TR/15/85 May 1985

Moment Properties of Estimators For A Type 1 Extreme Value Regression Model

by

A.A. Haddow and D.H. Young

Moment Properties Of Estimators For A Type 1 Extreme Value Regression Model

by

A. A. Haddow and D .H. Young

Summary

A regression model is considered in which the response variable has a type 1 extreme value distribution for smallest values. Small sample moment properties of estimators of the regression coefficients and scale parameter, based on maximum likelihood, ordinary least squares and best linear unbiased estimation using order statistics for grouped data, are presented, and evaluated, for the case of a single explanatory variable. Variance efficiency results are compared with asymptotic values.

Contents

1. Introduction

Consider the classical multiple linear regression model

$$Y_i = u_i = \varepsilon_i = \underset{\sim}{x}_i' \underset{\sim}{\beta} = \varepsilon_i \quad i = 1,\ldots,n \tag{1.1}$$

where $\underset{\sim}{x}_i' = (1, x_{.i1}, \ldots, x)$, $\underset{\sim}{\beta}' = (\beta_0, \beta_1, \ldots, \beta_k)$, the values $x_{i1}, \ldots, x_{ik}$ representing observations on k non-random explanatory variables for the ith individual. We shall assume that the true residuals $\{\varepsilon_i\}$ are independently and identically distributed with

$$E(\varepsilon_i) = 0, \quad var(\varepsilon_i) = \sigma^2, \quad i = 1,\ldots,n \tag{1.2}$$

If the distribution of the $\{\varepsilon_i.\}$ is not specified, the estimators of the regression coefficients are usually determined by ordinary least squares (OLS) and are given by the value of which minimises $(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta})'(\underset{\sim}{Y} - \underset{\sim}{X}\underset{\sim}{\beta})$,

where $\underset{\sim}{Y}' = (Y_1, \ldots, Y_n)$ and $\underset{\sim}{X}$ is the design matrix. Assuming that $\underset{\sim}{X}$ is of rank k+1, the OLS estimator is given by

$$\tilde{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y} \tag{1.3}$$

with

$$E(\tilde{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}, \quad cov(\tilde{\underset{\sim}{\beta}}) = \sigma^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1} \tag{1.4}$$

An unbiased estimator of σ^2 is given by the LS estimator

$$\tilde{\sigma}^2 = (n-k-1)(\underset{\sim}{Y} - \underset{\sim}{X}\tilde{\underset{\sim}{\beta}})'(\underset{\sim}{Y} - \underset{\sim}{X}\tilde{\underset{\sim}{\beta}}). \tag{1.5}$$

The justification for using least squares when the distribution of the $\{\varepsilon_i\}$ is unknown, is that the OLS estimators $\{\tilde{\beta}_r\}$ have minimum variance among estimators that are linear combinations of the $\{Y_i.\}$. Further the estimator $\tilde{\sigma}^2$ has minimum variance within the class of estimators which are quadratic forms in the $\{Y_i\}$. When the distribution of the $\{\varepsilon_i\}$ is specified, estimates of $\underset{\sim}{\beta}$ and σ^2 may be found by maximum likelihood (ML). The ML estimator of $\underset{\sim}{\beta}$, denoted by $\hat{\underset{\sim}{\beta}}$ will be a non-linear function of the $\{Y_i\}$ except for the normal case when $\hat{\underset{\sim}{\beta}}$ and $\tilde{\underset{\sim}{\beta}}$ are the same. The ML estimator $\hat{\sigma}^2$ of σ^2 is equal to for (n-2) $\tilde{\sigma}^2$ the normal case but in general is not a quadratic form in the $\{Y_i\}$. Calculation of the ML estimates usually requires an iterative solution but the extra computational effort is often worthwhile as appreciable loss of variance efficiency can occur if LS is used for non-normal cases.

In this report, we consider the case when the $\{Y_i\}$ have a type 1 extreme value distribution for smallest values with p. d. f.

$$f_{Y_i}(y) = \frac{1}{\theta} \exp\left\{ \frac{y - \underset{\sim}{x}_i' \underset{\sim}{\beta}}{\theta} - \gamma - \exp\left(\frac{y - \underset{\sim}{x}_i' \underset{\sim}{\beta}}{\theta} - \gamma \right) \right\}, \quad -\infty < y < \infty \qquad (1.6)$$

where $\gamma = 0.577216$ is Euler's constant and $\theta > 0$ is the common scale parameter . The mean and variance of the distribution are

$$E(Y_i) = \underset{\sim}{x}_i' \underset{\sim}{\beta} , \quad \text{var}(Y_i) = \frac{1}{6} \pi^2 \theta^2 \qquad (1.7)$$

respectively. The distribution is nonsymmetrical with skewness and kurtosis coefficients given by

$$\gamma_1(Y_i) = 1.29857 \qquad \gamma_2(Y_i) = 2.4. \qquad (1 .8)$$

The type 1 extreme-value (EV) distribution with p.d.f. given by (1.6) arises under certain conditions as the 1imiting distribution of the smallest value of a large number of independent and identically distributed random variables. It therefore often provides a useful approximation to the distribution of system life or system breaking strength when a system contains a large number of components and 'failure' occurs as soon as one component fails.

As is well-known, the type 1 EV distribution is closely related to the Weibull distribution. Thus the c.d.f. of Y_i. is

$$F_{Y_i}(y) = 1 - \exp\left\{ - \exp\left(\frac{y - \underset{\sim}{x}_i' \underset{\sim}{\beta}}{\theta} - \gamma \right) \right\} . \qquad (1.9)$$

If we put

$$W_i = \exp(Y_i) , \qquad i = 1,\ldots,n \qquad (1.10)$$

the random variable W. has the Weibull distribution with c.d.f.

$$F_{W_i}(w) = 1 - \exp\{ - (w/\delta)^{1/\theta}\} , 0 < w < \infty \qquad (1.11)$$

where the scale parameter 5 is

$$\delta = \exp\{ - (\underset{\sim}{x}_i' \underset{\sim}{\beta} + \gamma\theta) \} . \qquad (1.12)$$

We have

$$E(w_i) = \exp(\underset{\sim}{x}_i' \underset{\sim}{\beta}_w) , \ i = 1 ,. .. ,n \qquad (1.13)$$

where

$$\beta_{W_0} = \beta_o + \delta\theta + \log\Gamma(1+\theta) , \qquad \beta_{wr} = \beta_r , \qquad r = 1,...,k . \tag{1.14}$$

It follows that the regression model for the $\{Y_i\}$ with additive model $\mu_{i\cdot} = \underset{\sim}{x}_i' \underset{\sim}{\beta}$ for the means is equivalent to that based on the $\{W.\}$ with an multiplicative exponential model for the means. No special treatment is therefore required for the Weibull model.

