COWLES COMMISSION DISCUSSION PAPER: STATISTICS NO. 364

SUPPLY AND DEMAND STUDIES, USDA, RMA, RESEARCH PAPER NO. 2

Structural Estimates From Means of Subsets of Observations

Clifford Hildreth[1]

August 6, 1951

I

A method has been suggested by Wald [1] for fitting a linear relation between two variables when observations of both variables are subject to error.[2] It seems to me that procedures similar to Wald's may be useful in estimating structural parameters in some of the systems of equations encountered in econometric work. While a shock model is considered in most detail in the paper, the procedures may also be useful for certain types of error models and combined shock-error models.

Unfortunately it is difficult to say much about the general properties of estimators obtained by these procedures. A number of intuitive possibilities suggest themselves and these can be checked in special cases. Examples can also be investigated in some detail and the results used as a guide to intuition. Some investigation of special cases is presented in this paper; it is hoped that somewhat more general results can be achieved later.

To introduce the procedures in question we consider the following situation. An investigator wishes to estimate some or all of the unknown parameters of a system of equations given by —

$$(1.1) \qquad \varphi_i(y_t, z_t, u_t) = 0 \qquad i=1,2, \ldots, G$$
$$t=1,2, \ldots, T$$

The subscript i denotes successive equations of the system, the subscript t denotes successive observations in the investigator's sample. y_t is a vector of the values taken by the current endogenous variables in the t-th observation. If the system is complete, y_t will have G components. z_t is a vector of the values of the predetermined variables. u_t is a vector of unobserved random disturbances assumed to have a stable multivariate distribution that is independent of the predetermined variables. The disturbances are further assumed to have zero means and finite variances.

Consider the problem of estimating parameters of one equation, for convenience the first, assuming that this equation is linear in the observed variables — y_t, z_t,[3] and that the first element of u_t enters additively. The first equation could then be written —

$$(1.2) \qquad y_{1t} + \beta_{12} y_{2t} + \ldots + \beta_{1g} y_{gt} + \gamma_{11} z_{1t} + \gamma_{12} z_{2t} + \ldots + \gamma_{1k} z_{kt} + \gamma_{10} = u_{1t} .$$

y_{it}, $i=1,2, \ldots, g$, is an element of y_t not known a priori to have a zero coefficient in the first equation. z_{it}, $i=1,2, \ldots, k$, is an element of z_t not known to have a coefficient of zero in the first equation. If y_t has G elements and z_t has K elements, this means that G-g of the endogenous variables and K-k of the predetermined variables have zero coefficients in the first equation. u_{1t} is the first element of u_t.

β_{ij}, γ_{ij} are unknown coefficients assumed constant over the observations.

The coefficients in (1.2) have been normalized by setting the coefficient of y_{1t} equal to one. It is assumed that y_{1t} is known to have a non-zero coefficient.

The equation obtained by dropping the t subscripts and setting u_1 equal to zero will be called the equation of the true structural plane. It is given by —

$$(1.3) \qquad y_1 + \beta_{12} y_2 + \dots + \beta_{1g} y_g + \gamma_{11} z_1 + \gamma_{12} z_2 + \dots + \gamma_{1k} z_k + \gamma_{10} = 0$$

If estimates, say $\tilde{\beta}_{ij}$, $\tilde{\gamma}_{ij}$, of the coefficients are substituted for the true values we have an estimated structural plane —

$$(1.4) \qquad y_1 + \tilde{\beta}_{12} y_2 + \dots + \tilde{\beta}_{1g} y_g + \tilde{\gamma}_{11} z_1 + \tilde{\gamma}_{12} z_2 + \dots + \tilde{\gamma}_{1k} z_k + \tilde{\gamma}_{10} = 0.$$

The procedure to be considered for estimating the g+k unknown parameters is as follows. The observations, y_t, z_t, are divided into g+k subsets according to the magnitudes of some or all of the elements of z_t. The mean of each observed variable appearing in (1.2) is computed for each subset of the observations. Let $\bar{y}_{ij}$ be the mean of y_i in the j-th subset and $\bar{z}_{ij}$ be the mean of z_i in the j-th subset. The estimates are then obtained from the following set of equations —

$$(1.5) \qquad \bar{y}_{1j} + \tilde{\beta}_{12} \bar{y}_{2j} + \dots + \tilde{\beta}_{1g} \bar{y}_{gj} + \tilde{\gamma}_{11} \bar{z}_{1j} + \tilde{\gamma}_{12} \bar{z}_{2j} + \dots + \tilde{\gamma}_{1k} \bar{z}_{kj}$$
$$+ \tilde{\gamma}_{10} = 0 \qquad j = 1, 2, \dots, (g+k)$$

This determines an estimated structural plane in the variable space (a Euclidean space of g+k dimensions with axes representing values of $y_1, y_2, \dots, y_g, z_1, z_2, \dots, z_k$). The estimated plane passes through the (g+k) points given by the subset means. It is unique if the points do not lie on a plane of less than (g+k-1) dimensions. Equivalently, unique estimates of coefficients are obtained if the equations of (1.5) are linearly independent.

Thus far, the procedure has not been completely specified. Exactly how the observations are to be partitioned has not been stated. I doubt that a best procedure can be specified for models as general as those considered above. I would expect that, in a particular application, in a the best way to partition the observations would depend upon the investigator's a priori knowledge, i.e., on the special properties of his particular model.4/

There is, however, the following general observation on consistency of the estimates. Since the partition is independent of the values taken by the disturbance, u_{1t}, in the various observations, the values of the disturbance in each subset represent independent drawings from a stable population. As the number of observations in a given subset increases, the mean value of the disturbance in that subset approaches zero as a probability limit and the point in the variable space given by the subset means of observed variables approaches the true structural plane in a probability sense. Let g+k=h and —

$$(1.6)\quad M=\begin{pmatrix} \bar{y}_{21} & \bar{y}_{31} & \cdots & \bar{y}_{g1} & \bar{z}_{11} & \bar{z}_{21} & \cdots & \bar{z}_{k1} & 1 \\ \bar{y}_{22} & \bar{y}_{32} & \cdots & \bar{y}_{g2} & \bar{z}_{12} & \bar{z}_{22} & \cdots & \bar{z}_{k2} & 1 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \bar{y}_{2h} & \bar{y}_{3h} & & \bar{y}_{gh} & \bar{z}_{1h} & \bar{z}_{2h} & & \bar{z}_{kh} & 1 \end{pmatrix}$$

$$(1.7) \quad \alpha' = \begin{pmatrix} \beta_{12} \\ \beta_{13} \\ \vdots \\ \beta_{1g} \\ \gamma_{11} \\ \gamma_{12} \\ \vdots \\ \gamma_{1k} \\ \gamma_{10} \end{pmatrix} \qquad \tilde{\alpha}' = \begin{pmatrix} \tilde{\beta}_{12} \\ \tilde{\beta}_{13} \\ \vdots \\ \tilde{\beta}_{1g} \\ \tilde{\gamma}_{11} \\ \tilde{\gamma}_{12} \\ \vdots \\ \tilde{\gamma}_{1k} \\ \tilde{\gamma}_{10} \end{pmatrix}$$

$$(1.8) \quad \bar{y}'_{1.} = \begin{pmatrix} \bar{y}_{11} \\ \bar{y}_{12} \\ \vdots \\ \bar{y}_{1h} \end{pmatrix}$$

The equations of (1.5) can then be written --

$$(1.9) \qquad M \tilde{\alpha}' = \bar{y}'_{1.}$$

Assume that as the total number of observations approaches infinity, subsets are chosen in such a way that the number of observations in each subset also tends toward infinity. Assume, also, that as the number of observations grows large, each $\bar{z}_{ij}$ approaches a definite limit (if z_i is a fixed variable) or probability limit (if z_i is a random variable). Let plim M be the matrix obtained from M by substituting plim $\bar{y}_{ij}$ (probability limit of $\bar{y}_{ij}$) for $\bar{y}_{ij}$ ($i=2,3, \ldots,g$; $j=1,2, \ldots h$) and lim $\bar{z}_{ij}$ or plim $\bar{z}_{ij}$

for $\bar{z}_{ij}$ $(i=1,2, \ldots,k; j=1,2, \ldots,h)$. If plim M is nonsingular the estimates of coefficients obtained from (1.9) are consistent.

I conjecture that even if the $\bar{z}_{ij}$ do not approach limits, one could sometimes devise methods of partitioning the observations so that M would be nonsingular for any size sample and the estimates would still be consistent. I have not yet attempted a proof.

II

As a special case, consider the following two equation model —

$$(2.1) \qquad y_{1t} + \beta_{12} y_{2t} + \gamma_{11} z_{1t} + \gamma_{10} = u_{1t}$$

$$(2.2) \qquad \beta_{21} y_{1t} + y_{2t} + \gamma_{22} z_{2t} + \gamma_{20} = u_{2t}$$

An estimator of β_{12} obtained by the procedures of Section I will be examined and compared with the maximum likelihood estimator under the assumption that u_{1t}, u_{2t} are drawn from a bivariate normal. For this comparison the z's will be treated as fixed variates. $\tilde{\beta}_{12}$, the estimator obtained from means of subsets, will be obtained by solving —

$$(2.3) \qquad \begin{aligned} \bar{y}_{11} + \tilde{\beta}_{12} \bar{y}_{21} + \tilde{\gamma}_{11} \bar{z}_{11} + \tilde{\gamma}_{10} &= 0 \\ \bar{y}_{12} + \tilde{\beta}_{12} \bar{y}_{22} + \tilde{\gamma}_{11} \bar{z}_{12} + \tilde{\gamma}_{10} &= 0 \\ \bar{y}_{13} + \tilde{\beta}_{12} \bar{y}_{23} + \tilde{\gamma}_{11} \bar{z}_{13} + \tilde{\gamma}_{10} &= 0. \end{aligned}$$

Let $Y_{ij} = \bar{y}_{ij} - \bar{y}_i$, the difference between the mean of y_i in the j-th subset and the mean of y_i in the complete sample, and define Z_{ij} analogously.

