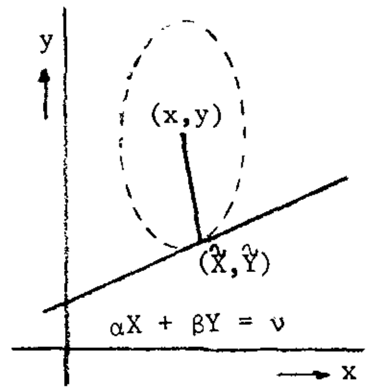

fig. 4.

$$\beta\sigma_\eta^2(X' - x) = \alpha\sigma_\varepsilon^2(Y' - y) \qquad (2.2.8.)$$

From (2.1.12.) and (2.1.15.) we get

$$\tilde{X} = x - \alpha\sigma_\varepsilon^2 \frac{(\alpha x + \beta y - \nu)}{\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2} \qquad (2.2.9.)$$

and similarly

$$\tilde{Y} = y - \beta\sigma_\eta^2 \frac{(\alpha x + \beta y - \nu)}{\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2} \qquad (2.2.10.)$$

from which we get the constant of the ellipse, giving the squared distance from (x, y) to the line, as

$$d^2 = \frac{(\alpha x + \beta y - \nu)^2}{\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2} \qquad (2.2.11.)$$

When there are n independent observations (x_i, y_i) each observation is projected in the same direction conjugate to the line on to a point (X_i, Y_i) on the line. The resulting sum of squared distances is

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} \frac{(\alpha x + \beta y - \nu)^2}{\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2} \tag{2.2.12.}$$

This must be minimised with respect to the line parameters giving the line of closest fit to a system of ellipses centred at the observation points.

An equivalent statement is that the ratio

$$\frac{\overline{(\alpha x + \beta y - \nu)^2}}{\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2} \tag{2.2.13.}$$

must be minimised with respect to the line parameters, the bar denoting mean value over the observations.
Since the line parameters are homogeneous and only ratios have a significance, the minimisation problem can be put in the following form which we will call

THE PRINCIPLE OF GENERALISED LEAST SQUARES: the line parameters of the maximum likelihood solution may be obtained as the solution of the minimisation problem

> $\overline{(\alpha x + \beta y - \nu)^2}$ is a minimum with respect to α, β, ν
> subject to the constraint
> $(\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2) = \text{const.}$

In interpreting this principle, note that in view of the constraint on (X, Y) we have

$$\alpha x + \beta y - \nu = \zeta \tag{2.2.14.}$$

where

$$\zeta = \alpha\varepsilon + \beta\eta \tag{2.2.15.}$$

The variance of ζ is

$$\sigma_\zeta^2 = \alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2 \qquad (2.2.16.)$$

may be regarded as that part of the error which measures deviation from the given line.

2.3. Derivation of Line Parameters Using Generalised Least Squares

We shall rederive the solution for the line parameters using the minimisation formulation of generalised least squares. First we write

$$\begin{aligned}\overline{(\alpha x + \beta y - \nu)^2} &= \alpha^2\overline{x}^2 + 2\alpha\beta\overline{\overline{xy}} + \beta^2\overline{y}^2 \\ &\quad - 2\alpha\nu\overline{x} - 2\beta\nu\overline{y} + \nu^2\end{aligned} \qquad (2.3.1.)$$

Minimisation with respect to ν which is unconstrained immediately gives

$$\nu = \alpha\overline{x} + \beta\overline{y} \qquad (2.3.2.)$$

The estimated line thus has the form

$$\alpha(x - \overline{X}) + \beta(y - \overline{Y}) = 0 \qquad (2.3.3.)$$

and passes through the centroid of observations. The quadratic in α and β becomes

$$\begin{aligned}\overline{(\alpha x + \beta y - \nu)^2} &= \overline{\{\alpha(x - \overline{x}) + \beta(y - \overline{y})\}^2} \\ &= \alpha^2 s_{xx} + 2\,\alpha\beta s_{xy} + \beta^2 s_{yy}\end{aligned} \qquad (2.3.4.)$$

where

$$s_{xx} = \overline{(x - \overline{x})^2} \qquad (2.3.5.)$$

$$s_{xy} = \overline{(x - \overline{x})(y - \overline{y})} \qquad (2.3.6.)$$

$$s_{yy} = \overline{(y - \overline{y})^2} \qquad (2.3.7.)$$

The minimisation problem now becomes

$$\begin{cases} (\alpha^2 s_{xx} + 2\,\alpha\beta s_{xy} + \beta^2 s_{yy}) \quad \text{minimum with respect} \\ \text{to } \alpha,\ \beta \text{ subject to} \\ (\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2) = \text{const.} = s^2 \end{cases}$$

This is a well known minimisation problem. It may be solved either trigonometrically or by the use of a Lagrange multiplier.

