Sankhyā : The Indian Journal of Statistics
1984, Volume 46, Series B, Pt. 1, pp. 90–107.

GROUPING AND RANK ESTIMATORS IN EVMS

By MRINAL BHAUMIK
Indian Statistical Service
and
MANORANJAN PAL
Indian Statistical Institute

SUMMARY. This paper first discusses various grouping/rank estimation methods in the EVMs. In doing so it gives an exposition of some of the recent work done in this area and compares them. An improved estimator is proposed and is compared with other estimators. It is shown that this estimator is uniformly better than the other estimators in the sense of minimizing asymptotic variance within the class of estimators.

Second, to prove an important result of the paper, it is assumed that the ranks of the true values of the regressor are known from an independent source. In this case, an optimum minimum variance rank estimator is found. The optimisation of the class of estimators proposed in this paper leads to a simple and elegant approach which can easily be applied in other similar situations.

1. Introduction

Various alternative methods of estimation in "Errors-in-Variables" (EIV's) models have been suggested by researchers in this field. These are based on different sets of assumptions depending on the specific situation. Thus some assume X, the true regressor, to be stochastic (structural models) while others do not assume so (functional models). Consistent estimators can be obtained if

(i) One has prior knowledge about the values of the error variances of the respective true variables, or if

(ii) "instrumental variables" (IV's) are available with specific properties.

Introduction of lagged values of regressors/regressand may also be helpful in finding consistent estimates of the parameters. But what happens if no such prior information is available ? The question has long been answered by Reiersol (1950), who has proved that no consistent estimators of the regression parameters are possible in the standard two-variable "EIV models" (EVM's) where the associated variables are assumed to be normally distributed.[1] The OLS estimate is obviously biased and the bias does not

[1] In the multivariate EVM also such a result exists (see Pal, 1982).

AMS (1980) *subject classification :* Primary 62J05, Secondary 62F10.

Key words and phrases : Errors-in-variables, Grouping estimators, Rank estimators, Instrumental variables

decrease to zero as the sample size increases indefinitely. If one regresses x (the observed regressor) on y (the observed regressand), then the "reverse least squares" (RLS) estimator is also biased and inconsistent. However, the limiting values of the two estimators bound the regression coefficient. The existence of such bounds seems to be the only comfort in unidentified models.

The grouping estimators in an unidentified model first proposed by Wald (1940) is also biased, but is much simpler than the other methods and hence/or otherwise may easily be applied in many practical cases. One more advantage of the grouping method is that it can allow for errors in the regressor to a certain extent : If the rankings/groupings of x_i's are the same as those of the underlying values (X_i's) then these estimators are consistent. It is, however, true that there is some loss of efficiency if OLS method is valid (i.e., if the regressor is free from errors).

This paper first discusses various grouping/rank estimation methods in EVM's. In doing so it gives an exposition of some of the recent work done in this area and compares them. An improved estimator is proposed and is compared with other estimators. It is shown that this estimator is uniformly better than the other estimators in the sense of minimizing asymptotic variance within the class of estimators.

Second, to prove an important result of the paper, it is assumed that ranks of the true values of the regressor are known from an independent source.[2] In this case, an optimum minimum variance rank-estimator is found. The optimisation of the class of estimators proposed in this paper leads to a simple and elegant approach which can easily be applied in other similar situations.

2. The survey

For the standard two-variable EVM, Wald (1940) proposed grouping estimators of the regression coefficients in which the observed values (x_i, y_i), $i = 1, 2, \ldots, n$ are divided into two equal groups according to the rank of the x_i's, and the centres of gravities of the two groups in the scatter diagram are then joined by a straight line to find the slope estimator. Later, Bartlett (1949) suggested the use of three groups with equal number of observations in each group according to the order of x_i's. Here, the centres of gravities of the two extreme groups are joined by a straight line to estimate the parameters. Suppose the group means of the two extreme groups are $(\bar{x}_i, \bar{y}_i)$

[2] e.g., successive time series observations on growth of a plant must be of increasing order leading to the obvious ranks 1, 2, ... etc.

and $(\bar{x}_3, \bar{y}_3)$ respectively. The grouping estimator of β, the scale parameter, is then defined by

$$b = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1}. \quad \ldots (1)$$

It is not necessary to take equal number of observations in each group. This choice is, however, optimal in the Gauss-Markov set-up if the X's are equi-spaced. In general, there is a problem of optimal choice of the three groups if one decides to use an estimator of the form of b. Theil and Van Yzeren (1956) have obtained the optimum group proportions for different distributions of X. The optimum proportions turned out to be approximately 0·3 : 0·4 : 0·3 for a variety of distributions considered by them.[3] The prevailing opinion seems to be that the three group estimator of Bartlett (with equal groups) is nearly optimal in almost all cases. The conclusion is, however, based on inadequate amount of investigation. The distribution of X examined so far (mainly by Theil and Van Yzeren) are mostly symmetric or negatively skewed. In most empirical application in economics the X-distribution is highly positively skewed. Hence the rule specified by Bartlett and Theil and Van Yzeren may not be applicable in all cases.[4]

In Pal and Bhaumik (1981) the optimum group proportions assuming that X follows (i) the lognormal distribution, and (ii) the gamma distribution, were found by minimizing $V(b)$ (variance of b). The most important finding is that the optimum proportions in the three groups (in ascending order of X) are quite stable around the values (0·40 : 0·45 : 0·15) for the commonly occuring lognormal or gamma type distributions of X. The estimators would be highly efficient (about 80 per cent) relative to OLS if the group proportions are near the optimum values. Further, the gain in efficiency appears to be considerable if one uses the optimum group proportions instead of equal groups as in common Bartlett estimator.