A set of equations equivalent to (2.3) is then given by --

$$(2.4)\qquad Y_{11} + \tilde{\beta}_{12} Y_{21} + \tilde{\gamma}_{11} Z_{11} = 0$$

$$Y_{12} + \tilde{\beta}_{12} Y_{22} + \tilde{\gamma}_{11} Z_{12} = 0$$

$$(2.5)\qquad \bar{y}_1 + \tilde{\beta}_{12} \bar{y}_2 + \tilde{\gamma}_{11} \bar{z}_{11} + \tilde{\gamma}_{10} = 0$$

Since we will not in this paper be interested in the estimate of $\tilde{\gamma}_{10}$, (2.5) may be ignored. From (2.4) we have --

$$(2.6)\qquad \tilde{\beta}_{12} = - \frac{Y_{11} Z_{12} - Y_{12} Z_{11}}{Y_{21} Z_{12} - Y_{22} Z_{11}}.$$

To investigate the distribution of $\tilde{\beta}_{12}$, it will be convenient to use the reduced form of (2.1), (2.2) --

$$(2.7)\qquad y_{1t} = \pi_{11} z_{1t} + \pi_{12} z_{2t} + \pi_{10} + v_{1t}$$

$$(2.8)\qquad y_{2t} = \pi_{21} z_{1t} + \pi_{22} z_{2t} + \pi_{20} + v_{2t}$$

where the π's are ordinary regression coefficients and the v's are normally distributed disturbances with zero means and finite variances. The following relations hold between coefficients and disturbances of the reduced form and those of the structural equations.

$$(2.9)\qquad \pi_{11} = \frac{-\gamma_{11}}{1-\beta_{12}\beta_{21}} \qquad \pi_{21} = \frac{\beta_{21}\gamma_{11}}{1-\beta_{12}\beta_{21}}$$

$$\pi_{12} = \frac{\beta_{12}\gamma_{22}}{1-\beta_{12}\beta_{21}} \qquad \pi_{22} = \frac{-\gamma_{22}}{1-\beta_{12}\beta_{21}}$$

$$\pi_{10} = \frac{-\gamma_{10}+\beta_{12}\gamma_{20}}{1-\beta_{12}\beta_{21}} \qquad \pi_{20} = \frac{\beta_{21}\gamma_{10}-\gamma_{20}}{1-\beta_{12}\beta_{21}}$$

$$v_1 = \frac{u_1-\beta_{12} u_2}{1-\beta_{12}\beta_{21}} \qquad v_2 = \frac{-\beta_{21} u_1 + u_2}{1-\beta_{12}\beta_{21}}$$

From (2.7), (2.8) we have --

(2.10) $$Y_{ij} = \pi_{i1} Z_{1j} + \pi_{i2} Z_{2j} + V_{ij} \qquad i=1,2;\ j=1,2,3$$

where

(2.11) $$V_{ij} = \bar{v}_{ij} - \bar{v}_i$$

From (2.6) and (2.1) --

(2.12) $$\tilde{\beta}_{12} = -\frac{\pi_{12}(Z_{21} Z_{12} - Z_{22} Z_{11}) + (V_{11} Z_{12} - V_{12} Z_{11})}{\pi_{22}(Z_{21} Z_{12} - Z_{22} Z_{11}) + (V_{21} Z_{12} - V_{22} Z_{11})}$$

Let

(2.13) $$c = (Z_{21} Z_{12} - Z_{22} Z_{11})$$

$$w_1 = (V_{11} Z_{12} - V_{12} Z_{11})$$

$$w_2 = (V_{21} Z_{12} - V_{22} Z_{11})$$

$$w_3 = -(w_1 + \beta_{12} w_2)$$

and note from (2.9) that $\pi_{12} = -\beta_{12}\ \pi_{22}$. (2.12) can then be put in the form --

(2.14) $$\tilde{\beta}_{12} = \frac{\beta_{12}\pi_{22}\, c - w_1}{\pi_{22}\, c + w_2}$$

$$= \beta_{12} - \frac{w_1 + \beta_{12} w_2}{\pi_{22}\, c + w_2} \qquad = \beta_{12} + \frac{w_3}{\pi_{22}\, c + w_2}$$

Since the predetermined variables are being treated as fixed variates and since w_1, w_2, w_3 are linear combinations of normal variables, $\tilde{\beta}_{12}$ is a ratio of two normal variables. It is known[5] that the distribution of such a ratio has no moments. In this case the maximum likelihood estimator is also a ratio of two normally distributed variables. Thus, for finite samples, the two estimators cannot be compared on the basis of moments. For an assumed pattern of observations on the z's and assumed structural parameters, the deciles of the distributions of the two estimators are compared in Section V below. It is intended that a comparison of the asymptotic efficiencies of the estimators will be attempted later.

III

Let $\hat{\beta}_{12}$ be the maximum likelihood estimator of β_{12} in the structural system given by (2.1), (2.2). It is known that —

$$(3.1) \qquad \hat{\beta}_{12} = -\frac{\hat{\pi}_{12}}{\hat{\pi}_{22}} = -\frac{m_{y_1z_2}m_{z_1z_1} - m_{y_1z_1}m_{z_1z_2}}{m_{y_2z_2}m_{z_1z_1} - m_{y_2z_1}m_{z_1z_2}}$$

where

$$(3.2) \qquad m_{y_iz_j} = \frac{1}{T}\sum_{t=1}^{T}(y_{it}-\bar{y}_i)(z_{jt}-\bar{z}_j)$$

$$m_{z_iz_j} = \frac{1}{T}\sum_{t=1}^{T}(z_{it}-\bar{z}_i)(z_{jt}-\bar{z}_j)$$

and $\hat{\pi}_{12}$, $\hat{\pi}_{22}$ are the least-squares estimates (in this case they are also maximum likelihood estimates) of $\hat{\pi}_{12}$, $\hat{\pi}_{22}$.

From reduced form equations (2.7), (2.8) we have —

$$(3.3) \qquad m_{y_iz_j} = \pi_{i1}\, m_{z_1z_j} + \pi_{i2}\, m_{z_2z_j} + m_{v_iz_j}$$

and

$$(3.4)\quad \hat{\beta}_{12} = -\frac{\pi_{12}(m_{z_1z_1}\, m_{z_2z_2} - m_{z_1z_2}^2) + (m_{v_1z_2}\, m_{z_1z_1} - m_{v_1z_1}\, m_{z_1z_2})}{\pi_{22}(m_{z_1z_1}\, m_{z_2z_2} - m_{z_1z_2}^2) + (m_{v_2z_2}\, m_{z_1z_1} - m_{v_2z_1}\, m_{z_1z_2})}$$

Let

$$(3.5)\quad d = (m_{z_1z_1}\, m_{z_2z_2} - m_{z_1z_2}^2)$$

$$r_1 = (m_{v_1z_2}\, m_{z_1z_1} - m_{v_1z_1}\, m_{z_1z_2})$$

$$r_2 = (m_{v_2z_2}\, m_{z_1z_1} - m_{v_2z_1}\, m_{z_1z_2})$$

$$r_3 = -(r_1 + \beta_{12}\, r_2)$$

(3.4) can then be rewritten —

$$(3.6)\quad \hat{\beta}_{12} = \frac{\beta_{12}\pi_{22}\, d - r_1}{\pi_{22}\, d + r_2} = \beta_{12} - \frac{r_1 + \beta_{12}\, r_2}{\pi_{22}\, d + r_2} = \beta_{12} + \frac{r_3}{\pi_{22}\, d + r_2}$$

The r's are normally distributed random variables and the similarity in form between $\hat{\beta}_{12}$ and $\tilde{\beta}_{12}$ is apparent [compare (3.6) and (2.14)].

IV

To compare the estimators further, it is useful to obtain the parameters of the distributions of the w's and r's in terms of the parameters of the reduced form.