When no explanatory variables are present, many investigations have been made to assess and compare the properties of various estimators for the location and scale parameters of the type 1 EV distribution. A useful survey is given by Mann (1968). Lawress (1982) discussed statistical inference procedures for the type 1 EV regression model and also gives asymptotic efficiency results for LS estimation relative to ML estimation. However, little work appears to have been done to assess the small sample properties of the estimators in the regression case.

This report focuses attention on the moment properties of estimators for the parameters in the regression model. In section 2, four computational procedures for determining the ML estimators are described and some findings relating to their computational efficiency are given. Approximations to the biases and variances of the ML estimators are given in section 3 and evaluated by simulation for the case of a single explanatory variable. The OLS estimators are considered in section 4 and in section 5 we discuss the best linear unbiased estimators (BLUE'S) based on order statistics when grouped data are available. Finally, in section 6 small sample variance efficiency results as obtained by simulation are compared with asymptotic results.

2. Computational Procedures For Maximum Likelihood Estimation

In this section, we describe four computational methods for determining the maximum likelihood estimates. The first two methods use the Newton-Raphson and Fisher's scoring approach respectively, while the last two methods are designed to facilitate the use of the statistical package GLIM which provides ML fits of generalised linear models.

We first present results for the first and second order derivatives of the logarithm of the density given by (1.6). For notational convenience we set

$$z_i = \theta^{-1}(y_i - \underset{\sim}{x}_i') - \gamma \,, \quad i = 1, \ldots, n \tag{2.1}$$

and we have

$$\log f_{Y_i}(y_i) = \theta^{-1} \exp(z_i - e^{z_i}) \tag{2.2}$$

with

$$\partial z_i/\partial \beta_r = -x_{ir}/\theta \,, \quad \partial z_i./\partial\theta = -(z_i + \gamma)/\theta \,.$$

We set

$$U_r^{(i)} = \frac{\partial \log f_{Y_i}(y_i)}{\partial \beta_r} \,, \quad U_\theta^{(i)} = \frac{\partial \log f_{Y_i}(y_i)}{\partial \theta} \tag{2.3}$$

$$V_{rs}^{(i)} = \frac{\partial^2 \log f_{Y_i}(y_i)}{\partial \beta_r \, \partial \beta_s} \,, \quad V_{r\theta}^{(i)} = \frac{\partial^2 \log f_{Y_i}(y_i)}{\partial \beta_r \, \partial \theta} \,, V_{\theta\theta}^{(i)} = \frac{\partial^2 \log f_{Y_i}(y_i)}{\partial \theta^2} \tag{2.4}$$

for r,s = 0,1,...,k, and a straightforward calculation gives

$$U_r^{(i)} = x_{ir}\, \theta^{-1}(e^{z_i} - 1) \,, U_\theta^{(i)} = \theta^{-1}\{(z_i + \gamma)(e^{z_i} - 1) - 1\} \tag{2.5}$$

$$V_{rs}^{(i)} = x_{ir}\, x_{is}\, \theta^{-2} e^{z_i} \,, \quad V_{r\theta}^{(i)} = -\, x_{ir}\, \theta^{-2}\{e^{z_i} - 1 + e^{z_i}(z_i + \gamma)\} \tag{2.6}$$

$$V_{\theta\theta}^{(i)} = -\,\theta^{-2}\{(z_i + \gamma)^2 e^{z_i} + 2(z_i + \gamma)\{e^{z_i} - 1) - 1)\} \tag{2.7}$$

If $L(\underset{\sim}{\beta}, \theta) = \sum_i \log f_{Y_i}(y_i)$ denotes the log-likelihood we have

$$\frac{\partial L(\underset{\sim}{\beta}, \theta)}{\partial \beta_r} = \sum_i U_r^{(i)} \,, \qquad \frac{\partial L(\underset{\sim}{\beta}, \theta)}{\partial \theta\,.} = \sum_i U_0^{(i)} \,. \tag{2.8}$$

Thus the ML estimates are given by the solution of the k+2 equations

$$\sum_i x_{ir}(e^{\hat{z}_i} - 1) = 0 \,, \qquad \sum_i (\hat{z}_i + \gamma)(e^{\hat{z}_i} - 1) = n \tag{2.9}$$

for r = 0,1,...,k, where

$$\hat{z}_i = \hat{\theta}^{-1}(y_i - \underset{\sim}{x}_i' \underset{\sim}{\beta}) - \gamma \,, \qquad i = 1, \ldots, n \tag{2.10}$$

The second order derivatives of the log-likelihood are given by

$$\frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_r\partial\beta_s} = \sum_i V_{rs}^{(i)} , \qquad \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_r\partial\theta} = \sum_i V_{r\theta}^{(i)} , . \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\theta^2} = \sum_i V_{\theta\theta}^{(i)} . \tag{2.11}$$

(i) Newton-Raphson Method

Using the Newton-Raphson approach to find an iterative solution to the likelihood equations, we set

$$\underset{(k+2)x1}{\underset{\sim}{D}_1^{(\ell)}} = \begin{bmatrix} \partial L(\underset{\sim}{\beta},\theta) / \partial\beta_0 \\ \partial L(\underset{\sim}{\beta},\theta) / \partial\beta_k \\ \partial L(\underset{\sim}{\beta},\theta) / \partial\theta \end{bmatrix}_{\underset{\sim}{\beta} = \underset{\sim}{\hat{\beta}}^{(\ell)},\ \theta = \hat{\theta}^{(\ell)}} \tag{2.12}$$

$$\underset{(k+2)x(k+2)}{\underset{\sim}{D}_1^{(\ell)}} = - \begin{bmatrix} \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta} & \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_0\partial\beta_1} & \cdots & \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_0\partial\beta_k} & \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_0\partial\beta_\theta} \\ \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_0\partial\beta_k} & \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_1\partial\beta_k} & \cdots & \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_k^2} & \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_k\partial\theta} \\ \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_0\partial\theta} & \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_1\partial\theta} & \cdots & \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\beta_k\partial\theta} & \frac{\partial^2 L(\underset{\sim}{\beta},\theta)}{\partial\theta^2} \end{bmatrix}_{\underset{\sim}{\beta} = \underset{\sim}{\hat{\beta}}^{(\ell)},\ \theta = \hat{\theta}^{(\ell)}} ,$$

(2.13)

where $\underset{\sim}{\hat{\beta}}^{(\ell)}, \hat{\theta}^{(\ell)}$ denotes the approximations to $\underset{\sim}{\hat{\beta}}$ and $\hat{\theta}$ at the stage of iteration. The new approximations are given by

$$\begin{bmatrix} \underset{\sim}{\hat{\beta}}^{(\ell+1)} \\ \hat{\theta}^{(\ell+1)} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{\hat{\beta}}^{(\ell)} \\ \hat{\theta}^{(\ell)} \end{bmatrix} - \{ \underset{\sim}{D}_2^{(\ell)} \}^{-1} \underset{\sim}{D}_1^{(\ell)} . \tag{2.14}$$