A similar investigation was carried out by the author and others (see Sil, Bhaumik and Pal, 1981) for the case where the disturbances are heteroscedastic. Actually, we studied the set-up where

$$V(\varepsilon_i/X_i) = \lambda X_i^p, \quad \lambda > 0, \quad \ldots (2)$$

[3] For the case where errors in X-values are absent, the Bartlett estimator for three equal groups had been proposed by Nair and Shrivastava (1942), who considered equi-spaced X-values. Nair and Banerjee (1943) later showed that this remains efficient even if errors are present.

[4] Only Gibson and Jowett (1957) studied, among other types, two particular forms of gamma distribution.

as considered by Lancaster (1968), who examined the efficiency of Wald's and Bartlett's (equal-groups) estimators *vis-a-vis* GLS destimator. Here also the optimum estimators are found to be highly efficient *vis-a-vis* "generalized least squares" (GLS) estimator which is BLUE. Their efficiency is about 80 per cent, and further, the increase in efficiency over Bartlett's estimator is quite considerable. Unlike the homoscedastic case, the optimum group proportions in the heteroscedastic case are highly dependent on p, the degree of heteroscedasticity. As p increases, the first group proportion (lowest X-values) decreases and the third group proportion (highest X-values) increases. But the sum of the two extreme group proportions is fairly stable. This means that the choice of optimum group proportions should be made in the light of some approximate idea regarding p. In other respect, the optimum group proportions seem to be nearly stable with respect to the types of distributions or parameters. The efficiencies of the grouping estimators in most of the cases decrease as the coefficient of varation of X increases.

The grouping estimators have some obvious advantages. One little known point is that variance of the GLS estimator contains terms like $E(X^{-p})$ which may not exist in all cases. There may be trouble even with OLS. The distribution of X may be such that $E(X^{1+p})$ or $E(X^{2+p})$ does not exist whereas $E(X^p)$ exists. Moreover, Lancaster showed the superiority of grouping estimators over OLS estimator in the heteroscedastic set-up for some values of p.

The grouping methods due to Wald, Bartlett, and Theil and Van Yzeren are special cases of IV method. Durbin (1954) proposed an estimator where the IV's are ranks. Theil (1950), however, proposed a different type of estimator which is neither a grouping estimator nor a ranking estimator, but is worth mentioning in this paper. He defined

$$\Delta(i,j) = (y_i - y_j)/(x_i - x_j), \quad i = 1, 2, \ldots, j; \quad j = 2, 3, \ldots, n; \quad i \neq j. \quad \ldots \quad (3)$$

The sample median of the $\binom{n}{2}$ values of $\Delta(i, j)$ was proposed as an estimator of β the regression coefficient.

Dorff and Gurland (1961), Housner and Brennan (1948), Richardson and Wu (1970) etc. were among others who tried their hands on the grouping problem. Our subsequent discussion will reveal some of their work whenever necessary.

3. The model

The classical two variable EVM forms the basis of several investigations and receives considerable attention. This model is specified as

$$\left.\begin{aligned} Y_i &= \alpha+\beta X_i+\varepsilon_i, \qquad i = 1, 2, \ldots, n \\ x_i &= X_i+u_i \\ y_i &= Y_i+v_i, \end{aligned}\right\} \qquad \ldots \quad (4)$$

where α and β are parameters to be estimated. ε_i is the random normal disturbance with mean zero and independent of other variables not involving Y_i. X_i and Y_i are true but unknown variables. The observed variables are x_i and y_i with unknown errors u_i and v_i respectively. u_i, v_i are EIV's independent of each other and of X_i, Y_i and ε_i. We assume that the successive observations are drawn independently. Moreover, ε_i, u_i and v_i are assumed to be normally distributed, independent of each other and having zero means and finite variances.

The above mentioned model is homoscedastic, because $V(\varepsilon_i \mid X_i)$ is constant for all $i = 1, 2, \ldots, n$. This model can further be generalized by taking heteroscedastic disturbance of the following type :

$$V(\varepsilon_i \mid X_i) = \lambda X_i^p, \quad \lambda > 0, \qquad \ldots \quad (5)$$

where λ and p are constants. Obviously, taking $p = 0$, one may reduce it to the homoscedastic case. The investigation by Katona *et al.* (1954), Goldberger (1964), and Lancaster (1968) reported p to be positive and especially between 1 and 2 in all cases, though theoretically it may take negative values also.