The trigonometrical representation method: we put

$$\alpha\sigma_\varepsilon = s \cos\theta \tag{2.3.10.}$$

$$\beta\sigma_\eta = s \sin\theta \tag{2.3.11.}$$

when the constraint is automatically satisfied. Then we must minimise

$$\frac{s_{xx}}{\sigma_\varepsilon^2}\cos^2\theta + 2\,\frac{s_{xy}}{\sigma_\varepsilon\sigma_\eta}\cos\theta\,\sin\theta + \frac{s_{yy}}{\sigma_\eta^2}\sin^2\theta \tag{2.3.12.}$$

The condition $\partial/\partial\theta = 0$ gives

$$-\frac{s_{xx}}{\sigma_\varepsilon^2}\,2\cos\theta\sin\theta + 2\,\frac{s_{xy}}{\sigma_\varepsilon\sigma_\eta}(-\sin^2\theta + \cos^2\theta) +$$

$$+ 2\,\frac{s_{yy}}{\sigma_\eta^2}\ \sin\theta\,\cos\theta = 0 \tag{2.3.13.}$$

or

$$\left(\frac{s_{yy}}{\sigma_\eta^2} - \frac{s_{xx}}{\sigma_\varepsilon^2}\right)\sin 2\theta + 2\frac{s_{xy}}{\sigma_\varepsilon\sigma_\eta}\cos 2\,\theta = 0 \tag{2.3.14.}$$

which is the same as (1.3.1.), (1.3.4.).

The Lagrange multiplier method: using a Lagrange parameter μ the minimisation problem becomes

$$\alpha^2 s_{xx} + 2\alpha\beta s_{xy} + \beta^2 s_{yy} - \mu\ (\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2)\ \min \tag{2.3.15.}$$

Equating derivatives with respect to α and β to zero we get

$$(s_{xx} - \mu\sigma^2_\varepsilon)\alpha + s_{xy}\ \beta = 0 \tag{2.3.16.}$$

$$s_{yx}\ \alpha + (s_{yy} - \mu\sigma^2_\eta)\ \beta = 0 \tag{2.3.17.}$$

For a non-zero solution it is necessary that

$$\begin{vmatrix} (s_{xx} - \mu\sigma^2_\varepsilon) & s_{xy} \\ s_{yx} & (s_{yy} - \mu\sigma^2_\eta) \end{vmatrix} = 0 \tag{2.3.18.}$$

giving μ as one of the roots of

$$\mu^2 - \mu\left(\frac{s_{xx}}{\sigma^2_\varepsilon} + \frac{s_{yy}}{\sigma^2_\eta}\right) + \frac{(s_{xx}s_{yy} - s_{xy})^2}{\sigma^2_\varepsilon\ \sigma^2_\eta} = 0 \tag{2.3.19.}$$

which are

$$\mu = \tfrac{1}{2}\left\{\frac{s_{xx}}{\sigma^2_\varepsilon} + \frac{s_{yy}}{\sigma^2_\eta} \pm \sqrt{\left(\frac{s_{xx}}{\sigma^2_\varepsilon} - \frac{s_{yy}}{\sigma^2_\eta}\right)^2 + 4\,\frac{s^2_{xy}}{\sigma^2_\sigma\sigma^2_\eta}}\right\} \tag{2.3.20.}$$

For each of these roots, values of α and β may be found satisfying the linear equations above and for those particular values we see that

$$\mu = \frac{\alpha^2 s_{xx} + 2\alpha\beta s_{xy} + \beta^2 s_{yy}}{\alpha^2\sigma^2_\varepsilon + \beta^2\sigma^2_\eta} \tag{2.3.21.}$$

and thus the two roots μ give respectively the maximum and minimum values of the ratio on the right which are achieved for the corresponding values α and β. Since we are looking for the minimum value of the ratio, the root with the negative sign must be chosen.

The corresponding value of the ratio β/α is

$$\frac{\beta}{\alpha} = - \frac{(s_{xx} - \mu\sigma_{\varepsilon}^2)}{s_{xy}}$$

$$= \frac{\frac{s_{yy}}{k^2} - s_{xx} + \sqrt{(\frac{s_{yy}}{k^2} - s_{xx})^2 + \frac{4s_{xy}^2}{k^2}}}{s_{xy}} \qquad (2.3.22.)$$

By comparing the ratio β/α of eq. (2.3.22.) with $1/a$ from equation (1.2.38.) we conclude that they are equal.