(ii) Fisher's Scoring Method

A simple and well-known modification to the Newton-Raphson approach is to use Fisher's scoring method in which the elements in $\underset{\sim}{D}_2$ are replaced at each stage by the current estimates of their expected values. We first need some simple expectation results for the random variables $Z_i = \theta^{-1} (Y_i - x_i' \underset{\sim}{\beta}) - \gamma$, i = 1,.......n, which are independently and identically distributed with p.d.f.

$$f_z(z) = \exp(z - e^z) , -\infty < z < \infty . \tag{2.15}$$

The moment generating function of Z is

$$M_Z(t) = \int_{-\infty}^{\infty} \exp\{z(t+1) - e^Z\}dt = \Gamma(1+t) \;, \qquad (2.16)$$

giving

$$E(Z^r e^{Zt}) = d^r M_z(t)/dt^r = d^r\,\Gamma(1+t)/dt^r \qquad (2.17)$$

Using the results that

$$\frac{\Gamma'(x)}{\Gamma(x)} = \frac{d\log\,\Gamma(x)}{dx} \;,\; \frac{\Gamma''(x)}{\Gamma(x)} = \frac{d^2\log\,\Gamma(x)}{dx^2} + \left\{\frac{d\log\,\Gamma(x)}{sx}\right\}^2$$

where dlogr(x)/dx and $d^2\log\Gamma(x)/dx^2$ are the digamma and trigamma functions respectively, we obtain

$$E(e^Z) = \Gamma(2) = 1 \;,\; E(Ze^Z) = \Gamma'(2) = 0.422784 \;,\; E(Z^2e^Z) = \Gamma''(2) = 0.823680 \;. \qquad (2.18)$$

Setting

$$I_{rs} = E(-\sum_i V_{rs}^{(i)} \;,\; I_{r\theta} = E(-\sum_i V_{r\theta}^{(i)}) \;,\; I_{\theta\theta} = E(-\sum_i V_{\theta\theta}^{(i)}) \qquad (2.19)$$

and using (2.6), (2.7) and (2.18), the elements in the information matrix are given by

$$I_{rs} = \theta^{-2}\sum_i x_{ir}x_{is} \;,\; I_{r\theta} = \theta^{-2}\sum_i x_{ir} \;,\; I_{\theta\theta} = 2.644934n\,\theta^{-2} \qquad (2.20)$$

for r,s, = 0,1,...,k. These elements replace those in $\underset{\sim}{D}_2$ and are estimated at the ℓth stage using the current approximation $\hat{\theta}^{(\ell)}$ to $\hat{\theta}$, the elements being independent of $\underset{\sim}{\beta}$ Iteration continues using the scheme given by (2.14) until satisfactory convergence is obtained.

The calculation of the ML estimates by the Newton-Raphson or Fisher's scoring method is straightforward using a matrix inversion subroutine. For large scale simulation work in GLIM, we have examined two other computational approaches for finding the ML estimates, which we now describe.

(iii) Two-Stage Method

The log-likelihood under the type 1 EV regression model is

$$L(\underset{\sim}{\beta}\,\theta) = -\,n\log\theta - n\gamma + \theta^{-1}\sum_i (y_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}) - \sum_i \exp\left(\frac{y_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\theta} - \gamma\right). \qquad (2.21)$$

Consider a fixed value of 0, say ℓ. Putting

$$\beta_{*0} = \gamma + \theta_0^{-1},\beta_0, \qquad \beta_{*r} = \theta_0^{-1}\beta_r, \qquad r = 1,\ldots,k$$

$$y_{*i} = \exp(y_i/\theta_0),\; u_{*i} = \exp(\underset{\sim}{x}_i'\underset{\sim}{\beta}_*), \qquad i = 1,\ldots, n$$

we have

$$L(\underset{\sim}{\beta}_*,\theta_0) = \sum_i \log y_{*i} - n\log\theta_0 - \sum_i \log u_{*i} - \sum_i (y_{*i}/u_{*i}) \qquad (2.22)$$

Maximising $L(\underset{\sim}{\beta},\theta_0)$ with respect to $\underset{\sim}{\beta}$ is therefore equivalent to maximising

$$L(\underset{\sim}{\beta}_*,\theta_0) = -\sum_i \log u_{*i} + y_{*i}/u_{*i}) \qquad (2.23)$$

with respect to $\underset{\sim}{\beta}_*$. $L_1(\underset{\sim}{\beta}_*,\theta_0)$ is the log-likelihood treating the $\{y_{*i}.\}$ as observations on independent exponentially distributed observations with means $\{\mu_{*i}\}$, where $\log\mu_{*i}\ \underset{\sim}{x}_i'\ \underset{\sim}{\beta}_*$ The value $\hat{\underset{\sim}{\beta}}_* = \hat{\underset{\sim}{\beta}}_*(\theta_0)$ which maximises $L_1(\underset{\sim}{\beta}_*,\theta_0)$ may therefore be obtained using GLIM by specifying an exponential error distribution and a logarithmic link function.

Treating $\underset{\sim}{\beta}$ as known, the value $\hat{\theta} = \hat{\theta}(\underset{\sim}{\beta})$ of 8 which maximises $L(\underset{\sim}{\beta},\theta)$ is from (2.9) a solution of the equation $g\ (\hat{\theta},\underset{\sim}{\beta}) = 0$ where

$$g(\hat{\theta},\underset{\sim}{\beta}) = \sum_i (y_i - \underset{\sim}{x}_i'\underset{\sim}{\beta})\left\{ xp\left(\frac{y_i - \underset{\sim}{x}_i'\underset{\sim}{\beta}}{\theta} - \gamma\right) - 1\right\} - n\hat{\theta}. \qquad (2.24)$$

Using a Newton-Raphson approach to find an iterative solution for 0, if θ ℓ. denotes the ℓ th approximation to $\hat{\theta}$, we have

$$\hat{\theta}_{\ell+1} = \hat{\theta}_\ell - g(\hat{\theta}_\ell,\underset{\sim}{\beta})/g(\hat{\theta}_\ell,\underset{\sim}{\beta}), \qquad \ell = 1,2,\ldots. \qquad (2.25)$$

Since in practice both $\underset{\sim}{\beta}$ and are unknown, the following two—stage iteration method may be used. If $\hat{\underset{\sim}{\beta}}, \hat{\theta}_1$ denote preliminary approximations to $\hat{\underset{\sim}{\beta}}$ and $\hat{\theta}$ respectively, the new approximation $_2$ is given by $\hat{\theta}_1 = g(\hat{\theta}_1,\hat{\underset{\sim}{\beta}}_1)/g'(\hat{\theta}_1,\hat{\underset{\sim}{\beta}}_1)$. The transformed observations $y_{*i} = \exp(y_i/\hat{\theta}_2)$ are then used in a GLIM fit to obtain $\hat{\underset{\sim}{\beta}}_{*2}$ and hence $\hat{\underset{\sim}{\beta}}_2$. The same steps are used repeatedly until satisfactory convergence is obtained.