4. The proposed general grouping estimator

Let $X_1, X_2, \ldots, X_n$ be arranged in a non-decreasing order, i.e., $X_1 < X_2 < \ldots < X_n$. Let there be k groups with number of observations $n_1, n_2, \ldots, n_k$ respectively such that $\sum_{i=1}^{k} n_i = n$ and

$$\bar{X}_j = \frac{1}{n_j} \sum_{i \in I_j} X_i, \qquad \bar{Y}_j = \frac{1}{n_j} \sum_{i \in I_j} Y_i; \; j = 1, \ldots, k,$$

where, $I_j = \{n_1+\ldots+n_{j-1}+1, n_1+\ldots+n_{j-1}+2, \ldots, n_1+\ldots+n_j\}$. The proposed class of estimator is

$$\beta_a = \sum_1^k a_j \bar{Y}_j \Big/ \sum_1^k a_j \bar{X}_j \qquad \ldots \quad (6)$$

$$= \beta + \sum_1^k a_j \bar{\varepsilon}_j \Big/ \sum_1^k a_j \bar{X}_j, \qquad \ldots \quad (7)$$

such that $\sum_1^k a_j = 0$. Assuming X to be non-stochastic, asymptotic variance of β_a in the heteroscedastic model is

$$\bar{V}(\beta_a) = \sum_1^k a_j^2 V(\bar{\varepsilon}_j) \Big/ \Big(\sum_1^k a_j \bar{X}_j \Big)^2 \qquad \ldots (8)$$

$$= \sum_1^k \Big(a_j^2 \lambda \frac{1}{n_j^2} \sum_{i \in I_j} X_i^p \Big) \Big/ \Big(\sum_1^k a_j \bar{X}_j \Big)^2$$

$$= \sum_1^k w_j a_j^2 \Big/ \Big(\sum_1^k a_j \bar{X}_j \Big)^2, \qquad \ldots (9)$$

where $$w_j = \lambda n_j^{-2} \sum_{i \in I_j} X_i^p.$$

Given the values of w_j's, our job is to minimize $\bar{V}(\beta_a)$ with respect to a_j's such that $\sum_1^k a_j = 0$. Assuming continuity and differentiability of $\bar{V}(\beta_a)$ with respect to a_j's, the minimization condition reduces to (see appendix I)

$$a_j \propto p_j^*(\bar{X}_j - \bar{X}^*), \quad j = 1, \ldots, k, \qquad \ldots (10)$$

where $p_j^* = 1/w_j$ and $\bar{X}^* = \Big(\sum_1^k \bar{X}_j/w_j \Big) \Big/ \Big(\sum_1^k 1/w_j \Big)$. Optimum variance then becomes

$$\bar{V}(\beta_0) = \frac{\sum_1^k w_j p_j^{*2} (\bar{X}_j - \bar{X}^*)^2}{\Big\{ \sum_1^k p_j^* (\bar{X}_j - \bar{X}^*) \bar{X}_j \Big\}^2}$$

$$= \frac{\sum_1^k p_j^* (\bar{X}_j - \bar{X}^*)^2}{\Big\{ \sum_1^k p_j^* (\bar{X}_j - \bar{X}^*) \bar{X}_j \Big\}^2}, \text{ since } p_j^* = 1/w_j,$$

or $$\bar{V}(\beta_0) = \frac{1}{\sum_1^k p_j^* (\bar{X}_j - \bar{X}^*)^2}. \qquad \ldots (11)$$

and $$\beta_0 = \frac{\sum_1^k p_j^* (\bar{X}_j - \bar{X}^*)(\bar{Y}_j - \bar{Y})}{\sum_1^k p_j^* (\bar{X}_j - \bar{X}^*)^2}. \qquad \ldots (12)$$

Remark 1 : This is a weighted regression estimator where the weight for the j-th group is p_j^*. But there is still a difference. Instead of $\bar{X}$, here we have $\bar{X}^*$ a weighted mean of group means which is different from the grand mean (unless $p = 0$).

Remark 2 : Once the groups are specified, p_j^*, $\bar{X}_j$, $\bar{Y}_j$, $\bar{X}^*$, $\bar{Y}$ become known (assuming p to be known), and then one can easily calculate β_0 and its variance, which is optimum given the group specification.

Remark 3 : In fact, one can further optimize the estimator. Given a distribution of X and the number of groups to be taken, one can vary the group proportions so as to arrive at a further minimum variance. This global minimum will obviously give a better estimate.

Remark 4 : For the homoscedastic case, one simply puts $p = 0$ and gets

$$p_j^* a n_j/n = p_j \text{ (say) and } \bar{X}^* = \bar{X}$$

$$\bar{V}(\beta_0) = 1\Big/\sum_1^k p_j(\bar{X}_j - \bar{X})^2 \qquad \dots \quad (13)$$

and

$$\beta_0 = \sum_1^k p_j(\bar{X}_j - \bar{X})(\bar{Y}_j - \bar{Y})\Big/\sum_1^k p_j(\bar{X}_j - \bar{X})^2.$$

This is the usual weighted regression estimator as found by Dorff and Gurland (1961) and Richardson and W_u (1970).

Remark 5 : For $K = 2$ and $p = 0$; β_a becomes

$$\beta_a = \frac{a_1\bar{Y}_1 + a_2\bar{Y}_2}{a_1\bar{X}_1 + a_2\bar{X}_2},$$

such that $a_1 + a_2 = 0$, i.e.,

$$\beta_a = \frac{\bar{Y}_1 - \bar{Y}_2}{\bar{X}_1 - \bar{X}_2}. \qquad \dots \quad (14)$$

This does not involve any unknown parameter, and hence the question of optimization simply does not arise in this case. One can of course optimize it through variation of groups.