3. SOLUTION OF THE STRUCTURAL ESTIMATION PROBLEM BY A COMBINED BAYESIAN AND MAXIMUM LIKELIHOOD APPROACH

In the usual solution of the structural estimation problem, as described up to now, both incidental and structural parameters are estimated by the maximum likelihood method. True values (X_i, Y_i) are estimated by parallel projection from the observed points (x_i, y_i). It is clear that, for a large number of observations, the resulting configuration of fig. 1, far from being one of maximum likelihood, is extremely improbable. In view of this, it is not obvious why the calculation gives acceptable results in most (though not all) aspects. In order to explain this and give a more satisfactory theoretical basis to the solution, it is necessary to combine Bayesian and maximum likelihood methods of estimation, using Bayes for the incidental parameters and maximum likelihood for the structural parameters with a modified likelihood function. The present section will show how this can be done.

3.1. On Bayesian Estimation of the True Values

We first discuss the Bayesian estimation of the true values (X_i, Y_i) which are the incidental parameters in the problem. Let us consider the result of making one observation (x, y) of a pair of true values (X, Y). For convenience, we shall denote the totality of structural parameters by π:

$$\pi = (\alpha, \beta, \nu, \sigma_\varepsilon, \sigma_\eta) \qquad (3.1.1.)$$

In the Bayesian view, (x, y) and π are given and (X, Y) has a corresponding conditional distribution on the estimated line. The probability density of (X, Y) as proportional to the likelihood function i.e.

$$p(X,Y|\pi,x,y,) \propto \exp\left[-\tfrac{1}{2}\left\{\frac{(X - x)^2}{\sigma_\varepsilon^2} + \frac{(Y - y)^2}{\sigma_\eta^2}\right\} \right] \qquad (3.1.2.)$$

Now, since the vectors

$$(X - \tilde{X},\ Y - \tilde{Y}) \text{ and } (x - \tilde{X},\ y - \tilde{Y}) \qquad (3.1.3.)$$

are conjugate with respect to the ellipse centred at (x,y), as in obvious from fig. 4, we have

$$\frac{(X - \tilde{X})(\tilde{X} - x)}{\sigma_\varepsilon^2} + \frac{(Y - \tilde{Y})(\tilde{Y} - y)}{\sigma_\eta^2} = 0 \qquad (3.1.4.)$$

from which follows the identity

$$\frac{(X - x)^2}{\sigma_\varepsilon^2} + \frac{(Y - y)^2}{\sigma_\eta^2} = \frac{(X - \tilde{X})^2}{\sigma_\varepsilon^2} + \frac{(Y - \tilde{Y})^2}{\sigma_\eta^2} +$$

$$\frac{(x - \tilde{X})^2}{\sigma_\varepsilon^2} + \frac{(y - \tilde{Y})^2}{\sigma_\eta^2} \qquad (3.1.5.)$$

Since X and Y occur in only the first two terms on the right hand side, we find

$$p(X,Y|\pi,x,y) \sim \exp\left[-\tfrac{1}{2}\left\{\frac{X - \tilde{X})^2}{\sigma_\varepsilon^2} + \frac{(Y - \tilde{Y})^2}{\sigma_\eta^2}\right\}\right] \qquad (3.1.6.)$$

Since the probability distribution of (X,Y) is confined to a line which contains (X,Y), we see that the distribution is Gaussian with its mean at $(\tilde{X},\tilde{Y})$. Although it has the appearance of a two dimensional distribution, it is in reality one-dimensional since the vectors $(X - \tilde{X})$, $(Y - \tilde{Y})$ are proportional. To bring it to one-dimensional form it is convenient to introduce variables along the two conjugate directions. This is done as follows. It follows easily from the previous formulae that

$$x - \tilde{X} = \frac{\beta\sigma_\varepsilon^2}{S^2}\{\alpha(x - X) + \beta(y - Y)\} \qquad (3.1.7.)$$

$$y - \tilde{Y} = \frac{\beta\sigma_\eta^2}{S^2}\{\alpha(x - X) + \beta(y - Y)\} \qquad (3.1.8.)$$

Then by subtraction from x - X and y - Y we get

$$X - \tilde{X} = \frac{\beta\sigma_\varepsilon\sigma_\eta}{S^2}\{-\frac{\beta\sigma_\eta}{\sigma_\varepsilon}(x - X) + \frac{\alpha\sigma_\varepsilon}{\sigma_\eta}(y - Y)\} \qquad (3.1.9.)$$

$$Y - \hat{Y} = -\frac{\alpha\sigma_\varepsilon\sigma_\eta}{s^2}\{-\frac{\beta\sigma_\eta}{\sigma_\varepsilon}(x - X) + \frac{\alpha\sigma_\varepsilon}{\sigma_\eta}(y - Y)\} \tag{3.1.10.}$$