(iv) The Roger/Peacock Method

An alternative method specially designed for work in GLIM was proposed by Roger and Peacock (1982). Their method copes with censored observations and has the advantage over the preceding method of providing the estimated asymptotic covariance matrix of the ML estimates directly from the output of a GLIM fit. To apply their method, we put

$$\alpha = 1/\theta \ , \ \mu_i^* = \exp\left(\frac{y_i - \underset{\sim}{x}_i' \underset{\sim}{\beta}}{\theta} - \gamma \right), \quad i = 1 ,...,n . \tag{2.26}$$

We have

$$\log \ \mu_i^* = \alpha \ y_i + \underset{\sim}{x}_i^* \underset{\sim}{\beta}^* , \qquad i = 1 ,.. , ,n \tag{2.27}$$

where $\underset{\sim}{\beta}^{*\prime} = \{ - (\alpha \beta_0 + \gamma) , - \alpha \beta_1 ,...,-\alpha \beta_k, \}$ and the log-likelihood may be written as

$$L(\underset{\sim}{\beta}^* ,\alpha) = \sum_{i=1}^{n} (\log \mu_i^* - \mu_i^*) + n \log \alpha \ . \tag{2.28}$$

Let $\hat{\alpha}$, $\hat{\underset{\sim}{\beta}}^*$ denote the ML estimates of α and. $\underset{\sim}{\beta}^*$ To use GLIM, we may proceed as follows. Suppose that we have n independent Poisson random variables $Z_1 ,...,Z_n$ with means $\mu_1^* ,,..,\mu^*_n$ satisfying (2.27), and an independent binomial random variable Z_{n+1} based on n trials and 'success[1] probability α: The log-likelihood for realised values z_1 , ,,.,z_n and z_{n+1} is

$$L_1(\underset{\sim}{\beta}^* ,\alpha) = \sum_{i=1}^{n} \log\left(\frac{e^{-\mu_i^*} \mu_i^{*z_i}}{z_i} \right) + \log\left\{ \binom{n}{z_{n+1}} \alpha^{z_{n+1}} (1-\alpha)^{n-z_{n+1}} \right\} \tag{2.29}$$

$$= \text{constant} + \sum_{i=1}^{n} (z_i \log\mu_i^* - \mu_i^*) + z_{n+1} \log \alpha + (n - z_{n+1}) \log (1-\alpha) \ .$$

For the specific realisation $z_i = 1$, i = 1 ,. . . ,n and $z_{n+1} = n$, we have $L_1(\underset{\sim}{\beta}^* ,\alpha) = L(\underset{\sim}{\beta}^* ,\alpha)$ + constant. Hence maximisation of $L(\underset{\sim}{\beta}^* ,\alpha)$ is equivalent to maximisation of the log-likelihood based on the random variables Z_1.,...,Z_n and Z_{n+1} , , when the realised values are z.=1, i = 1 , . . . ,n for the Poisson variables and z_{n+1} = n for the binomial variable. Roger and Peacock give a GLIM program, containing five macros including four macros for fitting a user-defined model, which may be used to find the ML estimates.

Some idea of the variations in computer time required to determine the ML estimates can be obtained from table 1 which shows the computer processing units (CPU) for each of the four methods of estimation for 10, 20 and 50 sets of sample data generated by simulation for n = 25. The results indicate that the Fisher-scoring method requires less time than the Newton-Raphson method and gives considerable savings over the two-stage method. Comparisons with the Roger/Peacock method are more difficult to make, as the stopping rule for convergence is not controlled by the user but is implemented in the GLIM system. Further, a run with 100 sets of data showed that in a few cases the Roger/Peacock method failed to produce a solution because negative values were produced for the total deviance.

Table

CPU usage for four methods for obtaining ML estimates

Run Size	Roger/Peacock	Fisher scoring	Newton-Raphson	Two Stage
10	59 sec	74 sec	83 sec	115 sec
20	108 sec	150 sec	151 sec	216 sec
50	255 sec	334 sec	458 sec	626 sec

3. Moment Properties Of The Maximum Likelihood Estimators

The exact moments of the ML estimators are unknown, but approximations to their biases, variances and covariances correct to $0(n^{-1})$ can be obtained by standard methods.

Without loss of generality, we shall assume that the values of the x's are centred such that $\sum_{i=1}^{n} x_{ir} = 0$ for r = 1,...,k. In this case, use of (2.20) shows that theinformation matrix may be written as

$$\underset{\sim}{I} = \theta^{-2}\begin{pmatrix} \underset{\sim 1}{I} & \underset{\sim}{O} \\ \underset{\sim}{O} & \underset{\sim 2}{I} \end{pmatrix} \qquad (3.1)$$

where $\underset{\sim 1}{I}$ refers to θ and β_0 irefers $\underset{\sim 2}{I}$ to $\beta_1 ..., \beta_k$ and

$$\underset{\sim 1}{I} = \begin{pmatrix} 2.64493n & n \\ -n & n \end{pmatrix} \quad \underset{\sim 2}{I} = \begin{bmatrix} \sum_i x_{i1}^2 & \sum_i x_{i1}x_{i2} & \cdots & \sum_i x_{i1}x_{ik} \\ \sum_i x_{i1}x_{i2} & \sum_i x_{i2}^2 & \cdots & \sum_i x_{i2}x_{ik} \\ \sum_i x_{i1}x_{ik} & \sum_i x_{i2}x_{ik} & & \sum_i x_{ik}^2 \end{bmatrix}. \qquad (3.2)$$

Since

$$\underset{\sim}{I}^{-1} = n^{-1}\begin{pmatrix} 0..6079 & -0.6079 \\ -0.6079 & 0..6079 \end{pmatrix}$$

we have the standard approximations

$$\operatorname{var}(\hat{\beta}_0) \approx \frac{1.6079\,\theta^2}{n}\ ,\ \operatorname{var}(\hat{\theta}) \approx \frac{0.6079\,\theta^2}{n}\ ,\ \operatorname{cov}(\hat{\beta}_0\ ,\hat{\theta}) \approx \frac{-\,0.6079\,\theta^2}{n}. \tag{3.3}$$

The approximate covariance matrix for $\underset{\sim 1}{\hat{\beta}}' = (\hat{\beta}_1,....,\hat{\beta}_k)$ is =

$$\operatorname{Cov}(\underset{\sim 1}{\hat{\beta}}) \approx \theta^2\ ((\sum_i\ x_{ir}\, x_{is}\,))^{-1}\ . \tag{3.4}$$