Remark 6 : Dorff and Gurland investigated some special situations in the homoscedastic case. These situations help us to have a better insight to the estimators considered by Wald and Bartlett (b_w and b_B, say).

(i) Suppose

$$\begin{aligned} X_i &= B - A && \text{if } i = 1, 2, \dots, N/3 \\ &= B && \text{if } i = N/3+1, \dots, 2N/3 \\ &= B + A && \text{if } i = 2N/3+1, \dots, N, \end{aligned}$$

then Bartlett's estimator is optimum.

(ii) Again if

$$\begin{aligned} X_i &= A && \text{if } i = 1, 2, \dots, N/2 \\ &= B && \text{if } i = N/2+1, \dots, N. \end{aligned}$$

Then Wald's estimator is optimum.

(iii) Suppose X_i's are equally spaced (i.e., $X_{i+1}-X_i =$ constant $\forall$ i), then Bartlett's estimator is optimum among its class.

Dorff and Gurland (1961) also pointed out that $\bar{V}(b_D)$ (also $\bar{V}(b_H)$: due to Housner and Brennan)[5] does not greatly exceed $\bar{V}(\beta_0)$ unless the X_i's are highly skewed or bunched. $b_H (= b_D)$ is in this respect more robust than b_W and b_B.

Kiefer and Wolfowitz (1956) have suggested a ML estimator for the non-normal model with one regressor (see also Wolfowitz, 1952). Neyman (1951) also provided a consistent estimator for the non-normal structural model with one regessor. The formulae and approach suggested by them are too complicated to be useful in practice. The practical methods like moment or cumulant estimators have a large variance in some cases. These observation suggest that grouping estimator may be used in these cases.

The aim of the present investigation is to find optimal weighted grouping estimators for lognormal and gamma-type distributions of X for a wide range of parameters. Earlier, grouping estimators were optimized only through group variations. The present study also allows weights to each group. The estimator thus obtained must be superior to the other grouping estimators. For illustration we take two distributions for X to see how much better the present method is. We take only three groups for this purpose.

5. An illustration

Table 1 and Table 2 show the efficiencies of the weighted optimum grouping estimators and the usual unweighted optimum Bartlett type grouping estimators with respect to OLS estimator (V_L/V^* and V_L/V_0 respectively) for different parameters of lognormal and gamma distributions. It should be noted here that both Bartlett and Theil and Van Yzeren considered the classical set-up for two-variable linear regression where x is free from errors. The same approach is followed in this investigation also. Note also that if x is free from error OLS estimator is BLUE. The optimum weights of the second group (A_2) are also tabulated. The weights for other groups can be found once we note that $A_1+A_2+A_3 = 0$ and A_3 is normalized to 1. F_1^* and $1-F_2^*$ are the proportions of first and last groups respectively for the weighted grouping optimum estimator. For Bartlett type optimum grouping estimator, the corresponding figures are tabulated under the heading F_1^0 and $1-F_2^0$ respectively.

The following observations may be made from Table 1 and Table 2 :

[5] b_D is the Durbin's rank-estimator. This, in fact, was earlier proposed by Housner and Brennan (1948).

TABLE 1. EFFICIENCIES OF DIFFERENT ESTIMATES WHERE REGRESSOR $X \sim \Lambda(\cdot, \sigma^2)$

σ^2	LR_π and CV_π	γ_1^+ and γ_2^+	F_1^* and $1-F_2^*$	F_1^0 and $1-F_2^0$	V^* and V_L	efficiency vis-a-vis OLS: Wald and Bartlett equal group	efficiency vis-a-vis OLS: weighted optimum and Bartlett optimum	percentage gain in efficiency: weighted optimum over Bartlett optimum	percentage gain in efficiency: weighted optimum over Bartlett equal group	A_2
0.2	0.25 0.47	1.52 4.35	0.53 0.10	0.35 0.17	0.629 0.500	0.54 0.68	0.80 0.77	3.9	17.6	1.6
0.3	0.30 0.59	1.98 7.71	0.60 0.07	0.38 0.15	0.363 0.287	0.50 0.63	0.79 0.76	3.9	25.4	1.4
0.4	0.34 0.70	2.45 12.27	0.66 0.05	0.42 0.13	0.236 0.184	0.46 0.58	0.78 0.73	6.8	34.5	1.3
0.5	0.38 0.81	2.94 18.51	0.68 0.05	0.43 0.12	0.163 0.127	0.42 0.54	0.77 0.72	6.9	42.6	1.2
0.6	0.42 0.91	3.47 27.08	0.70 0.05	0.44 0.10	0.118 0.090	0.38 0.50	0.76 0.70	8.6	52.0	1.1

π : LR = Lorenz ratio, CV = coefficient of variation

$+$: $\gamma_1 = \mu_3/\mu_2^{3/2}$ and $\gamma_2 = \mu_4/\mu_2 - 3$ where μ_i is the ith central moment