Now put

$$\zeta = \alpha(x - X) + \beta(y - Y) \tag{3.1.11.}$$

$$\omega = \beta\frac{\sigma_\eta}{\sigma_\varepsilon}(x - X) + \alpha\frac{\sigma_\varepsilon}{\sigma_\eta}(y - Y) \tag{3.1.12.}$$

Then

$$x - \hat{X} = \frac{\alpha\sigma_\varepsilon^2}{s^2}\zeta \tag{3.1.13.}$$

$$y - \hat{Y} = \frac{\beta\sigma_\eta^2}{s^2}\zeta \tag{3.1.14.}$$

and

$$X - \tilde{X} = \beta\omega \tag{3.1.15.}$$

$$Y - \tilde{Y} = -\alpha\omega \tag{3.1.16.}$$

Notice that, in terms of the error variables ε and η we can write

$$\zeta = \alpha\varepsilon + \beta\eta \tag{3.1.17.}$$

$$\omega = -\beta\frac{\sigma_\eta}{\sigma_\varepsilon}\varepsilon + \alpha\frac{\sigma_\varepsilon}{\sigma_\eta}\eta \tag{3.1.18.}$$

from which we see that ζ and ω are uncorrelated components of the error with variances

$$\sigma_\varepsilon^2 = \alpha^2\sigma_\varepsilon^2 + \beta\sigma_\eta^2 = s^2 \tag{3.1.19.}$$

$$\sigma_\omega^2 = \beta^2\sigma_\eta^2 + \alpha^2\sigma_\varepsilon^2 = s^2 \tag{3.1.20.}$$

Further we get

$$\frac{(X - \hat{X})^2}{\sigma_\varepsilon^2} + \frac{(Y - \hat{Y})^2}{\sigma_\eta^2} = \frac{\omega^2}{\left(\frac{\sigma_\varepsilon\sigma_\eta}{s}\right)^2} \tag{3.1.21.}$$

$$\frac{(X - x)^2}{\sigma_\varepsilon^2} + \frac{(Y - y)^2}{\sigma_\eta^2} = \frac{\zeta^2}{s^2} \tag{3.1.22.}$$

from which it follows, using (3.1.5.) that

$$\frac{(X - x)^2}{\sigma_\varepsilon^2} + \frac{(Y - y)^2}{\sigma_\eta^2} = \frac{\zeta^2}{s^2} + \frac{\omega^2}{\left(\frac{\sigma_\varepsilon \sigma_\eta}{s}\right)^2} \tag{3.1.23.}$$

The likelihood function may consequently be split up into the product of two-dimensional Gaussian probability densities as follows:

$$\frac{1}{2\pi\sigma_\varepsilon\sigma_\eta} \exp\left[-\tfrac{1}{2}\left\{ \frac{(x - X)^2}{\sigma_\varepsilon^2} + \frac{(y - Y)^2}{\sigma_\eta^2} \right\}\right]$$

$$= \frac{1}{\sqrt{2\pi}.s} \exp\left[-\tfrac{1}{2}\left\{ \frac{\alpha(x - X) + \beta(y - Y)}{s}\right\}^2\right] \cdot$$

$$\cdot \frac{1}{\sqrt{2\pi}\cdot\left(\frac{\sigma_\varepsilon\sigma_\eta}{s}\right)} \exp\left[-\tfrac{1}{2}\left\{\frac{-\beta \frac{\sigma_\eta^2}{s^2}(x - X) + \alpha \frac{\sigma_\varepsilon^2}{s^2}(y - Y)}{\left(\frac{\sigma_\varepsilon\sigma_\eta}{s}\right)}\right\}^2\right] \tag{3.1.24.}$$

The two one-dimensional distributions occuring here are along the conjugate directions. Note that when (X,Y) lies on the estimated line, the first of the one-dimensional densities is independent of X and Y.

3.2. Decomposition Rule for Bayesian Structural Estimation

The likelihood function on which the theory of the previous two sections is based can be defined, in the Bayesian form, by the equation

$$p(X,Y,\pi|x,y) = L(x,y;\ X,Y,\pi)p(X,Y,\pi) \tag{3.2.1.}$$

giving the posterior density of the parameters, both incidental and structural, in terms of the prior density. The likelihood function may be written as the ratio

$$L(x,y,X,Y,\pi) = \frac{p(x,y|X,Y,\pi)}{p(x,y|\pi)} \tag{3.2.2.}$$

where

$$p(x,y|X,Y,\pi) = \frac{1}{2\pi\sigma_\varepsilon\sigma_\eta} \exp\left[-\tfrac{1}{2}\left\{\frac{(x-X)^2}{\sigma_\sigma^2} + \frac{(y-Y)^2}{\sigma_\eta^2}\right\}\right] \qquad (3.2.3.)$$