To obtain the approximate biases, we first need the third order derivatives of the logarithm of the density given by (1.6). Setting

$$W^{(i)}_{rst} = \frac{\partial^3 \log f_{Y_i}(y_i)}{\partial\beta_r\partial\beta_s\partial\beta_t}, W^{(i)}_{rs\theta} = \frac{\partial^3 \log f_{Y_i}(y_i)}{\partial\beta_r\partial\beta_s\partial\theta}, W^{(i)}_{r\theta\theta} = \frac{\partial^3 \log f_{Y_i}(y_i)}{\partial\beta_r\partial\theta^2}, W^{(i)}_{\theta\theta\theta} = \frac{\partial^3 \log f_{Y_i}(y_i)}{\partial\theta^3} \tag{3.5}$$

and using the second order derivatives given in (2.6) and (2.7) we obtain

$$W^{(i)}_{rst} = \theta^{-3}\ x_{ir}\, x_{is}\, x_{it}\, e^{z_i}\quad ,\quad W^{(i)}_{rs\,\theta} = \theta^{-3} x_{ir}\, x_{is}\, e^{z_i}(z_i + \gamma + 2) \tag{3.6}$$

$$W^{(i)}_{r\theta\theta} = \theta^{-3}\ x_{ir}\, [\, e^{z_i}\ \{z_i. + \gamma\,)^2 + 4\,(z_i + \gamma) + 2\,\}\ -\ 2] \tag{3.7}$$

$$W^{(i)}_{\theta\theta\theta} = \theta^{-3}\ [\, e^{z_i}\ x\{(z_i + \gamma)^3 + 6\,(z_i + \gamma)^2 + 6(z_i + \gamma)\,\}\ -6(z_i + \gamma)\ -2]\ . \tag{3.8}$$

We set

$$K_{rst} = E(\sum_i\ W^{(i)}_{rst}),\ K_{rs\theta}\ E(\sum_i\ W^{(i)}_{rs\theta}\ ,\ )K_{r\theta\theta} = E(\sum_i\ W^{(i)}_{r\theta\theta})\ ,\ K_{\theta\theta\theta} = E(\sum_i\ W^{(i)}_{\theta\theta\theta}) \tag{3.9}$$

Using (2.17) we obtain

$$E\{(Z + \gamma)\exp(Z:)\}\ -\ \Gamma'(2) + \gamma = 1 \tag{3.10}$$

$$E\{(Z + \gamma)^2\exp(Z)\} = \Gamma''(2) + 2\gamma\Gamma'(2) + \gamma^2 = 1.64494 \tag{3.11}$$

$$E\{(Z+\gamma)^3\exp(Z)\}\ -\ \Gamma'''(2) + 3\gamma\Gamma''(2) + 3\gamma^{2\Gamma'}(2) + \gamma^3 = 2.53070 \tag{3.12}$$

and hence

$$K_{rst} = \theta^{-3} \sum_i x_{ir} x_{is} x_{it} , \quad K_{rs\theta} = 3\theta^{-3} \sum_i x_{ir} x_{is} \tag{3.13}$$

$$K_{r\theta\theta} = 5.64494 \ \theta^{-3} \sum_i x_{ir} , \quad K_{\theta\theta\theta} = 16.4.003 \ \theta^{-3} n \tag{3.14}$$

Finally, we require the values of the quantities

$$J_{r,st} = E(\sum_i U_r^{(i)} V_{st}^{(i)} = -\theta^{-3} \sum_i x_{ir} x_{is} x_{it} \ E\left\{e^Z (e^Z - 1)\right\} \tag{3.15}$$

$$J_{r,s\theta} = E(\sum_i U_r^{(i)} V_{s\theta}^{(i)} = -\theta^{-3} \sum_i x_{ir} x_{is} \ E\left[(e^Z - 1)\left\{e^Z - 1 + e^Z (Z + \gamma)\right\}\right] \tag{3.16}$$

$$J_{r,\theta\theta} = E(\sum_i U_r^{(i)} V_{\theta\theta}^{(i)}$$

$$= -\theta^{-3} \sum_i x_{ir} E\left[(e^Z - 1)\left\{(Z + \gamma)^2 e^Z + 2(Z + \gamma) e^Z - 1) - 1\right\}\right] \tag{3.17}$$

$$J_{\theta,st} = E(\sum_i U_r^{(i)} V_{st}^{(i)} = -\theta^{-3} \sum_i x_{is} x_{it} E\left[e^Z\left\{(Z + \gamma)(e^Z - 1) - 1\right\}\right] \tag{3.18}$$

$$J_{\theta,s\theta} = E(\sum_i U_\theta^{(i)} V_{s\theta}^{(i)}$$

$$= -\theta^{-3} \sum_i x_{is} E\left[\left\{(Z + \gamma)(e^Z - 1) - 1\right\} \left\{e^Z - 1 + (Z + \gamma)(e^Z - 1) e^Z\right\}\right] \tag{3.19}$$

$$J_{\theta,\theta\theta} = E(\sum_i U_\theta^{(i)} V_{\theta\theta}^{(i)})$$

$$= -n\theta^{-3} E\left[\left\{(Z + \gamma)(e^Z - 1) - 1\right\} \left\{(Z + \gamma)^2 e^Z + 2(Z + \gamma)(e^Z - 1) - 1\right\}\right] \tag{3.20}$$

Using (2.17) we have $E\{Z^r \exp(2z)\} = d^r (x)/dx^r$ evaluated at $x = 3$ which gives

$$E(e^{2Z}) = r(3) = 2 , \ E(Ze^{2Z}) - T'(3) = 1.84557$$

$$E(Z^2 e^{2Z}) = \Gamma''(3) = 2.49293 , \ E(Z^3 e^{2Z}) = \Gamma'''(3) = 3.44997 .$$

Use of these results in (3.15) to(3.20) gives after simplification

$$J_{r,st} = -\theta^{-3} \sum_i x_{ir} x_{is} x_{it} , \quad J_{r,s\theta} = -3\theta^{-3} \sum_i x_{ir} x_{is} \tag{3.21}$$

$$J_{r,\theta\theta} = -5.64493 \ \theta^{-3} \sum_i x_{ir} , \quad J_{\theta,st} = -\theta^{-3} \sum_i x_{is} x_{it} \tag{3.22}$$

$$J_{r,st} = -3.64493 \ \theta^{-3} \sum_i x_{is} , \quad J_{\theta,\theta\theta} = -11.11040 \ \theta^{-3} n. \tag{3.21}$$

If $b_r = E(\hat{\beta}_r) - \beta_r$ and $b_\theta = E(\hat{\theta}) - \theta$ denote the biases of $\hat{\beta}_r$ and θ, respectively, r = 0,1,..., k, then from Cox and Snell (1968) the bias approximations correct to 0(n) are

$$b_r = \frac{1}{-2} \sum_s \sum_t \sum_u I^{rs} I^{tu} (K_{stu} + 2 J_{t,su}) \tag{3.24}$$

$$b_\theta = \frac{1}{-2} \sum_s \sum_t \sum_u I^{\theta s} I^{tu} (K_{stu} + 2 J_{t,su}) \tag{3.25}$$

where the summations are over s,t ,u = 0,1,...,k, θ and I^{rs}, $I^{\theta S}$ etc. denote elements in the inverse of the information matrix.