TABLE 2: EFFICIENCIES OF DIFFERENT ESTIMATES WHERE REGRESSOR $X \sim (., r)$

r	LR^{π}; and CV^{π}	γ_1^+ and γ_2^+	F_1^* and $1-F_2^*$	F_1^0 and $1-F_2^0$	V^* and V_L	efficiency *vis-a-vis* OLS: Wald and Bartlett equal group	efficiency *vis-a-vis* OLS: Weighted optimum Bartlett optimum	percentage gain in efficiency: weighted optimum over Bartlett optimum	percentage gain in efficiency: weighted optimum over Bartlett equal group	A_2
1.5	0.42 0.82	1.63 4.00	0.56 0.11	0.40 0.17	0.846 0.667	0.52 0.66	0.79 0.76	3.9	19.7	1.6
2.0	0.37 0.71	1.41 3.00	0.54 0.11	0.39 0.17	0.614 0.500	0.55 0.69	0.81 0.79	2.5	17.4	1.7
2.5	0.34 0.63	1.26 2.40	0.51 0.12	0.38 0.18	0.491 0.400	0.57 0.71	0.82 0.80	2.5	15.5	1.7
3.0	0.31 0.58	1.15 2.00	0.49 0.12	0.36 0.20	0.409 0.333	0.58 0.72	0.82 0.80	2.5	13.9	1.7
3.5	0.29 0.53	1.07 1.72	0.47 0.13	0.36 0.20	0.350 0.250	0.59 0.74	0.82 0.80	2.5	10.8	1.8

π : LR = Lorenz ratio, CV = coefficient of variation

$+$: $\gamma_1 = \mu_3/\mu_2^{3/2}$ and $\gamma_2 = \mu_4/\mu_2^2 - 3$ where μ_i is the ith central moment.

(1) The increase in efficiency due to weights is about 3 to 9 per cent over optimum Bartlett estimator. The efficiency in both the cases decreases as LR or CV increases, but the gain in efficiency increases as LR or CV increases.

(2) As σ^2 of lognormal distribution increases, the lorenz ratio (LR), coefficient of variation (CV), coefficient of skewness (γ_1) and Kurtosis (γ_2) increase, whereas r of gamma distribution increases, LR, CV, γ_1 and γ_2 decrease. The optimum first group proportions (F_1^* and F_1^0) increases as σ^2 in lognormal distribution increases and r in gamma distribution decreases. Hence it can be thought to be a property of LR or CV.[6] i.e., As LR or CV increases the proportion of first group increases and the proportion of third group decreases for both the distributions.

(3) The optimum weighted first/third group proportions are respectively uniformly higher/lower than those of the corresponding optimum Bartlett group proportion for the lognormal and gamma distributions. The optimum weight of the second group varies with LR or CV in case of lognormal distribution whereas it is not so noticeable in the gamma distribution.

(4) In Engel curve analysis, a typical parameter for σ^2 is 0·3 and in that case $F_1^* = 0{\cdot}60$, $1-F_2^* = 0{\cdot}07$ and $A_2 = 1{\cdot}4$, i.e., one should take 60 percent in the first group, 33 percent in the second group and the remaining 7 per cent in the last group, the weights being $-2{\cdot}4$, $1{\cdot}4$ and 1 respectively in the first, second and third groups.[7]

6. Estimation when the ranks of the true regressor are known

In this section we take only the homoscedastic model, i.e., the standard two-variable EVM. Moreover, we suppose that there exists an IV Z such that

$$\text{rank}(X_i) = \text{rank}(Z_i) \forall_i.$$

Without loss of generality we may assume $Z_1 \leqslant Z_2 \leqslant \dots \leqslant Z_n$, and hence $X_1 \leqslant X_2 \leqslant \dots \leqslant X_n$. The proposed estimator is

$$b = \frac{\sum_1^n (X_i - \bar{X}) y_i}{\sum_1^n (X_i - \bar{X}) x_i} \qquad \dots (15)$$

$$= \frac{\sum_1^n l_i y_i}{\sum_1^n l_i x_i} \text{ such that } \Sigma l_i = 0. \qquad \dots (16)$$

---

[6] γ_1 and γ_2 can not be taken as indicators. This becomes evident once we take symmetric distribution.

[7] The heteroscedastic model was also investigated. But the increase in efficiency in this case was not very substantial.

Our problem is to minimize $\bar{V}(b)$ (the asymptotic variance of b) such that $\Sigma l_i = 0$ and $l_1 < l_2 < \ldots < l_n$. Now,

$$\bar{V}(b) \propto \frac{V(l)}{\mathrm{cov}^2(X, l)} \quad \ldots \quad (17)$$

the estimated $\hat{V}(b)$ is then

$$\hat{V}(b) \propto \frac{V(l)}{\mathrm{cov}^2(x, l)}. \quad \ldots \quad (18)$$

In the classical IV set up one assumes

$$X_i = \gamma_0 + \gamma_1 Z_i + w_i, \quad \ldots \quad (19)$$

with standard assumptions. The ML estimate of this model is the median of OLSE, Reverse LSE and IVE if all of them have the same sign (Leamer, 1978). But in our case the relation between X_i and Z_i is not known. As a first approximation one may assume equation (19) to be true and arrive at the ML estimate. As a second approximation one may take

$$X_i = -p_1 Z_i + p_2 Z_i^2/_2 + e_i. \quad \ldots \quad (20)$$

Even if we assume that all e_i's are zero, the condition for rank (X_i) = rank (Z_i) for all i is that

$$\min\{Z_i\} > p_1/p_2. \quad \ldots \quad (21)$$

One may take similar relations between X and Z and arrive at optimum estimators.