and

$$p(x,y|\pi) = 1 \qquad (3.2.4.)$$

A certain amount of difficulty arises in using these equations because of the occurence of singular and improper probability densities: the probability distribution of (X,Y) is confined to a line and the density of (x,y) before the occurence of (X,Y), is uniform over the whole plane. These difficulties can be avoided by considering only probability ratios which can be rather easily interpreted. In ratio form we write Bayes' rule as

$$\frac{p(X,Y,\pi|x,y)}{p(X,Y,\pi)} = L(X,Y;\ X,Y,\pi) = \frac{p(x,y|X,Y,\pi)}{p(x,y|\pi)} \qquad (3.2.5.)$$

We shall now show how Bayes' rule in this form may be decomposed into two similar Bayes' rules, one for the estimation of the incidental parameters and one for the estimation of the structural parameters.
The Bayes' rule for the estimation of the incidental parameters has already been given in 3.1. It may be written, in ratio form

$$\frac{p(X,Y|x,y,\pi)}{p(X,Y|\pi)} = L(\pi,x,y;\ X,Y) = \frac{p(x,y|X,Y,\pi)}{p(x,y|\pi)} \qquad (3.2.6.)$$

which, in the case when the prior probability distribution of (X,Y) along the line, can be identified with the equation

$$\frac{\exp\left[-\tfrac{1}{2}\left\{\frac{(X-\tilde{X})^2}{\sigma_\varepsilon^2} + \frac{(Y-\tilde{Y})^2}{\sigma_\eta^2}\right\}\right]}{1} = \frac{\exp\left[-\tfrac{1}{2}\left\{\frac{(x-X)^2}{\sigma_\varepsilon^2} + \frac{(y-Y)^2}{\sigma_\eta^2}\right\}\right]}{\exp\left[-\tfrac{1}{2}\left\{\frac{(x-\hat{X})^2}{\sigma_\varepsilon^2} + \frac{(y-\hat{Y})^2}{\sigma_\eta^2}\right\}\right]} \qquad (3.2.7.)$$

which comes from (3.1.5.).

To relate this result to (3.2.5.) we write

$$\frac{p(X,Y,\pi|x,y)}{p(X,Y,\pi)} = \frac{p(X,Y|x,y,\pi)}{p(X,Y|\pi)} \cdot \frac{p(\pi|x,y)}{p(\pi)} \tag{3.2.8.}$$

thus introducing an extra term corresponding to Bayesian estimation of the parameters π expressed by the equation

$$p(\pi|x,y) = L(x,y;\pi)\ p(\pi) \tag{3.2.9.}$$

The ratio form for the Bayes' rule for structural parameter estimation is

$$\frac{p(\pi|x,y)}{p(\pi)} = L(x,y;\pi) = \frac{p(x,y|\pi)}{p(x,y)} \tag{3.2.10.}$$

Taking into account the decomposition (3.1.24.) of the likelihood function we get the following:

DECOMPOSITION RULE FOR BAYESIAN STRUCTURAL ESTIMATION: the Bayesian estimation of the parameters in the structural estimation problem may be decomposed into

(a) <u>estimation of the incidental parameters given the structural parameters</u>

$$p(X,Y|x,y,\pi) = L(x,y,\pi;\ X,Y)\ p(X,Y|\pi) \tag{3.2.11.}$$

where

$$L(x,y,\pi;\ X,Y) \propto \exp -\tfrac{1}{2}\left\{\frac{\frac{\beta\sigma_\eta^2}{s^2}(X-x) - \frac{\alpha\sigma_\varepsilon^2}{s^2}(Y-y)}{\left(\frac{\sigma_\varepsilon\sigma_\eta}{s}\right)}\right\}^2 \tag{3.2.12.}$$

(b) <u>Estimation of the structural parameters</u>

$$p(\pi|x,y) = L(x,y;\pi)\ p(\pi) \tag{3.2.13.}$$

where

$$L(x,y;\pi) \propto \exp -\tfrac{1}{2}\{\frac{\alpha x + \beta y - \nu}{s}\}^2 \qquad (3.2.14)$$

Note that in (b) the variables X and Y have been illiminated

Despite the similar appearance of the two parts (a) and (b), they must be given somewhat different interpretations as we shall now see in connexion with repeated observations.