From (2.11) —

$$(4.1)\quad v_{11} = \bar{v}_{11} - \bar{v}_1 = \frac{2\,\bar{v}_{11} - \bar{v}_{12} - \bar{v}_{13}}{3}$$

$$v_{12} = \bar{v}_{12} - \bar{v}_1 = \frac{2\,\bar{v}_{12} - \bar{v}_{11} - \bar{v}_{13}}{3}$$

$\bar{v}_{1i}$ has the normal distribution with zero mean and variance $\frac{3\sigma^2_{v_1}}{T}$. Since $\bar{v}_{1i}$ is independent of $\bar{v}_{1j}$ for $i \neq j$, we have —

(4.2) $$\sigma^2_{V_{11}} = \frac{2\sigma^2_{v_1}}{T}$$

$$\sigma^2_{V_{12}} = \frac{2\sigma^2_{v_1}}{T}$$

$$\sigma_{V_{11}V_{12}} = -\frac{\sigma^2_{v_1}}{T}$$

(4.3) $$\sigma^2_{w_1} = Z^2_{12}\,\sigma^2_{V_{11}} - 2Z_{12}\,Z_{11}\,\sigma_{V_{11}V_{12}} + Z^2_{11}\,\sigma^2_{V_{12}}$$

$$= \frac{2}{T}\,(Z^2_{12} + Z_{12}\,Z_{11} + Z^2_{11})\,\sigma^2_{v_1}$$

$$= \frac{2a}{T}\,\sigma^2_{v_1}$$

where

(4.4) $$a = (Z^2_{12} + Z_{12}\,Z_{11} + Z^2_{11}).$$

The following can be similarly obtained —

(4.5) $$\sigma^2_{w_2} = \frac{2a}{T}\,\sigma^2_{v_2}$$

(4.6) $$\sigma_{w_1w_2} = \frac{2a}{T}\,\sigma_{v_1v_2}$$

(4.7) $$\sigma^2_{w_3} = \frac{2a}{T}\,(\sigma^2_{v_1} + 2\beta_{12}\,\sigma_{v_1v_2} + \beta^2\,\sigma^2_{v_2}) = \frac{2a}{T}\,\sigma^2_{v_3}$$

where

(4.8) $$v_3 = -(v_1 + \beta_{12}\,v_2)$$

(4.9) $$\sigma_{w_2w_3} = \frac{2a}{T}\,\sigma_{v_2v_3}.$$

To find the parameters of the distributions of r_1, r_2, r_3 we need to know the distribution of moments, $m_{v_i z_j}$ $i,j=1,2$, involving a disturbance and a predetermined variable. v_{it} is known to be normal, independent of z_{js}, and independent of $v_{jr}(i,j=1,2;\ s=1,2,\ \ldots,\ T;\ r\neq t)$. Since, in addition, the z's are being treated as fixed variates we know that $m_{v_i z_j}$ is normally distributed and that —

(4.10) $$E(m_{v_i z_j}) = 0$$

(4.11) $$\mathrm{Var}\ (m_{v_i z_j}) = \frac{m_{z_j z_j}}{T}\,\sigma^2_{v_i}$$

(4.12) $$\mathrm{Cov}\ (m_{v_i z_j},\ m_{v_i z_i}) = \frac{m_{z_i z_j}}{T}\,\sigma^2_{v_i}$$

(4.13) $$\mathrm{Cov}\ (m_{v_i z_j},\ m_{v_j z_j}) = \frac{m_{z_j z_j}}{T}\,\sigma_{v_i v_j}$$

Using the above and the definitions given in (3.5) we have —

(4.14) $$\sigma^2_{r_1} = \frac{m_{z_1 z_1}^{d}}{T}\,\sigma^2_{v_1}$$

(4.15) $$\sigma^2_{r_2} = \frac{m_{z_1 z_1}^{d}}{T}\,\sigma^2_{v_2}$$

(4.16) $$\sigma^2_{r_3} = \frac{m_{z_1 z_1}^{d}}{T}\,(\sigma^2_{v_1} + 2\beta_{12}\,\sigma_{v_1 v_2} + \beta_{12}^2\ \sigma^2_{v_2}) = \frac{m_{z_1 z_1}^{d}}{T}\,\sigma^2_{v_3}$$

(4.17) $$\sigma_{r_1 r_2} = \frac{m_{z_1 z_1}^{d}}{T}\ \sigma_{v_1 v_2}$$

(4.18) $$\sigma_{r_2 r_3} = \frac{m_{z_1 z_1}^{d}}{T}\ \sigma_{v_2 v_3}$$

Let

$$(4.19) \qquad \tilde{\mathcal{E}} = \tilde{\beta}_{12} - \beta_{12} = \frac{w_3}{\pi_{22}\, c + w_2}$$

$$(4.20) \qquad \hat{\mathcal{E}} = \hat{\beta}_{12} - \beta_{12} = \frac{r_3}{\pi_{22}\, d + r_2}$$

The cumulative distribution functions of $\tilde{\mathcal{E}}$ and $\hat{\mathcal{E}}$ can be written as integral equations. For $\tilde{\mathcal{E}}$ we have —

$$(4.21) \qquad P(\tilde{\mathcal{E}} < \mathcal{E}) = F(\mathcal{E}) = P[w_3 < \mathcal{E}(\pi_{22}\, c + w_2),\ w_2 > -\pi_{22}\, c]$$

$$+ P[w_3 > \mathcal{E}(\pi_{22}\, c + w_2),\ w_2 < -\pi_{22}\, c]$$

Let

$$(4.22) \qquad x_1 = \frac{w_3 - \mathcal{E} w_2}{\sqrt{\sigma_{w_3}^2 - 2\mathcal{E}\sigma_{w_2 w_3} + \mathcal{E}^2 \sigma_{w_2}^2}} = \frac{w_4}{\sigma_{w_4}}$$

$$x_2 = \frac{w_2}{\sigma_{w_2}}$$

where

$$(4.23) \qquad w_4 = w_3 - \mathcal{E} w_2 = -[w_1 + (\beta_{12} + \mathcal{E}) w_2]$$

$$v_4 = v_3 - \mathcal{E} v_2 = -[v_1 + (\beta_{12} + \mathcal{E}) v_2]$$

$$\sigma_{w_4}^2 = \sigma_{w_3}^2 - 2\mathcal{E}\sigma_{w_2 w_3} + \mathcal{E}^2 \sigma_{w_2}^2 = \frac{2a}{T}\sigma_{v_4}^2$$

x_1, x_2 have a bivariate normal with zero means, unit variances and —

$$(4.24) \qquad \rho_{x_1 x_2} = \sigma_{x_1 x_2} = \frac{\sigma_{w_2 w_3} - \mathcal{E}\sigma_{w_2}}{\sigma_{w_2}\, \sigma_{w_4}} = \frac{\sigma_{v_2 v_4}}{\sigma_{v_2}\, \sigma_{v_4}}$$

We now have —

$$(4.25) \qquad F(\mathcal{E}) = P\left[x_1 < \frac{\mathcal{E}\pi_{22}c}{\sigma_{w_4}},\ x_2 > -\frac{\pi_{22}c}{\sigma_{w_2}}\right] + P\left[x_1 > \frac{\mathcal{E}\pi_{22}c}{\sigma_{w_4}},\ x_2 < -\frac{\pi_{22}c}{\sigma_{w_2}}\right]$$

$$= \int_{-\frac{\pi_{22}c}{\sigma_{w_2}}}^{\infty} \int_{-\infty}^{\frac{\mathcal{E}\pi_{22}c}{\sigma_{w_4}}} h(x_1, x_2)\,dx_1\,dx_2 + \int_{-\infty}^{-\frac{\pi_{22}c}{\sigma_{w_2}}} \int_{\frac{\mathcal{E}\pi_{22}c}{\sigma_{w_4}}}^{\infty} h(x_1, x_2)\,dx_1\,dx_2$$

where $h(x_1, x_2)$ is the density function of x_1, x_2. If π_{22}, c, σ_{w_2}, σ_{w_4} are known, $F(\mathcal{E})$ can be computed for various values of $\mathcal{E}$ from Pearson's <u>Tables for Statisticians and Biometricians, Part II</u>. This is done for a hypothetical example in the next section.

As yet, little has been said about the basis for partitioning the observations. Let

$$(4.26) \qquad \tilde{\ell}_1 = \frac{\mathcal{E}\pi_{22}c}{\sigma_{w_4}} = \frac{\sqrt{T}\,\mathcal{E}\pi_{22}c}{\sqrt{2a}\,\sigma_{v_4}}$$

$$\tilde{\ell}_2 = \frac{\pi_{22}c}{\sigma_{w_2}} = -\frac{\sqrt{T}\,\pi_{22}c}{\sqrt{2a}\,\sigma_{v_2}}$$

In the expressions on the right, only c and a depend on the method of partitioning. Typically we may expect that increases in the absolute values of $\tilde{\ell}_1$ and $\tilde{\ell}_2$ will decrease $F(\mathcal{E})$ for $\mathcal{E} < 0$ and increase $F(\mathcal{E})$ for $\mathcal{E} > 0$, thus concentrating the probability associated with $\tilde{\beta}_{12}$ more closely abo[ut] the true value.[6] This suggests that a method of partitioning that maximizes $\left|\frac{c}{\sqrt{a}}\right|$ may be close to optimal in a wide variety of circumstances.

For $\hat{\mathcal{E}}$ we have —

$$(4.27) \qquad G(\mathcal{E}) = P(\hat{\mathcal{E}} < \mathcal{E})$$

$$= P[r_3 < \mathcal{E}(\pi_{22}\, d + r_2),\ r_2 > -\pi_{22}\, d]$$

$$+ P[r_3 > \mathcal{E}(\pi_{22} d + r_2),\ r_2 < -\pi_{22}\, d]$$

where $G(\mathcal{E})$ gives the cumulative distribution of $\hat{\mathcal{E}}$. By a transformation analogous to (4.22) —

$$(4.28) \qquad G(\mathcal{E}) = \int_{-\frac{\pi_{22} d}{\sigma_{r_2}}}^{\infty} \int_{-\infty}^{\frac{\mathcal{E}\pi_{22}\, d}{\sigma_{r_4}}} h(x_1, x_2)\, dx_1, dx_2$$

$$+ \int_{-\infty}^{-\frac{\pi_{22}\, d}{\sigma_{r_2}}} \int_{\frac{\mathcal{E}\pi_{22}\, d}{\sigma_{r_4}}}^{\infty} h(x_1, x_2)\, dx_1 dx_2$$

where

$$(4.29) \qquad r_4 = r_3 - \mathcal{E} r_2 = -[r_1 + (\beta_{12} + \mathcal{E}) r_2]$$

$$\sigma^2_{r_4} = \sigma^2_{r_3} - 2\,\mathcal{E}\sigma_{r_2 r_3} + \mathcal{E}^2\ \sigma^2_{r_2} = \frac{m_{z_1 z_1}\, d}{T}\ \sigma^2_{v_4}$$

For given $\mathcal{E}$, (4.28) differs from (4.25) only in the limits over which the integrals are taken.