It may so happen that the functional relation between X and Z is not known. The value of Z may even be unknown. The only known thing is the rank of X_i for all i. In this case one has to minimize $\bar{V}(b)$ subject to $X_1 \leqslant \ldots \leqslant X_n$. In other words,

$$\text{minimize } \frac{\Sigma l_i^2}{(\Sigma x_i l_i)^2} \text{ subject to } \Sigma l_i = 0 \text{ and } l_1 < l_2 \leqslant \ldots \leqslant l_n.$$

Lemma 1 : *Suppose* $l_1^* < l_2^* < \ldots < l_k^*$ *are the distinct values (such that* $l_j^* = l_{r_1+\ldots+r_{j-1}+1} = l_{r_1+\ldots+r_{j-1}+2} = \ldots = l_{r_1+\ldots+r_j}\ \forall j = 1, \ldots, k,$ and $\sum_{j=1}^{k} r_j = n$*) which minimizes* $\sum_1^n l_i^2 \Big/ \left(\sum_1^n x_i l_i\right)^2$ *subject to* $\sum_1^n l_i = 0$ *and* $\sum_1^n x_i = 0$ *then*

$$l_j^* = \frac{X_{r_1+\ldots+r_{j-1}+1} + \ldots + X_{r_1+\ldots+r_j}}{r_j}$$

Proof: See appendix I.

Lemma 2 : *Suppose $l_1^* < \ldots < l_k^*$ are the distinct values of l's thus attained as stated in Lemma 1, then*

$$\overline{V}(b) = 1 \Big/ \left(\sum_1^k l_i^{*2} r_i \right)^2.$$

The proof is straight forward.

So, minimization of $\Sigma l_i^2/(\Sigma x_i l_i)^2$ subject to $\Sigma l_i = 0$ and $Ex_i = 0$ reduces to maximization of $(\Sigma r_i l_i^{*2})$ i.e., maximization of Σl_i^2 such that

(i) $\Sigma l_i = 0$ ... (22)

(ii) $l_1 \leqslant \ldots \leqslant l_n$ ... (23)

(iii) whenever $l_i < l_{i+1} = \ldots = l_j < l_{j+1}$, ... (24)

$$l_{i+1} = \ldots = l_j = (x_{i+1} + \ldots + x_j)/(j-i).$$

Theorem : *Consider the sequence of cumulative averages*

$$\overline{T}_1 = x_1,\ \frac{x_1+x_2}{2}\ (= \overline{T}_2),\ \frac{x_1+x_2+x_3}{3}\ (= \overline{T}_3), \ldots, \frac{\sum_1^n x_i}{n}\ (= \overline{T}_n).$$

Suppose the minimum of $\overline{T}_1, \overline{T}_2, \ldots, \overline{T}_n$ is attained at $i_1, i_2, \ldots, i_{j_1}$ places such that

$$\overline{T}_{i_1} = \ldots = \overline{T}_{i_{j_1}} \text{ and } i_1 < i_2 < \ldots < i_{j_1}\ (= r_1, \text{ say}).$$

Put
$$l_1 = l_2 = \ldots = l r_1 = \frac{x_1 + \ldots + x_{r_1}}{r_1}\ (= \bar{x}_1, \text{ say}).$$

Next consider the sequence

$$\overline{T}^{(1)}_{r_1+1} = x_{r_1+1},\ \frac{x_{r_1+1} + x_{r_1+2}}{2}\ (= \overline{T}^{(1)}_{r_1+2}), \ldots, \frac{x_{r_1+1} + \ldots + x_n}{n - r_1}\ (= \overline{T}^{(1)}_n)$$

We then find the minimum of $\overline{T}^{(1)}_{r_1+1}, \ldots, \overline{T}^{(1)}_n$. Suppose it is attained at $i_1', i_2', \ldots, i_{j_2}'$ places such that $\overline{T}^{(1)}_{i'_1} = \ldots = T^{(1)}_{i'_{j_2}}$ and $i_1' < i_2' < \ldots < i_{j_2} (= r_1 + r_2$, say). Put

$$l_{r_1+1} = \ldots = l_{r_1+r_2} = \frac{x_{r_1+1} + \ldots + x_{r_1+2}}{r_2}\ (= \bar{x}_2, \text{ say}) \text{ and so on.}$$

Thus we get

$$\bar{x}_1\ (r_1 \text{ times}) < \bar{x}_2\ (r_2 \text{ times}) < \ldots < \bar{x}_k (r_k \text{ times}) \text{ for } l_i\text{'s}.$$

Correspondingly

$$\sum_1^n l_i^2 = \sum_1^k r_i \bar{x}_i^2 = V_1 \text{ (say).}$$

We do the same for the sequence

$$-x_1, -x_2, \ldots, -x_n,$$

and get corresponding Σl_i^2 to be V_2 (say).

Then, maximization of Σl_i^2 subject to (22), (23) and (24) mentioned above reduces to choosing the maximum value between V_1 and V_2.

Proof: See appendix II.