Simple expressions for the biases can be obtained straightforwardly for the case when the x's satisfy the conditions

$$\sum_{i=1}^{n} x_{ir} x_{is} \; 0 \text{ for all } r \neq s = 0,1,...,k . \tag{3.26}$$

The elements in the inverse of the information matrix are then

$$I^{00} = 1.6079 n^{-1} \theta^2 , \; I^{\theta\theta} = 0.6079 n^{-1} \theta^2 , \; I^{rr} = \theta^2 / \Sigma_i x_{ir}.$$

$$\text{for } r = 1 ,... ,k \tag{3.27}$$

$$I^{\theta\theta} = -0.6079 n^{-1} \theta, \; I^{r0} = I^{r\theta} = I^{rs} = 0 \;\text{ for }\; r \neq s - 1 , . . . ,k \tag{3.28}$$

with the regression coefficient estimators $\hat{\beta}_r$, r = 0,1,...,k being pair-wise asymptotically uncorrelated. We may write

$$b_c = \frac{1}{2} (I^{00} A_0 + I^{0\theta} A_\theta) , \; b_\theta = \frac{1}{2} (I^{\theta 0} A_0 + I^{\theta 0} A_\theta) , = \frac{1}{2} I^{rr} A_r \tag{3.29}$$

for r = 1,...,k, where

$$A_r = \sum_t \sum_u I^{tu} (K_{rtu} + 2J_{t,ru}) , \; A_\theta = \sum_t \sum_u I^{tu} (K_{\theta tu} + 2J_{t,\theta u}) \tag{3.30}$$

A straightforward calculation gives

$$A_0 = -(k + 1.3921)\theta^{-1} , \; A_r = -\theta^{-1} \sum_{t=1}^{k} \left(\sum_i X_{ir} X_{it}^2 / \sum_i X_{it}^2 \right) , \; A_\theta = -(3k + 3.9304)\theta^{-1} \tag{3.31}$$

from which we obtain

$$b_0 = \theta n^{-1} (0.1080k + 0.0754) , \tag{3.32}$$

$$b_r = -\frac{1}{2}\frac{\theta}{\sum_i X_{ir}^2}\sum_{t=1}^{k}\left(\sum_i X_{ir}X_{it}^2 / \sum_i X_{it}^2\right), \qquad r = 1,...K \tag{3.33}$$

$$b_\theta = -\theta n^{-1}(0.6079k + 0.7715) . \tag{3.34}$$

In the case of a single explantory variable x with values centred such that $\sum_i X_i = 0$, these bias expressions give

$$b_0 = 0.1834n^{-1}\theta , \; b_1 = -\frac{1}{2}\theta \sum_i X_i^3 / (\sum_i X_i^2)^2 , \; b_{\hat{\theta}} = -1.3794n^{-1}\theta . \tag{3.35}$$

An important property which can easily be deduced from the likelihood equations is that the random variables

$$\hat{\theta}/\theta = \theta^{(1)} \text{ say }, \; (\hat{\beta}_r - \beta_r)/\theta_r = \beta_r^{(1)} \text{ say }, \; r = 0,1,..., k \tag{3.36}$$

are distributed independently of $\underset{\sim}{\beta}$ and 0. Thus from (2.9) the likelihood

equations depend on $\hat{Z}_i$ which may be written as

$$\hat{Z}_i = (u_i - \underset{\sim}{x}_i' \underset{\sim}{\beta}^{(1)}) / \hat{\theta}^{(1)} - \gamma$$

where

$$u_i = \theta^{-1}(y_i - \underset{\sim}{x}_i' \underset{\sim}{\beta}) , \qquad i = 1,...,n . \tag{3.37}$$

The $\{u_i\}$ are independently and identically distributed with p.d.f.

$$f_U(u) = \exp(u - \gamma - e^{u-\gamma}) , \qquad -\infty < u < \infty$$

which is the standardised type 1 EV distribution for smallest values and which does not depend on $\underset{\sim}{\beta}$ or θ It follows that the joint dis-

tribution of the random variables $\hat{\theta}^{(1)}$ and $\hat{\underset{\sim}{\beta}}^{(1)}$ is the same as that

of the ML estimators of 9 and when the 'observations' $\{u_i\}$ have the p.d.f. given by (3.38). Consequently, $\hat{\underset{\sim}{\beta}}^{(1)}$ and $\hat{\theta}^{(1)}$ are distributed

independently of $\underset{\sim}{\beta}$ and θ.

It is straightforward to show that the above distribution property for $\hat{\theta}^{(1)}$ and $\hat{\underset{\sim}{\beta}}^{(1)}$ holds generally for the class of regression models with

$f_{Y_i}(y) = \theta^{-1} f\{y-\mu_i)/0\}$, where $u_i - \underset{\sim}{x}_i' \underset{\sim}{\beta}$ This generalises the result given by Antle and Bain (1969) for the case when no explanatory variables are present.

In order to examine the adequacy of the bias and variance approximations and to assess other moment properties of the ML estimators, a Monte-Carlo simulation investigation was made for the case of simple linear regression with grouped data, the model being

$$Y_{ij} = \beta_o + \beta_i x_i + \varepsilon_{ij}\ , \qquad i = 1, \ldots g \quad j = 1,\ldots,m_i \tag{3.39}$$

where $E(\varepsilon_{ij}) = 0$, $\text{var}\ (\varepsilon_{ij}) = \frac{1}{\theta}\pi^2\theta^2$ and the $\{Y_{ij}\}$ are independently distributed with p.d.f. given by

$$f_{Y_{ij}}(y) = \frac{1}{\theta}\exp\left\{\frac{y-\beta_0-\beta_1 X_i}{\theta} - \gamma - \exp\left(\frac{y-\beta_0-\beta_1 X_i}{\theta} - \gamma\right)\right\},\ -\infty < y < \infty\ . \tag{3.40}$$

Equally spaced values of x were taken with $X_i = i - \frac{1}{2}(g+1)$, i - 1,...,g. Equal sample sizes $m_i = m = 1,2,5(5)20$ for i = 1 ,. .. ,g were used with g = 5,10. Without loss of generality, the values $\underset{\sim}{\beta} = \underset{\sim}{0}$ and $\theta = 1$ were used, the y-variate observations for the simulation being generated by sampling the distribution with density given by (3.38). The ML estimates were found correct to at least four decimal places using Fisher's scoring method. In each case, two independent runs each of size 2000 were made and the moment results then averaged over the two runs.

Values of the biases and variances of the ML estimates as obtained by simulation are shown in tables 2, 3, 4 for β_0, β_1 and θ respectively. The approximations to the biases and variances are also shown.