Let

$$(4.30) \qquad {}_1 = \frac{\mathcal{E}\pi_{22}\, d}{\sigma_{r_4}} = \frac{\sqrt{T}\,\mathcal{E}\,\pi_{22}\sqrt{d}}{\sqrt{m_{z_1 z_1}}\,\sigma_v}$$

$$\hat{\ell}_2 = -\frac{\pi_{22}\, d}{\sigma_{r_2}} = -\frac{\sqrt{T}\,\pi_{22}\sqrt{d}}{\sqrt{m_{z_1 z_1}}\,\sigma_{v_2}}$$

Then, from (4.5), (4.15), (4.23), (4.30)

$$(4.31) \qquad \ell = \frac{\hat{\ell}_1}{\ell_1} = \frac{\hat{\ell}_2}{\ell_2} = \frac{\sqrt{2a\, d}}{c\sqrt{m_{z_1 z_1}}} \quad .$$

In Appendix I it is shown that —

$$(4.32) \qquad |\ell| \geqq 1.$$

From this we can expect the distribution of $\hat{\beta}_{12}$ to ordinarily exhibit at least as great a concentration of probability around the true parameter value as $\tilde{\beta}_{12}$. Aside from simplicity of computation, $\hat{\beta}_{12}$ is ordinarily an equally good or better estimator for the type of model considered.

V

It would not be surprising to find that the maximum likelihood estimator has generally better properties than the proposed alternative for many other types of models. Part of the motivation for studying the properties of an alternative estimator arises from the fact that in many of the models that have been developed to represent practical economic situations it is difficult or impossible to obtain maximum likelihood estimates either because of the cost and complexity of the calculations or because the investigator can not specify his model narrowly enough to determine the likelihood function. Alternative methods for some types of models have been suggested. These are usually called limited information maximum likelihood or quasi maximum likelihood and have been discussed by Anderson and Rubin [12], [13] and by Anderson [14]. Ultimately, it will be interesting to compare estimates based on means of subsets with these other alternatives in some of the situations where full maximum likelihood estimates are not practical. These situations usually involve rather complicated models and the comparisons are likely to also be complicated. There is the further difficulty that it is difficult to know which of many possible situations are likely to be of much interest.

The following two simple examples may, however, be suggestive of situations in which estimates based on subset means may be useful:

<u>Example I</u>

Consider a two equation model like that given by (2.1), (2.2)--

$$y_{1t} + \beta_{12} y_{2t} + \gamma_{11} z_{1t} + \gamma_{10} = u_{1t} \tag{2.1}$$

$$\beta_{21} y_{1t} + y_{2t} + \gamma_{22} z_{2t} + \gamma_{20} = u_{2t} \tag{2.2}$$

Suppose the investigator cannot observe z_{2t} but that he knows that z_{1t}, z_{2t} are connected by a non-linear relation (which may or may not involve random disturbances). In this case I know of no way of applying limited information or quasi-maximum likelihood methods yet the method based on means of subsets would generally give consistent estimates. The observations could be partitioned according to the magnitude of z_{1t} alone.

Example II

Suppose the first equation is unchanged but the second involves an unknown non-linear relation between y_{1t}, y_{2t}, u_{2t}, say —

(2.1) $$y_{1t} + \beta_{12} y_{2t} + \gamma_{11} z_{1t} + \gamma_{10} = u_{1t}$$

(5.1) $$\varphi(y_{1t}, y_{2t}, u_{2t}) = 0$$

Again I do not believe other methods that have been discussed could be applied, but estimates based on means of subsets would be consistent.

In the absence of general statements about properties of the proposed estimators, it may be of some use to compare the distributions of $\tilde{\mathcal{E}}$ and $\hat{\mathcal{E}}$ for particular examples of the simple type of model considered in Section IV. One such example has been worked out. For this example it was necessary to specify the structural coefficients, the distribution of disturbances, and the pattern of observations on z_{1t}, z_{2t}. The model was given by (2.1), (2.2) and the following values for parameters were assumed —

(5.2)
$$\begin{aligned}
\beta_{12} &= 1 \\
\gamma_{11} &= 1 \\
\gamma_{10} &= 0 \\
\beta_{21} &= -1 \\
\gamma_{22} &= 1 \\
\gamma_{20} &= 0 \\
\sigma^2_{u_1} &= 5 \\
\sigma^2_{u_2} &= 5 \\
\sigma_{u_1 u_2} &= 1
\end{aligned}$$

The reduced form is then given by --

(2.7) $$y_{1t} = \pi_{11} z_{1t} + \pi_{12} z_{2t} + \pi_{10} + v_{1t}$$

(2.8) $$y_{2t} = \pi_{21} z_{1t} + \pi_{22} z_{2t} + \pi_{20} + v_{2t}$$

where

(5.3) $$\pi_{11} = -1/2$$

$$\pi_{12} = 1/2$$

$$\pi_{10} = 0$$

$$\pi_{21} = -1/2$$

$$\pi_{22} = -1/2$$

$${}_{20} = 0$$

$$\sigma^2_{v_1} = 2$$

$$\sigma^2_{v_2} = 3$$

$$\sigma_{v_1 v_2} = 0.$$

30 observations on z_{1t}, z_{2t} were assumed. They are plotted in figure 1. The observations were partitioned into 3 subsets by imagining that the positive half of the z_1-axis was rotated counter clockwise. The observations corresponding to the first 10 points encountered were placed in subset 1, the next 10 in subset 2, and the last 10 in subset 3. The observations in each subset are listed below figure 1.

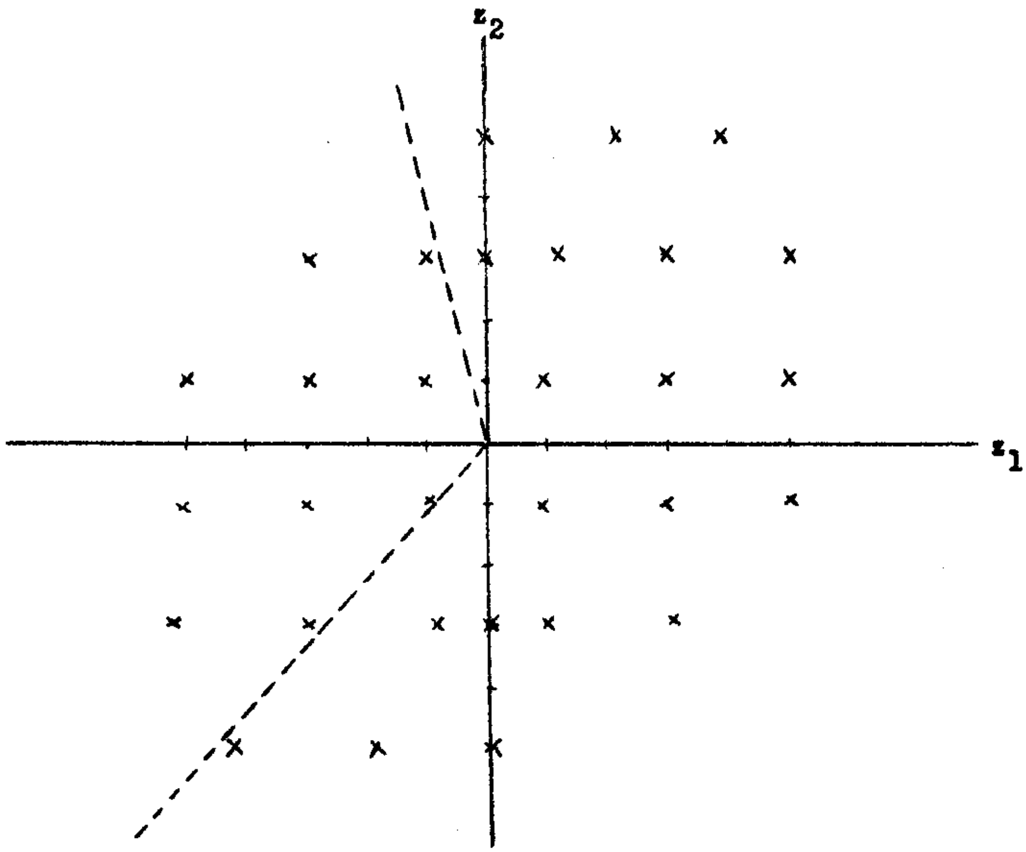

Figure 1.

subset 1	subset 2	subset 3
(1, 1)	(-1, 3)	(1, -1)
(3, 1)	(-3, 3)	(3, -1)
(5, 1)	(-1, 1)	(5, -1)
(0, 3)	(-3, 1)	(-1, -3)
(1, 3)	(-5, 1)	(0, -3)
(3, 3)	(-1,-1)	(1, -3)
(5, 3)	(-3,-1)	(3, -3)
(0, 5)	(-5,-1)	(-4, -5)
(2, 5)	(-3, -3)	(-2, -5)
(4, 5)	(-5, -3)	(0, -5)