In most of the cases it becomes clear which of V_1 and V_2 gives the minimum value as one compares the ranks of x_i''s and z_i's. But it is still a problem whether positive correlation between ranks of x_i''s and z_i''s leads to the minimum value at V_1.

Appendix I

(i) To minimize $V = \dfrac{\Sigma w_j a_j^2}{(\Sigma a_j \bar{X}_j)^2}$ subject to $\Sigma a_j = 0$.

Suppose, the minimum value is attained at $a_1^*, a_2^*, \ldots, a_n^*$. Define a new set of values $a_1', \ldots, a_n'$ such that

$$a_i' = a_i^* + \varepsilon$$
$$a_j' = a_j^* - \varepsilon$$
$$a_r' = a_r^* \text{ for } r \neq i \text{ or } j.$$

Hence
$$\frac{\Sigma w_1 a_1'^2}{(\Sigma a_1' \bar{X}_1)^2} = \frac{\Sigma w_1 a_1^{*2} + 2\varepsilon a_i^* w_i + 2\varepsilon^2 - 2\varepsilon a_j^* w_j}{\{\Sigma a_1 \bar{X}_1 + (\bar{X}_i - \bar{X}_j)\varepsilon\}^2}$$
$$= f(\varepsilon), \text{ say.}$$

Since $f(\varepsilon)$ is minimum at $a_1^*, \ldots, a_n^*$

$$\left.\frac{df(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = 0$$

or
$$a_i^* w_i - a_j^* w_j - \frac{\Sigma w_1 a_1^*}{\Sigma a_1^* \bar{X}_1} (\bar{X}_i - \bar{X}_j) = 0.$$

or
$$a_i^* w_i - C\bar{X}_i = a_j^* w_j - C\bar{X}_j = D \text{ (say),}$$

or
$$a_i^* = \frac{D + C\bar{X}_i}{w_i}.$$

Now, adding over all i one gets,

$$C = \Sigma a_j^* = D\Sigma 1/w_i + C\Sigma \bar{X}_i/w_i$$

or
$$D = -C(\Sigma \bar{X}_i/w_i)/\Sigma(1/w_i)$$

$\therefore$
$$a_i^* \alpha p_i^*(\bar{X}_i - \bar{X}^*)$$

where
$$p_i^* = 1/w_i \text{ and } \bar{X}^* = (\Sigma \bar{X}_i/w_i)/(\Sigma 1/w_i). \qquad \text{Q.E.D.}$$

(ii) *Proof of Lemma* 1 : Define a new set of values $l_1', \ldots, l_k'$ such that

$$l_i' = l_i^* + \varepsilon$$
$$l_j' = l_j^* - \varepsilon$$
$$l_t' = l_t^* \text{ for } t \neq i \text{ or } j.$$

Then,
$$\frac{\Sigma r_t l_t^{*2} + 2\varepsilon l_i^* r_i + 2\varepsilon^2 - 2\varepsilon l_j^* r_j}{\{\Sigma r_t l_t^* + (r_i \bar{x}_i - r_j \bar{x}_j)\varepsilon\}^2} = f(\varepsilon), \text{ say}$$

Now,
$$\left.\frac{df(\varepsilon)}{d\varepsilon}\right|_{\varepsilon=0} = 0 \text{ implies}$$

$$l_i^* r_i - l_j^* r_j - C(r_i \bar{x}_i - r_j \bar{x}_j) = 0 \text{ for some constant } C.$$

or
$$l_i^* r_i - C r_i \bar{x}_i = D, \text{ say.}$$

Summing over all i one gets

$$D = -C\Sigma r_i \bar{x}_i = 0$$

Hence,
$$l_i^* = \frac{C r_i \bar{x}_i}{r_i} = C\bar{x}_i \alpha \bar{x}_i. \qquad \text{Q.E.D.}$$

Appendix II

To maximize Σl_i^2 such that

(i) $\Sigma l_i = 0$

(ii) $l_1 \leqslant \ldots \leqslant l_n$

(iii) $l_i \leqslant l_{i+1} = \ldots = l_j < l_{j+1} \Longrightarrow l_{i+1} = \ldots = l_j$

$$= \frac{x_{i+1} + \ldots + x_j}{j - i}, \text{ for } j > i.$$

Suppose $x_1^*, x_2^*, \ldots, x_n^*$, are the values of $l_1, \ldots, l_n$ attributed as stated in the theorem. $x_1', \ldots, x_n'$ are some other values of $l_1, \ldots, l_n$ satisfying (i), (ii) and (iii). Obviously $x_1^*, \ldots, x_n^*$ are satisfy (i), (ii) and (iii).

From the construction it follows that

$$x_1^* = \ldots = x_{r_1}^* = \frac{x_1+\ldots+x_{r_1}}{r_1} = \bar{x}_1 \text{ (say)}$$

$$x_{r_1+1}^* = \ldots = x_{r_1+r_2}^* = \frac{x_{r_1+1}+\ldots+x_{r_1+r_2}}{r_2} = \bar{x}_2 \text{ (say)}.$$

$$x_{r_1+\ldots+r_{k-1}+1}^* + \ldots + x_n^* = \frac{x_{r_1+\ldots+r_{k-1}+1}+\ldots+x_n}{r_k} = \bar{x}_k \text{ (say)}.$$

Similarly for x_i''s we get

$$\bar{x}_1'(l_1 \text{ times}), \ldots, \bar{x}_m' (l_m \text{ times}) \text{ such that}$$

$$l_1+\ldots+l_m = n = r_1+\ldots+r_k.$$

Let us draw the step functions for x^*'s and x''s to make the picture clear.