The results show that the approximate biases given by (3.35) agree quite well with the biases obtained by simulation even when m is very small. Further the biases in $\hat{\theta}$ are appreciably larger than the biases of the ML estimates of the regression coefficients. The large sample approximation for the variance of $\underset{\sim}{\hat{\beta}_0}$ works well for all m but the approximating variance of $\underset{\sim}{\hat{\beta}_0}$ underestimates the true variance when m = 1,2. In the case of $\hat{\theta}$, the approximating variance given by (3.3) generally agrees well with the values obtained by-simulation, but there is some overestimation of the variance when m = 1 and g = 5.

Table 2

Biases x $10^2 \theta^{-1}$ and variances x $10^{2} \theta^{-2}$ of the ML estimates of β

	m	Bias Simul	Bias Approx (3 .35)	Variance Simul	Variance Approx (3.3)
g=5	1	3.07	3.67	32.423	32.158
	2	1 .34	1 .83	15.631	16.079
	5	0.73	0.73	6.575	6.432
	10	-0.08	0.37	3.158	3.216
	15	0.35	0.25	2.118	2.144
	20	0.15	0.18	1.580	1 .608
g = 10	1	2.53	1 .83	15.807	16.079
	2	0.77	0.92	7.853	8.040
	5	0.40	0.37	3. 152	3.216
	10	0.02	0. 18	1 .591	1 .608
	15	0, 19	0.12	1 .075	1 .072
	20	-0.20	0.09	0.825	0.804

Table 3

Biases × $10^2 \theta^{-1}$ and variances x $10^2 \theta^{-2}$ of the ML estimates of β_1

	m	Bias Simul	Approx (3 .35)	Variance Simul	Approx (3.4)
g=5	1	-0.48	0.00	14.241	10.000
	2	0.40	0.00	5.705	5 .000
	5	-0.29	0.00	2. 130	2.000
	10	-0.14	0.00	1 .017	1 .000
	15	-0.06	0.00	0.681	0.667
	20	0.13	0.00	0.507	0.500
g-10	1	-0.12	0.00	1 .491	1 .212
	2	0.00	0.00	0.670	0.606
	5	0.12	0.00	0,260	0.242
	10	0.08	0.00	0.124	0.121
	15	0.12	0.00	0.082	0.081
	20	0.07	0.00	0.061	0.061

Table 4

Biases x $10^2\ \theta^{-1}$ and variances x $10^2\ \theta^{-2}$ of the ML estimates of θ

	m	Bias Simul	Bias Approx (3.35)	Variance Simul	Variance Approx (3.3)
g=5	1	-29.87	-27.59	10.357	12.158
	2	-14.51	-13.79	5.743	6.079
	5	-5.78	-5.52	2.417	2.432
	10	-2.61	-2.76	1.241	1.216
	15	-1 .89	-1.84	0.826	0.811
	20	-1.37	-1.38	0,598	0.608
g=10	1	-14.94	-13.79	5.944	6.079
	2	-7.05	-6.90	2.931	3.040
	5	-2.93	-2.76	1.238	1.216
	10	-1.31	-1.38	0.608	0.608
	15	-0.87	-0.92	0.411	0.405
	20	-0.67	-0.69	0.314	0.304

The shapes of the distributions of the ML estimates can be seen from the histogram plots which are shown in figures la,..., "If. Values of the skewness coefficient $\gamma_1(\hat{\beta}_0)$ t $\gamma_1(\hat{\beta}_0)$, $\gamma_1(\hat{\theta})$ and the kurtosis coefficients $\gamma_2(\hat{\beta}_0)$, $\gamma_2(\hat{\beta}_1)$ $\gamma_2(\hat{\theta})$ as obtained by simulation are also given correct to 2 decimal places . These must be treated with caution as their sampling errors are quite large even for the run-size of 4000 used in the investigation. The plots and the coefficients indicate that the distributions are reasonably close to the normal even for these moderately small values of m.