This assumed sample gave the subset means and sample moments listed below —

(5.4)

$$\bar{z}_{11} = 2.4$$

$$\bar{z}_{21} = 3.0$$

$$\bar{z}_{12} = -3.0$$

$$\bar{z}_{22} = 0$$

$$\bar{z}_{13} = .6$$

$$\bar{z}_{23} = -3.0$$

$$m_{z_1 z_1} = 9$$

$$m_{z_1 z_2} = 3$$

$$m_{z_2 z_2} = 9$$

and the following values to be used in obtaining $F(\mathcal{E})$ and $G(\mathcal{E})$ —

(5.5)

$$a = z_{12}^2 + z_{12}\, z_{11} + z_{11}^2 = 7.56$$

$$c = z_{21}\, z_{12} - z_{22}\, z_{11} = -9$$

$$d = m_{z_1 z_1}\, m_{z_2 z_2} - m_{z_1 z_2}^2 = 72$$

$$\sigma_{w_2}^2 = \frac{2a}{T}\,\sigma_{v_2}^2 = 1.512$$

$$\sigma_{w_4}^2 = \frac{2a}{T}\,\sigma_{v_4}^2 = 1.008 + 1.512\,(1+\mathcal{E})^2$$

$$\tilde{\lambda}_1 = \frac{\mathcal{E}\,\pi_{22}\, c}{\sigma_{w_4}} = \frac{6.34\,\mathcal{E}}{\sqrt{2+3(1+\mathcal{E})^2}}$$

$$\tilde{\lambda}_2 = -\frac{\pi_{22}\, c}{\sigma_{w_2}} = -3.66$$

$$\sigma^2_{r_2} = \frac{m_{z_1 z_1} d}{T} \sigma^2_{v_2} = 64.8$$

$$\sigma^2_{r_r} = \frac{m_{z_1 z_1} d}{T} \sigma^2_{v_4} = 4.32 + 6.48\,(1+\mathcal{E})^2$$

$$\hat{\ell}_1 = \frac{\mathcal{E}\,\pi_{22}\,d}{\sigma_{r_4}} = \frac{-7.74\,\mathcal{E}}{\sqrt{2+3(1+\mathcal{E})^2}}$$

$$\hat{\ell}_2 = -\frac{\pi_{22}\,d}{\sigma_{r_2}} = 4.47$$

$$\ell = \frac{\hat{\ell}_1}{\bar{\ell}_1} = \frac{\hat{\ell}_2}{\bar{\ell}_2} = -1.22$$

$$\rho_{x_1 x_2} = \frac{\sigma_{v_2 v_4}}{\sigma_{v_2}\sigma_{v_4}} = \frac{-\sqrt{3}(1+\mathcal{E})}{\sqrt{2+3(1+\mathcal{E})^2}}$$

Using the above expressions for limits and for $\rho_{x_1 x_2}$, $F(\mathcal{E})$ and $G(\mathcal{E})$ were calculated[7] for various values of $\mathcal{E}$. The result is plotted in figure 2.

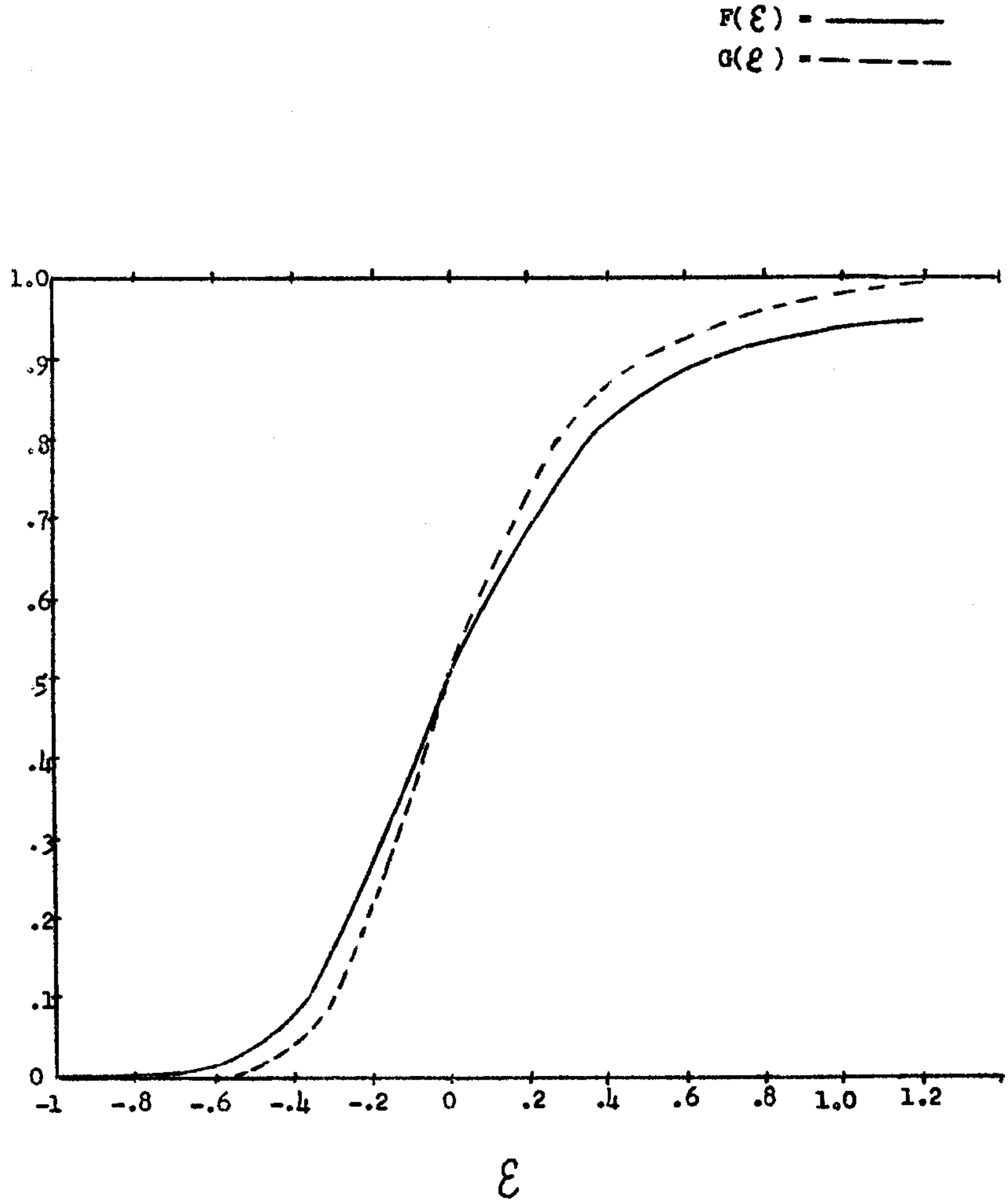
F($\mathcal{E}$) =
G($\mathcal{E}$) =
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
-1
-.8
-.6
-.4
-.2
0
.2
.4
.6
.8
1.0
1.2
$\mathcal{E}$

Figure 2

From figure 2 the decile boundaries of $\tilde{\mathcal{E}}$ and $\hat{\mathcal{E}}$ can be approximated. These are compared in the table below.

P	$\mathcal{E}$ for which $F(\mathcal{E})=P$	$\mathcal{E}$ for which $G(\mathcal{E})=P$
.1	-.36	-.30
.2	-.26	-.22
.3	-.17	-.14
.4	-.08	-.07
.5	.00	.00
.6	.10	.08
.7	.21	.17
.8	.37	.29
.9	.66	.52

1. I have benefited from a number of discussions with John Gurland. Gurland is currently investigating some additional properties of the estimators proposed in this paper. William Kruskal furnished references on the distribution of the ratio of two normally distributed variables. Dan Waterman was consulted on several technical points.

2. Wald's procedure has also been discussed by Lindley [2], Nair and Shrivastava [3], Nair and Banerjee [4], Bartlett [5], and Thiel [6]. Thiel considers the problem of obtaining confidence intervals when there aremore than two variables and also for systems of equations with errors in variables but not in equations.

3. The discussion would be modified only slightly if the equation to be estimated were assumed linear in known functions of the observed variables.

4. Adaptation of the general procedure to particular models need not be confined to the manner of partitioning the observations. For example, if the investigator believed the disturbance in his equation to be highly non-normal, he might base his estimates on medians of subsets rather than on means. It is also possible that, in some models, omitting observations for which all of the predetermined variables have values near their means may improve efficiency.

5. The distribution of the ratio of two normally distributed variables has been discussed by Fieller [7], Craig [8], Geary [9], Bose [10], and Epstein [11].

6. A geometric argument to this effect is presented in Appendix I. I hope that a more precise statement can be made later.

7. These computations were made by William Parrish.

Appendix I

The first part consists of a simple and non-rigorous geometric argument that, when $|\mathcal{L}| > 1$, $G(\mathcal{E})$ is typically less than $F(\mathcal{E})$ for $\mathcal{E} < 0$ and for $\mathcal{E} > 0$ $G(\mathcal{E})$ is typically greater than $F(\mathcal{E})$. From (4.25) and (4.28) it is plain that $F(\mathcal{E})$ and $G(\mathcal{E})$ are unchanged if the signs of both limits are changed. There will thus be no loss of generality in assuming that both $\pi_{22}\,c$ and $\pi_{22}\,d$ are positive.