3.3. Bayesian Estimation for Repeated Observations

In the case of independent observations, Bayes' rule for combined incidental and structural parameters becomes

$$p\left((X_i, Y_i), i = 1,\ldots,n,\pi \,\middle|\, (x_i, y_i), i = 1,\ldots,n\right)$$

$$= \prod_{i=1}^{n} L(x_i, y_i; X_i, Y_i,\pi)\; p(\pi) \qquad (3.3.1.)$$

Splitting the equation by the decomposition rule we get, for the estimation of the incidental parameters given the structural parameters

$$p((X_i, Y_i), i = 1,\ldots,n|\pi,(x_i, y_i), i = 1,\ldots,n)$$

$$= \prod_{i=1}^{n} \{ L(\pi, x_i, y_i; X_i, Y_i)\; p(X_i, Y_i|\pi)\} \qquad (3.3.2.)$$

and, for the estimation of the structural parameters,

$$p(\pi|(x_i, y_i), i = 1,\ldots,n)$$

$$= \prod_{i=1}^{n} L\left((x_i, y_i),\pi\right)\; p(\pi) \qquad (3.3.3.)$$

The essential difference between these two last equations is as follows. When estimating the incidental parameters (the true values) there are just as many parameters as observations. Further, each observation provides only information about the corresponding pair of true values. Hence continued observation provides no better information about the individual values of these parameters although information about the statistical distribution of them may be obtained. On the other hand, the structural parameters do not change with each observation and it is reasonable to expect that continued observation will provide more and more precise estimates. Hence it makes sense to use the method of maximum likelihood for the structural parameters although the Bayes method must be used for the incidental parameters.

Let us first consider the maximum likelihood method for the structural parameters. The likelihood function for the n observations is

$$\prod_{i=1}^{n} L\ ((x_i,\ y_i),\pi) \propto \exp\left[-\tfrac{1}{2}\sum_{i=1}^{n}\{\frac{\alpha x_i + \beta y_i - \nu}{s}\}^2\right] \qquad (3.3.4.)$$

The ln-likelihood function is consequently

$$\ln L = \text{const} - \tfrac{1}{2}\sum_{i=1}^{n}\left(\frac{\alpha x_i + \beta y_i - \nu}{s}\right)^2$$

$$= \text{const} - \frac{n}{2}\cdot\frac{\alpha^2 s_{xx} + 2\alpha\beta s_{xy} + \beta^2 s_{yy}}{\alpha^2\sigma_\varepsilon^2 + \beta^2\sigma_\eta^2} \qquad (3.3.5.)$$

where the constant will depend in variances. Maximisation with respect to the line parameters leads to the method of generalised least squares already discussed. The determination of the variance however needs a special discussion.

As regards the incidental parameters, we find, using the expression for the likelihood function and a uniform prior distribution of the $(X_i,\ Y_i)$ along the estimated line,

$$p((X_i,\ Y_i),\ i = 1,\ldots,n|\pi,(x_i,\ y_i),i = 1,\ldots,n)$$

$$\propto \exp\left[-\tfrac{1}{2}\sum_{i=1}^{n}\left\{\frac{\frac{\beta\sigma_\eta^2}{s^2}(X_i - \tilde{X}_i) - \frac{\alpha\sigma_\varepsilon^2}{s^2}(Y_i - \tilde{Y}_i)}{(\frac{\sigma_\varepsilon\sigma_\eta}{s})}\right\}^2\right] \qquad (3.3.6.)$$

If we now put

$$\omega_i = \frac{1}{\left(\frac{\sigma_\varepsilon \sigma_\eta}{s}\right)} \left\{ \frac{\beta\sigma_\eta^2}{s^2} (X_i - \tilde{X}_i) - \frac{\alpha\sigma_\varepsilon^2}{s^2} (Y_i - \tilde{Y}_i) \right\} \tag{3.3.7.}$$

and change the variables from x_i, y_i to ζ_i, ω_i then, in the new variables where again the prior distribution is uniform,

$$p\left(\omega_i \mid \pi, (x_i, y_i)\right) \propto \exp\left[-\tfrac{1}{2} \sum_{i=1}^{n} \omega_i^2 \right] \tag{3.3.8.}$$

so that the ω_i have a spherically symmetrical Gaussian distribution with zero means and unit variances.

Now it is a property of the n-dimensional spherically symmetrical Gaussian distribution that, asymptotically, as $n \to \infty$, the distribution becomes concentrated uniformly over a hypersphere*. That this is so may be understood from the fact that

$$\frac{1}{n} \sum_{i=1}^{n} \omega^2{}_i \to E\left[\omega^2\right] = 1 \tag{3.3.9.}$$

so that asymptotically,

$$\sum_{i=1}^{n} \omega_i^2 \sim n \tag{3.3.10.}$$

meaning that $(\omega_1, \ldots, \omega_n)$ lies on a sphere of radius $\sqrt{n}$. The same property may be deduced more precisely by transforming to n-dimensional spherical coordinates and deriving that the quantity

$$R^2 = \sum_{i=1}^{n} \omega_i^2 \tag{3.3.11.}$$

has a χ^2 distribution with n degrees of freedom which, asymptotically, has a sharp peak at $R = \sqrt{n}$.