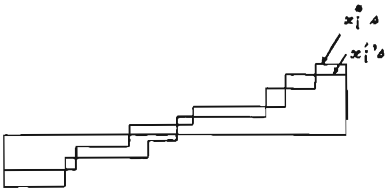

Fig. 1

Since Σl_i^2 is a convex function, any transfer from right to left will lower down it's value. Hence our job is to show that from x_i^*'s we can arrive at x_i''s by transferring values from right to left. Equivalently one may show that

$$\sum_{i=1}^{j} x_i^* \leqslant \sum_{i=1}^{j} x_i' \text{ for all } j = 1, \ldots, n.$$

It suffices to show that

$$\sum_{i=1}^{j} x_i^* \leqslant \sum_{i=1}^{j} x_i'$$

for only those j's for which

$$x_j' < x_j^* < x_{j+1}'.$$

Fix any such j. Suppose

$$j = r_1+\ldots+r_s+c; \quad c < r_{s+1}$$
$$= l_1+\ldots+l_l.$$

The condition $\sum_{i=1}^{j} x_i^* \leqslant \sum_{i=1}^{j} x_i'$

$$\iff \sum_{i=1}^{r_1+\cdots+r_s} x_i + c\bar{x}_{s+1} \leqslant \sum_{i=1}^{r_1+\cdots+r_s} x_i + x_{r_1+\cdots r_{s+1}} + \cdots + x_j$$

$$\iff \bar{x}_{s+1} \leqslant \frac{x_{r_1+\cdots r_{s+1}} + \cdots + x_j}{c}$$

which is obviously true from the way it was constructed. Q.E.D

References

Bartlett, M. S. (1949): Fitting of straight lines when both variables are subject to error. *Biometrics*, **5**, 207-212.

Dorff, M. and Gurland, J. (1961): Estimation of the parameters of a linear relation. *J. Roy. Stat. Soc.*, B, **23**, 160-170.

Durbin, J. (1954): Errors in variables. *Review of the Int. Stat. Inst.* **22**, 23-32.

Gibson, W. M. and Jowett, G. H. (1957): Three-group regression analysis Part 1. *Applied Statistics*, **6**, 114-122.

Goldberger, A. S. (1964): *Econometric Theory*, Wiley, New York

Housner, G. W. and Brennan, J. F. (1948): Estimation of linear trends, *Ann. Math. Stat.*, **19**, 380-388.

Katona, G., Klein, L. R., Lansing, J. B. and Morgan, J. N. (1954): Contributions of survey methods to Econometrics, Calumbia, Referred to in Lancaster (1968).

Keifer, J. and Wolfwitz, J. (1956): Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters *Ann. Math. Stat.*, **27**, 887-906.

Lancaster, T. (1968): Grouping estimators on Heteroscedastic data. *J. Am., Stat. Assoc.*, **63**, 182-191.

Leamer, E. E. (1978): Least squares versus IV's estimation in simple EVM. *Econometrica*, **46**, 961-968.

Nair, K. R. and Banerjee, K. S. (1943): A Note on fitting of straight lines if both variables are subject to error. *Sankhyā*, **6**, 331.

Neyman, J. (1951): Existence of consistence estimate of the directional parameter in a linear structural relation between two variables. *Ann. Math. Stat.*, **22**, 352-361.

Nair, K. R. and Shrivastava, M. P. (1942): On a simple method of curve fitting, *Sankhyā*, **6**, 121-132.

Pal, M. (1982): Some observations on identification in EVM's. *Discussion Paper* No. ERU/D/5/82, Indian Statistical Institute, Calcutta.

Pal, M. and Bhaumik, M. (1981): A note on Bartlett's method of grouping in regression analysis, *Sankhyā* B, **43**, 399-404.

Reiersol, O. (1950): Identifiability of a linear relation between variables which are subject to error, *Econometrica*, **18**, 375-389.

Richardson, D. H. and De-Min Wu (1970): Least squares and grouping method estimators in the EVM. *J. Am. Stat. Assoc.*, **65**, 724-748.

SIL, S., BHAUMIK, M. and PAL, M. (1981): A note on Bartlett's grouping estimators in an hetero-scedastic model. *Tech. Report No.* ERU/2/81, Indian Statistical Institute, Calcutta.

THEIL, H. (1950): A rank-invariant method of linear and polynomial regression analysis. *Indagationes Mathematicae*, **12**, 85-91.

THEIL, H. and J. VAN YZEREN (1956): On the efficiency of Wald's method of fitting straight lines. *Review of the Int. Stat. Inst.*, **24**, 17-26.

WALD, A. (1940): The fitting of straight lines if both variables are subject to error. *Ann. Math. Stat.*, **11**, 284-300.

WOLFOWITZ, J. (1952): Consistent estimators of the parameters of a linear structural relation, *Skandinavisk Aktuarietidskrift*, **35**, 132-151.

Paper received : November, 1982.