From (4.23), (4.26), (4.30) we have —

$$(1) \qquad \frac{\tilde{\ell}_1}{\tilde{\ell}_2} - \frac{\hat{\ell}_1}{\tilde{\ell}_2} = -\frac{\mathcal{E}\,\sigma_{v_2}}{\sqrt{\sigma^2_{v_1} + 2(\beta_{12} + \mathcal{E})\sigma_{v_1v_2} + (\beta_{12} + \mathcal{E})^2\,\sigma^2_{v_2}}}$$

and from (4.24) —

$$(2) \qquad \rho_{x_1x_2} = -\frac{\sigma_{v_1v_2} + (\beta_{12} + \mathcal{E})\,\sigma^2_{v_2}}{\sigma_{v_2}\sqrt{\sigma^2_{v_1} + 2(\beta_{12} + \mathcal{E})\sigma_{v_1v_2} + (\beta_{12} + \mathcal{E})^2\,\sigma^2_{v_2}}} \quad .$$

For the case where $\hat{\ell}_i > \tilde{\ell}_i$ $(i=1,2)$ and $\mathcal{E} > 0$, the areas over which $h(x_1,x_2)$ is integrated to obtain $F(\mathcal{E})$ and $G(\mathcal{E})$ are shown in figure 2. The shaded portion represents the area of integration for $F(\mathcal{E})$. The area for $G(\mathcal{E})$ differs in that the portions labeled + and + are included and the portions labeled - and - are excluded.

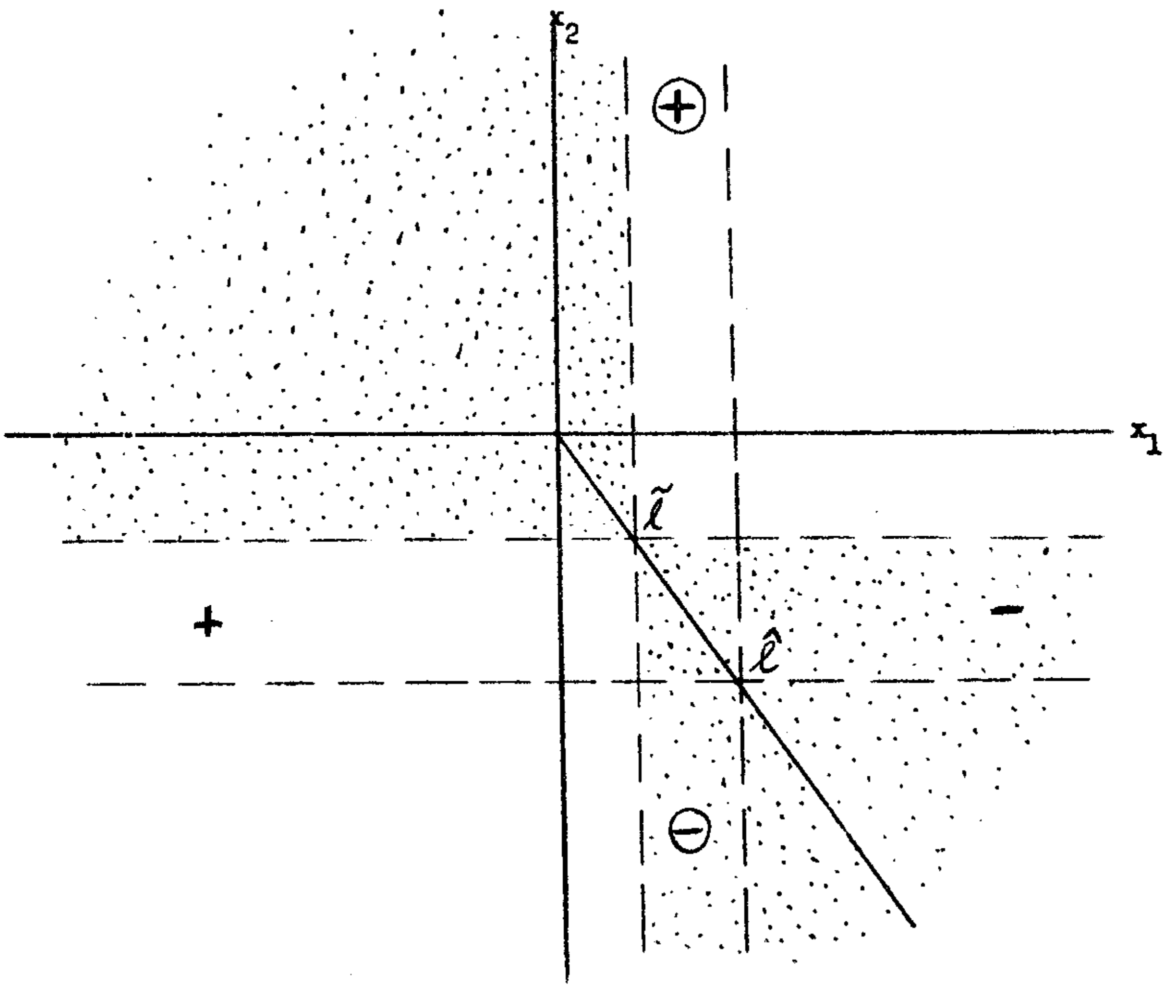

Figure 2

$\tilde{\ell} = (\tilde{\ell}_1 \tilde{\ell}_2)$, $\hat{\ell} = (\hat{\ell}_1 \hat{\ell}_2)$

The equation $h(x_1, x_2) =$ a constant defines a family of ellipses of constant density in the x_1, x_2 plane with centers at the origin. When $\rho > 0$, the major axis of these ellipses is the line $x_1 = x_2$, when $\rho < 0$ the major axis is given by $x_1 = -x_2$, when $\rho = 0$ the ellipses become circles. It is easy to see that when $\rho \lesseqgtr 1$, the integral over + exceeds the integral over −, the integral over + exceeds the integral over −, and $G(\mathcal{E}) > F(\mathcal{E})$.

The case where $\rho > 0$ is illustrated in figure 3 and seems indeterminate on the simple analysis used here. In this case the integral over + exceeds that over - but $\ominus$ exceeds $\oplus$.

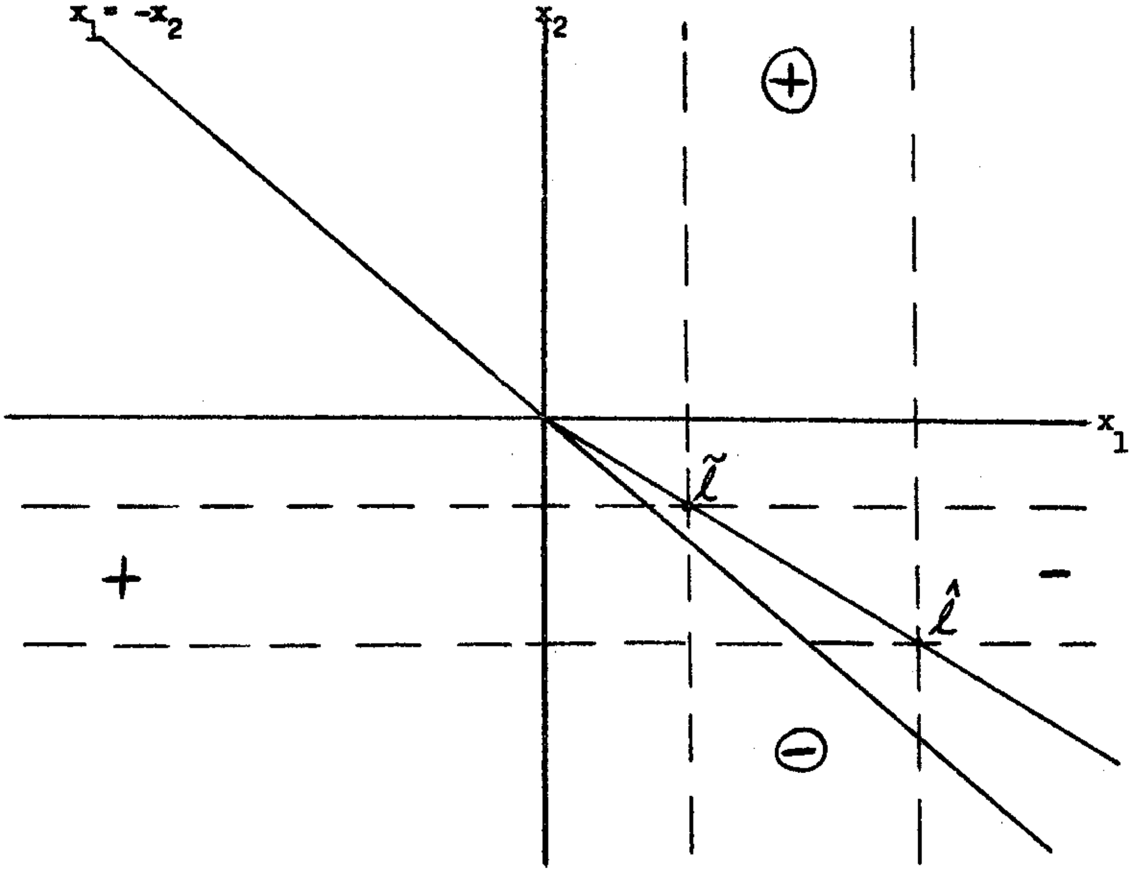

Figure 3

From (1) and (2) it follows that --

(3)
$$\lim_{\varepsilon \to \infty} \frac{\tilde{\ell}_1}{\tilde{\ell}_2} = \lim_{\varepsilon \to \infty} \frac{\hat{\ell}_1}{\hat{\ell}_2} = -1$$

$$\lim_{\varepsilon \to -\infty} \frac{\tilde{\ell}_1}{\tilde{\ell}_2} = \lim_{\varepsilon \to -\infty} \frac{\hat{\ell}_1}{\hat{\ell}_2} = 1$$

$$\lim_{\varepsilon \to \infty} \rho_{x_1 x_2} = -1$$

$$\lim_{\varepsilon \to -\infty} \rho_{x_1 x_2} = 1$$

The case where $\rho < 0$ and $\tilde{\ell}_1$ is approximately equal to $\tilde{\ell}_2$ is thus of some interest in the behavior of the estimators for large $\mathcal{E}$. The case where $\rho < 0$, $\tilde{\ell}_1 = \tilde{\ell}_2$ is shown in figure 4. Here again it is clear that the contribution of + and ⊕ exceed the diminution of - and ⊖ $G(\mathcal{E}) > F(\mathcal{E})$.