Such considerations provide the mathematical basis for the criticism of the use of the maximum likelihood estimates.

---

* P. Lévy: Leçons d'analyse fonctionelle. Paris 1922 (Gauthier-Villars).

These estimates correspond to assuming

$$\omega_i = 0 \qquad i = 1,\dots,n \tag{3.3.12.}$$

which certainly maximises $p\left(\omega_i | \sigma, (x_i, y_i)\right)$ but corresponds to a region which is extremely improbable since this is the centre of the sphere which asymptotically contains the whole of the ω-distribution. The only correct method in this situation is to abandon the use of maximum likelihood since there is no peak near to the maximum likelihood estimate which contains the greater part of the distribution. A similar situation occurs also in other contexts where the parameter or parameters to be estimated have a uniform distribution*.

* B.T. Pol'ak & Ya.Z. Tsypkin: Noise proof identification. IFAC Symp. on Identification & Syst. Parameter Estimation, Tiblisi, USSR, Sept. 1976.

4. SUMMARY AND CONCLUSIONS

As stated in the introduction, the present report is partly expository and partly original. The main expository part is the first section where an account has been given of the one-dimensional linear structural relationship problem with Gaussian errors. Attention has been given to topic points not readily available in the literature such as the convenient graphical presentation of the solution and the solution of Dent for the case of unknown variances. The second section, which contains a somewhat new presentation of known material, shows how the maximum likelihood solution gives rise to the generalised least squares principle.
The third section, which is original, has re-analysed the problem from a Bayesian viewpoint and shown how such an analysis leads to the introduction of a modified likelihood function for the estimation of the structural parameters. The use of this likelihood function immediately leads to the principle of generalised least squares.

In future work it is intended to show how a similar approach may be used for structural relation in the multidimensional case and in linear systems analysis.

5. APPENDIX

Monotonicity of the Estimate of the Slope

We here show that the slope of the estimated line is a monotonic function of the variance ratio k unless the observations are either uncorrelated or perfectly correlated.

In non-homogenous line coordinates, the slope is

$$a = \frac{s_{yy} - h^2 s_{xx} + \sqrt{(s_{yy} - k^2 s_{xx})^2 - 4\, k^2 s_{xy}^2}}{2 s_{xy}}$$

Differentiation gives

$$\frac{1}{2k}\,\frac{\partial a}{\partial k} = \frac{1}{2 s_{xy}} \{- s_{xx} + \frac{s_{yy} - k^2 s_{xx})(-s_{xx}) + 2 s_{xy}^2}{\sqrt{(s_{yy} - k^2 s_{xx})^2 + 4k^2 s_{xy}^2}}\}$$

so that if the right hand side is zero then

$$s_{xx}^2 \{(s_{yy} - k^2 s_{xx})^2 + 4k^2 s_{xy}^2\} = (s_{yy} - k^2 s_{xx})^2 s_{xx}^2 +$$

$$- 4 s_{xx} s_{xy}^2 (s_{yy} - k^2 s_{xx}) + 4 s_{xy}^4$$

from which it follows that

$$0 = 4 s_{xy}^2 (- s_{xx} s_{yy} + s_{xy}^2)$$

This equation implies that either the observations are uncorrelated or are perfectly correlated.

It is easy to verify that $k = 0$ and $k = \infty$ correspond to the two regression lines. Consequently, as k increases from 0 to ∞ the estimated line moves monotonically from one regression line to the other.

6. REFERENCES ON STRUCTURAL ESTIMATION

The following list gives the principal references on structural estimation in both the statistical and system-analysis literature. For further references, see Mandansky (1959), Moran (1971). The list is arranged in chronological order.

1. ADCOCK, A.D. (1877)
   The Analyst 1877, 4, 183-, 5, 53-

2. KUMMEL, C.H.
   The Analyst 1897, 6, 97-

3. PEARSON, K; On lines and planes of closest fit to systems of points in space. Phil.Mag. 1901, 2, 559-572

4. RHODES, E.C.; On lines and planes of closest fit.
   Phil. Mag. 1927, 7(3), 357-

5. UVEN, M.J. van; Adjustment of N points....
   Proc. Kon. Akad. Wetensch. Amst. 1930, 33, 143-, 307-

6. DENT, B.M.; On observation of points connected by a linear relation.
   Proc. Phys. Soc. 1935, 47, 92-108

7. ROOS, C.F.
   Metron. 1937, 13, (1), 3-

8. KOOPMANS, T.; Linear Regression Analysis of Economic Time Series.
   Haarlem 1937 (DeErven-Bohn)

9. WALD, A.; The fitting of straight lines if both variables are subject to error. Annals Math. Stat. 1940, 4, 284-300

10. LINDLEY, D.V.; Regression lines and the linear functional relationship.
    J. Roy. Stat. Soc. 1947, supplement, 9, 218-