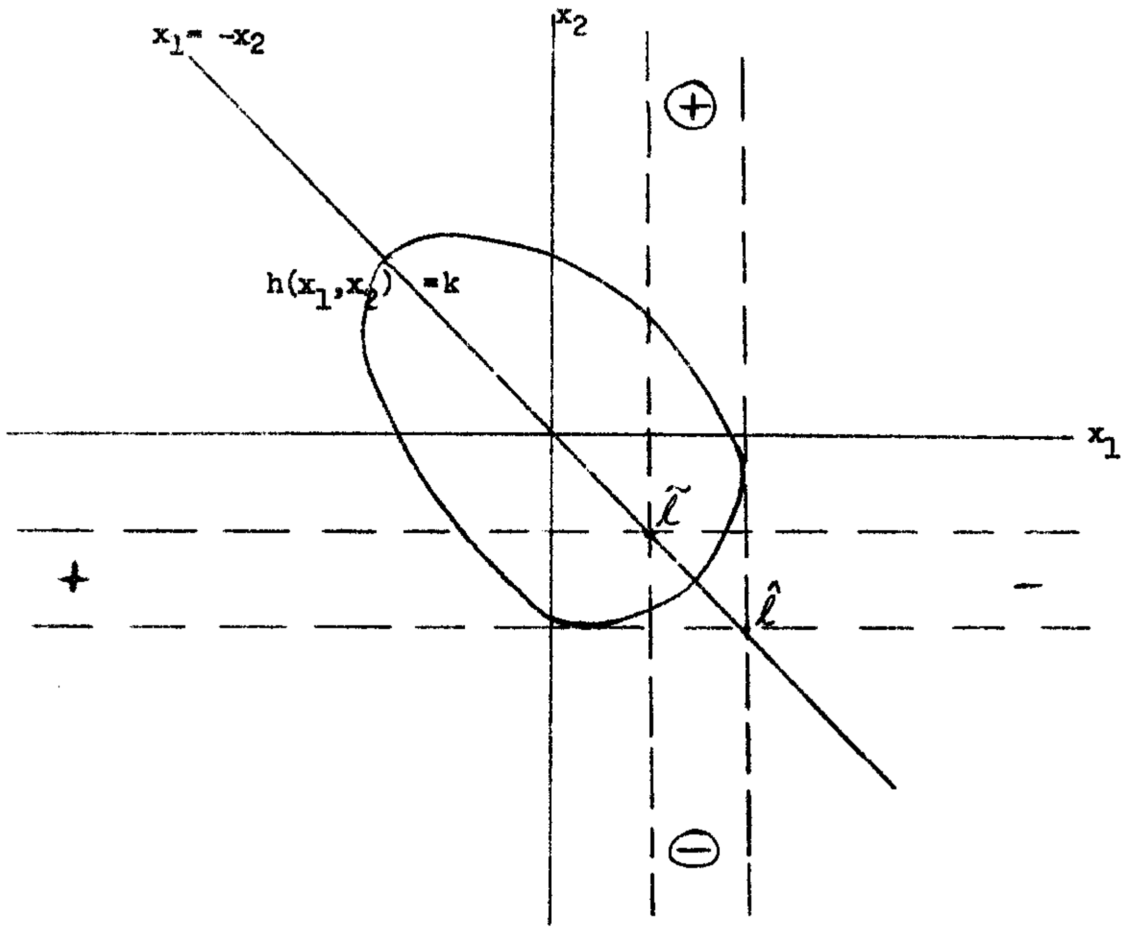

Figure 4

If $\tilde{\ell}_2$ is so small that $P(x_2 < \tilde{\ell}_2)$ is negligible, then $P(x_2 < \hat{\ell}_2)$ is negligible also and $|\hat{\ell}_1| > |\tilde{\ell}_1|$ implies $G(\mathcal{E}) > F(\mathcal{E})$ for $\mathcal{E} > 0$

unambiguously. For the example of Section IV, this was the case and I suspect that it will often prove to be the case for samples as large as 30.

Analogous arguments apply to the case where $\mathcal{E} < 0$. It is desirable to try to obtain more precise comparisons between $F(\mathcal{E})$ and $G(\mathcal{E})$, but I do believe the above arguments justify the statements in the text that for $|\ell| > 1$ and $\mathcal{E} > 0$ we may typically expect that $G(\mathcal{E}) > F(\mathcal{E})$, and for $\mathcal{E} < 0$ that $G(\mathcal{E}) < F(\mathcal{E})$.

It remains to show that $|\ell| \geqq 1$. From (4.31) —

$$(4) \qquad \ell^2 = \frac{2ad}{c^2\, m_{z_1 z_1}}$$

where

$$(4.4) \qquad a = (z_{12}^2 + z_{12}\, z_{11} + z_{11}^2)$$

$$(2.13) \qquad c = (z_{21}\, z_{12} - z_{22}\, z_{11})$$

$$(3.5) \qquad d = m_{z_1 z_1}\, m_{z_2 z_2} - m_{z_1 z_2}^2$$

Consider an arbitrary sample of T observations on z_{1t}, z_{2t} ($t=1,2, \ldots, T$) that has been partitioned into 3 subsets of $\frac{T}{3}$ observations each (for convenience let $T/3$ be an integer). Denote the arbitrary sample by S_3 and the value of ℓ for this sample by $\ell_{(3)}$. Let the observations be ordered so that t takes the values $1,2, \ldots, \frac{T}{3}$ in the first subset; $t= \frac{T}{3}+1, \frac{T}{3}+2, \ldots, \frac{T}{3}$ for observations in the second subset; and $t= \frac{T}{3}+1, \frac{T}{3}+2, \ldots, T$ in the third subset. We then have —

(5) $$\bar{z}_i = \frac{1}{T} \sum_{t=1}^{T} z_{it} \qquad i=1,2$$

(6) $$\bar{z}_{i1} = \frac{3}{T} \sum_{t=1}^{\frac{T}{3}} z_{it} \qquad i=1,2$$

$$\bar{z}_{i2} = \frac{3}{T} \sum_{t=\frac{T}{3}+1}^{2\frac{T}{3}} z_{it}$$

$$\bar{z}_{i3} = \frac{3}{T} \sum_{t=\frac{T}{3}+1}^{T} z_{it}$$

(7) $$Z_{ij} = \bar{z}_{ij} - \bar{z}_i \qquad i=1,2;\ j=1,2,3$$

Consider another sample of size T denoted by S_o and related to S_3 in the following way — each observed value of z_{1t} or z_{2t} in the first subset of S_o is equal to the subset mean of that variable in the first subset of S_3, and similarly for the second and third subsets. Thus there are just three distinct observations in S_o and each has been observed T/3 times. Let $\ell_{(o)}$ be the value of ℓ for the sample S_o, and let $d_{(o)}$, $m^{(o)}_{z_1z_1}$, $m^{(o)}_{z_1z_2}$, $m^{(o)}_{z_2z_2}$ also refer to values computed from S_o.

It can be shown that —

(8) $$\left|\ell_{(o)}\right| = 1$$

Since

(9) $$\ell^2_{(o)} = \frac{2a\, d_{(o)}}{c^2\, m^{(o)}_{z_1z_1}}$$

(10) $$m^{(o)}_{z_1 z_1} = 1/3\ (Z^2_{11} + Z^2_{12} + Z^2_{13})$$

From (5), (7) we know that

(11) $$Z_{13} = -Z_{11} - Z_{12}$$

and therefore

(12) $$m^{(o)}_{z_1 z_1} = 2/3\ (Z^2_{11} + Z^2_{12} + Z_{11}\ Z_{12}) = \frac{2a}{3}\ .$$

Similarly —

(13) $$m^{(o)}_{z_2 z_2} = 2/3\ (Z^2_{21} + Z^2_{22} + Z_{21}\ Z_{22})$$

(14) $$m^{(o)}_{z_1 z_2} = 1/3\ (2Z_{11}\ Z_{21} + 2Z_{12}\ Z_{22} + Z_{11}\ Z_{22} + Z_{12}\ Z_{21})$$

(15) $$d_{(o)} = m^{(o)}_{z_1 z_1}\ m^{(o)}_{z_2 z_2} - (m_{z_1 z_2})^2$$

$$= 1/3\ (Z^2_{11}\ Z^2_{22} + Z^2_{12}\ Z^2_{21} + 2Z_{11}\ Z_{12}\ Z_{21}\ Z_{22})$$

$$= 1/3\ c^2$$

(16) $$\ell^2_{(o)} = \frac{2a\ d_{(o)}}{c^2\ m^{(o)}_{z_1 z_1}} = \frac{2a\ 1/3\ c^2}{c^2\ 2/3\ a} = 1$$

and (8) follows.

Now consider a sample, S_1, in which the observations in the first subset are the same as those of the first subset of S_3 and the observations in the second and third subsets are respectively the same as those of the second and third subsets of S_o. Use $\ell_{(1)}$, $d_{(1)}$, $m^{(o)}_{z_1 z_1}$, $m^{(o)}_{z_1 z_2}$, $m^{(o)}_{z_2 z_2}$ for the relevant parameters of S_1. Let —

(17) $$\xi_{11} = \sum_{t=1}^{T/3} (z_{1t}-\bar{z}_{11})^2$$

$$\xi_{22} = \sum_{t=1}^{T/3} (z_{2t}-\bar{z}_{21})^2$$

$$\xi_{12} = \sum_{t=1}^{T/3} (z_{1t}-\bar{z}_{11})(z_{2t}-\bar{z}_{21})$$

From the fact that $\frac{\xi_{12}}{\xi_{11}\ \xi_{22}}$ is the sample correlation of z_1 and z_2 in the first subset we know that —

(18) $$\xi_{12}^2 \leqq \xi_{11}\ \xi_{22}$$

From the definition of S_1, we have —

(19) $$m^{(1)}_{z_1 z_1} = m^{(o)}_{z_1 z_1} + \xi_{11}$$

$$m^{1)}_{z_2 z_2} = m^{(o)}_{z_2 z_2} + \xi_{22}$$

$$m^{(1)}_{z_1 z_2} = m^{(o)}_{z_1 z_2} + \xi_{12}\ .$$

We wish to show that $\frac{d_{(1)}}{m^{(1)}_{z_1 z_1}} \geqq \frac{d_{(o)}}{m^{(o)}_{z_1 z_1}}$ and therefore that $\ell^2_{(1)} \geqq \ell^2_{(o)}$.

We have

(20) $$\frac{d_{(1)}}{m^{(1)}_{z_1 z_1}} = m^{(1)}_{z_2 z_2} - \frac{\left(m^{(1)}_{z_1\ z_2}\right)^2}{m^{(1)}_{z_1\ z_1}}$$

(21) $$\frac{d_{(o)}}{m^{(o)}_{z_1 z_1}} = m^{(o)}_{z_2 z_2} - \frac{\left(m^{(o)}_{z_1\ z_2}\right)^2}{m^{(o)}_{z_1 z_1}}$$

Subtracting (21) from (20) and using (19) yields —

$$\left(\frac{d_{(1)}}{m^{(1)}_{z_1z_1}} - \frac{d_{(o)}}{m^{(o)}_{z_1z_1}}\right) = \xi_{22} - \frac{\left(m^{(o)}_{z_1z_2} + \xi_{12}\right)^2}{m^{(o)}_{z_1z_1} + \xi_{11}} + \frac{\left(m^{(o)}_{z_1z_2}\right)^2}{m^{(o)}_{z_1z_1}}$$

Letting $D = \dfrac{d_{(1)}}{m^{(1)}_{z_1z_1}} - \dfrac{d_{(o)}}{m^{(o)}_{z_1z_1}}$,

(23) $$m^{(o)}_{z_1z_1}\left(m^{(o)}_{z_1z_1} + \xi_{11}\right) D =$$

$$\xi_{22}\left(m^{(o)}_{z_1z_1}\right) + \xi_{22}\,\xi_{11}\,m^{(o)}_{z_1z_1} + m^{(o)}_{z_1z_1}\left(m^{(o)}_{z_1z_2}\right)^2 + \xi_{11}\left(m^{(o)}_{z_1z_2}\right)^2$$

$$- m^{(o)}_{z_1z_1}\left(m^{(o)}_{z_1z_2}\right)^2 - 2\,\xi_{12}\,m^{(o)}_{z_1z_1}\,m^{(o)}_{z_1z_2} - \xi^2_{12}\,m^{(o)}_{z_1z_1}$$

$$= \left\{m^{(o)}_{z_1z_1}\left(\xi_{11}\,\xi_{22} - \xi^2_{12}\right)\right\} + \left\{\xi_{22}\left(m^{(o)}_{z_1z_1}\right)^2\right.$$

$$\left. - 2\,\xi_{12}\,m^{(o)}_{z_1z_1}\,m^{(o)}_{z_1z_2} + \xi_{11}\left(m^{(o)}_{z_1z_2}\right)^2\right\}$$

From (18) we know the first quantity in brackets is non-negative and that the second is greater than $\left(\sqrt{\xi_{22}}\,m^{(o)}_{z_1z_1} - \sqrt{\xi_{11}}\,m^{(o)}_{z_1z_2}\right)^2$ which in turn is non-negative. Since the coefficient of D in (23) is positive we have —

(24) $D > 0$

$$\frac{d_{(1)}}{m^{(1)}_{z_1 z_1}} > \frac{d_{(o)}}{m^{(o)}_{z_1 z_1}}$$

and from (16) and the definitions of $\ell_{(o)}$, $\ell_{(1)}$ —

(25) $$\ell^2_{(1)} \geqq \ell^2_{(o)} = 1.$$

We can now define a sample, S_2, whose observations are the same as those of S_3 in subsets 1 and 2 and the same as S_o in subset 3. We then have —

(26) $$m^{(2)}_{z_1 z_1} = m^{(1)}_{z_1 z_1} + \eta_{11}$$

$$m^{(2)}_{z_2 z_2} = m^{(1)}_{z_2 z_2} + \eta_{22}$$

$$m^{(2)}_{z_1 z_2} = m^{(1)}_{z_1 z_2} + \eta_{12}$$

where

(27) $$\eta_{11} = \sum_{t=\frac{T}{3}+1}^{2\frac{T}{3}} (z_{1t} - \bar{z}_{12})^2$$

$$\eta_{22} = \sum_{t=\frac{T}{3}+1}^{2\frac{T}{3}} (z_{2t} - \bar{z}_{22})^2$$

$$\eta_{12} = \sum_{t=\frac{T}{3}+1}^{2\frac{T}{3}} (z_{1t} - \bar{z}_{12})(z_{2t} - \bar{z}_{2t})$$

Steps (20) to (25) can be repeated to show that --

(28) $$\ell_{(2)}^{2} \geqq \ell_{(1)}^{2}$$

In a similar fashion, we could obtain --

(29) $$\ell_{(3)}^{2} \geqq \ell_{(2)}^{2} \geqq \ell_{(1)}^{2} \geqq \ell_{(o)}^{2} = 1.$$

From the proof it is apparent that the equalities hold only if S_3 consists of three sets of identical points with equal numbers of observations in each set.

REFERENCES

[1] WALD, A., "The Fitting of Straight Lines if Both Variables are Subject to Error," Annals of Mathematical Statistics 11, 1940 284 pp., 1940

[2] LINDLEY, D. V., "Regression Lines and the Linear Functional Relationship," Journal of the Royal Statistical Society (Supplement) 9, 1947, 218 pp.

[3] NAIR, K. R., and SHRIVASTAVA, M. P., "On a Simple Method of Curve Fitting," Sankhya 6, 1942, 121 pp.

[4] NAIR, K. R., and BANERJEE, K. S., "A Note on Fitting of Straight Lines if Both Variables are Subject to Error," Sankhya 6, 1942, 331 pp.

[5] BARTLETT, M. S., "Fitting a Straight Line When Both Variables are Subject to Error," Biometrics, Vol. 5, No. 3, 1949.

[6] THEIL, H., "A Rank-Invariant Method of Linear and Polynomial Regression Analysis," I and II, Proceedings Vol. LIII, Nos. 3 and 4, 1950.

________, "A Rank-Invariant Method of Linear and Polynomial Regression Analysis," III, Proceedings Vol. LIII, No. 9, 1950.

[7] FIELLER, E. C., "The Distribution of the Index in a Normal Bivariate Population," Biometrika, Vol. 24, p. 428.

[8] CRAIG, C. C., "On Frequency Distributions of the Quotient and of the Product of Two Statistical Variables," The American Mathematical Monthly, Vol. 49, p. 24.

[9] GEARY, R. C., "The Frequency Distribution of the Quotient of Two Normal Variates," Journal of the Royal Statistical Society, Vol. 93, p. 442.

[10] BOSE, R. C., "A Note on Two Series of Balanced Incomplete Block Designs," Bulletin of the Calcutta Mathematical Society, Vol. 34, 1942, p. 129.

[11] EPSTEIN, BENJAMIN, "Some Applications of the Mellin Transform in Statistics," Annals of Mathematical Statistics, Vol. 19, p. 370.

[12] ANDERSON, T. W. and RUBIN, HERMAN, "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," Annals of Mathematical Statistics, Vol. 20, p. 46.

[13] ANDERSON, T. W. and RUBIN, HERMAN, "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations," Annals of Mathematical Statistics, Vol. 21, p. 570.

[14] ANDERSON, T. W., "Estimation of the Parameters of a Single Equation by the Limited-Information Maximum Likelihood Method," Chapter IX of Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph No. 10, edited by Tjalling C. Koopmans.