11. NEYMAN & SCOTT, E.; Consistent estimates based on partially consistent observations.
Econometrica 1948, 16, 1-32

12. KOOPMANS, T.S. & REIERSØL, O.; The identification of structural characteristics. Ann. Math. Stat. 1950, 21, 165-181.

13. BERKSON, J.; Are there two regressions?
J. Amer. Stat. Ass. 1950, 45, 164-180

14. NEYMAN, J. & SCOTT, E.; On certain methods of estimating the linear structural relationship.
Ann. Math. Stat., 22, 252-361 (correction in 23, 135-)

15. NEYMAN, J.; Existence of consistent estimates of the directional parameter in a linear structural relation between two variables.
Ann. Math. Stat. 1951, 22, 497-512

16. KENDALL, M.G.; Regression, structure and functional relation.
Biometrika I, 1951, 38, 11-25; II, 1952, 39, 96-108

17. WOLFOWITZ, J.; Consistent estimators of the parameters of a linear structural relation.
Skand. Aktuarietids. 1952, 132-151

18. LINDLEY, D.V.; Estimation of a functional relationship.
Biometrika 1953, 40, 47-49

19. KIEFER, J. & WOLFOWITZ, J.; Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters.
Annals Math. Stat. 1956, 27, 887-906

20. BROWN, R.L.; The Bivariate structural relation.
Biometrika 1957, 44, 84-96

21. KENDALL, M.G. & STEWART, A.; The Advanced Theory of Statistics. London 1958 (Griffin) part 1, 2, 3

22. MADANSKY, A.; The fitting of straight lines when both variables are subject to error.
J. Amer. Stat. Ass. 1959, 54, 173-205.

23. GRAYBILL, F.A.; An Introduction to Linear Statistical Models. New York 1961, (McGraw-Hill) Vol. 1

24. VILLEGAS, C.; Maximum likelihood estimation of a linear functional relation.
Ann. Math. Stat. 1961, 32, 1048-1062

25. LEVIN, J.; Estimation of a system pulse transfer function in the presence of noise.
IEEE Trans. Aut. Control 1964, AC-9, 229-

26. SPRENT, P.; A generalised least squares approach to linear functional relationships.
J. Roy. Stat. Soc. 1966, 28B, 272-297

27. KENDALL, M.G. & STEWART, A.; The Advanced Theory of Statistics. 2nd ed. London 1967 (Griffin) Vol. 2

28. LINDLEY, D.V. & EL-SAYYAD; The Bayesian estimation of a linear functional relationship.
J. Roy. Stat. Soc. 1968, 30B, 190-202

29. SMITH, F.W.; System Laplace-transform estimation from sampled data. IEEE Trans. Aut. Control, 1968, AC-13, no. 1, 37-44

30. SOLARI, M.E.; The "maximum likelihood solution" of the problem of estimating a linear functional relation.
J. Roy. Stat. Soc. 31B, 1969, 372-375

31. ROGERS, A.E. & STEIGLITZ, K.; On system identification from noise-obscured input and output measurements.
Int. J. Control 1970, 12, no. 4, 625-635

32. BARNETT, V.D.; Fitting straight lines - the linear functional relationship problem.
J. Roy. Stat. Soc. 1970, 32B, 274-278

33. SPRENT, P.; The saddle point of the likelihood surface for a linear functional relationship.
J. Roy. Stat. Soc. 1970, 32B, 432-434

34. MORAN, D.P.P.; Estimating structural and functional relationships.
J. Mult. Anal., 1, 232-255

35. DOLBY, G.R.; Generalised least squares and maximum likelihood estimation of non-linear functional relationships.
J. Roy. Stat. Soc. 1972, 34B, 393-400

36. TAYLOR, J.; A method of fitting several linear functional relations and of testing for differences between them.
Appl. Statistics 1973, 22, 239-242

37. FLORENS, J.P., MOUCHART, M. & RICHARD, J.P.; Bayesian inference in error-in-variables models.
J. Mult. Anal. 1974, 4, 419-452

38. DOLBY, G.R. & FREEMAN, T.G.; Functional relationships having many independent variables and errors with multivariate normal distributions.
J. Mult. Anal. 1975, 5, 466-479

39. AKASHI, H. & MOUSTAFA, A.F.; Parameter identification of systems with noise in input and output.
Kyoto Univ. 1975, Mem. Fac. Eng. 184-201