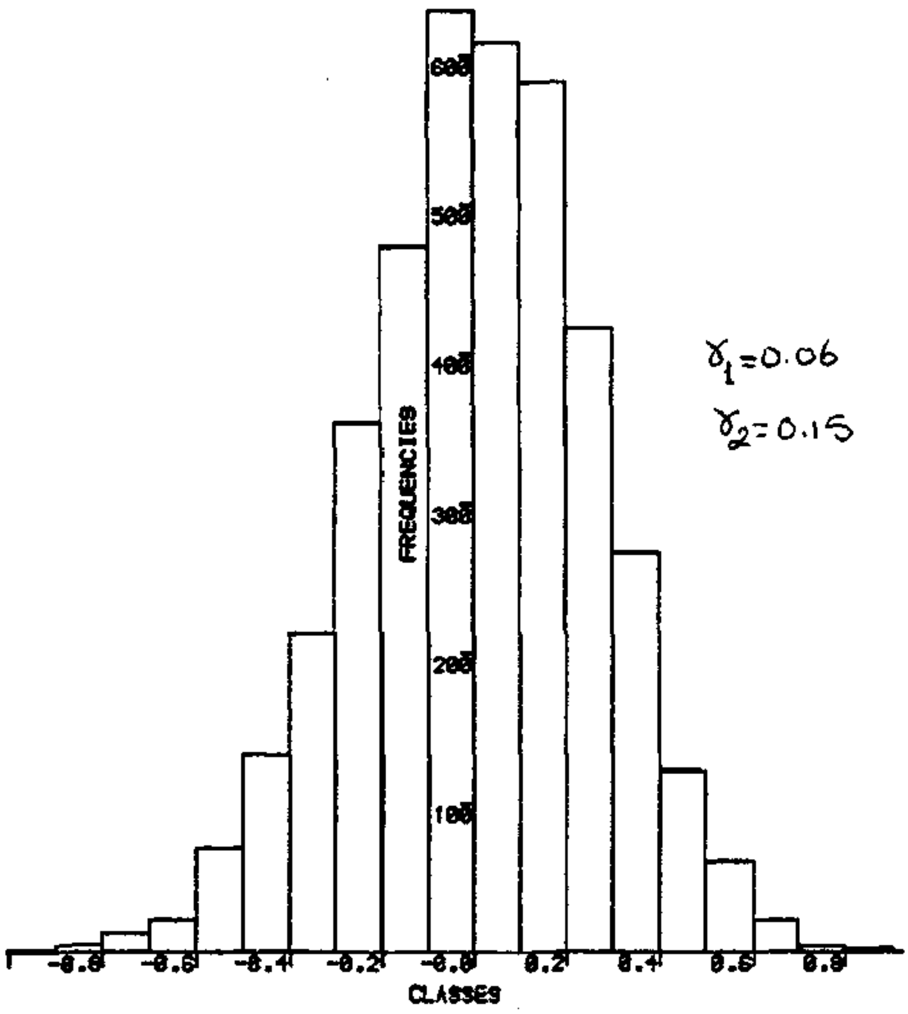

FIGURE (1.α) : HISTOGRAM PLOT FOR ML ESTIMATE $\hat{\beta\theta}$, m=5,g=5

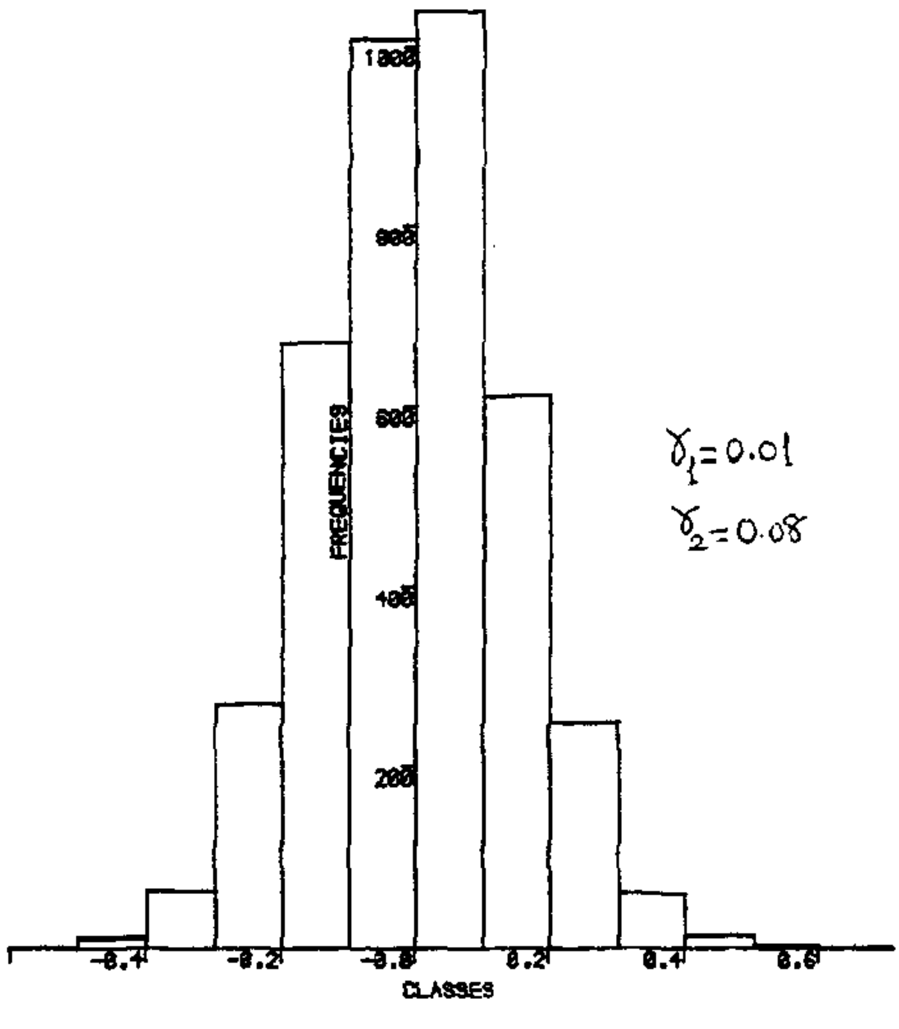

FIGURE (1.b) : HISTOGRAM PLOT FOR ML ESTIMATE $\hat{\beta}1$, m=5,g=5

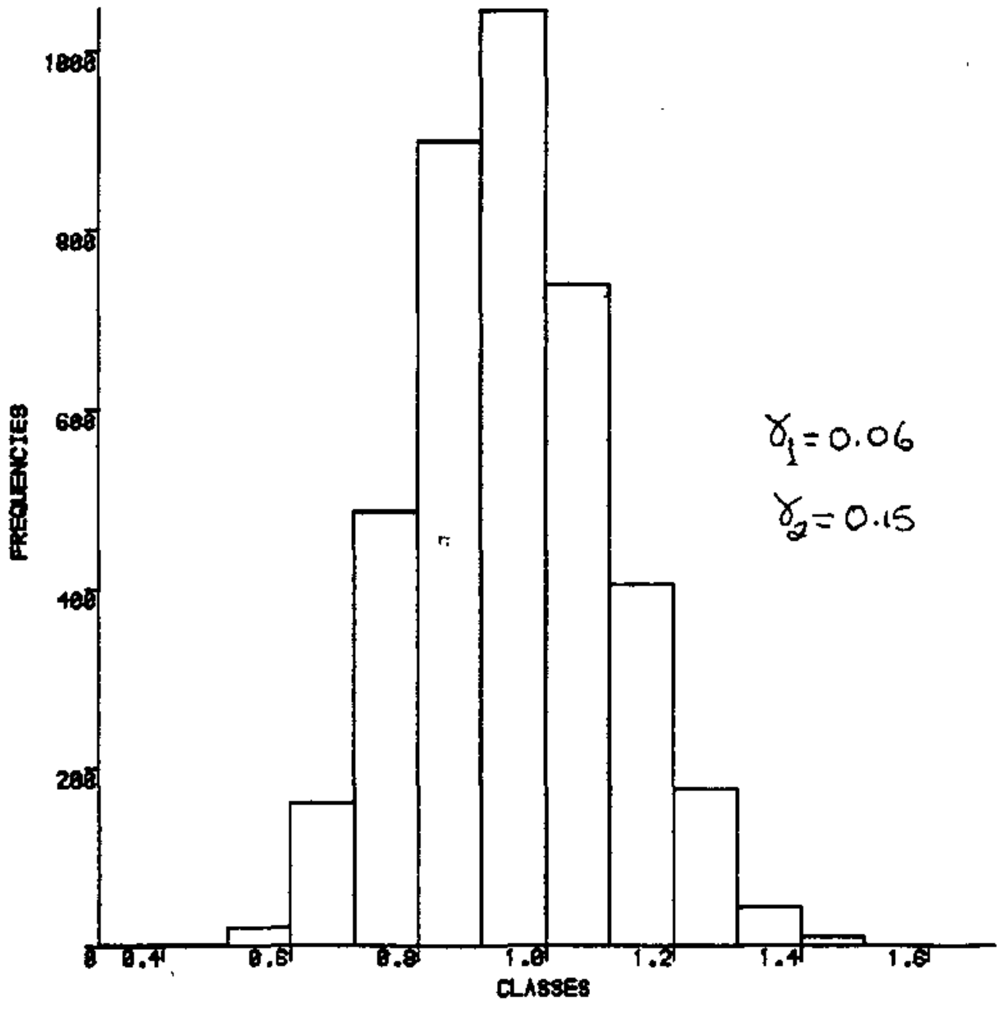

FIGURE (1.c) : HISTOGRAM PLOT FOR ML ESTIMATE $\hat{\theta}$, m=5,g=5

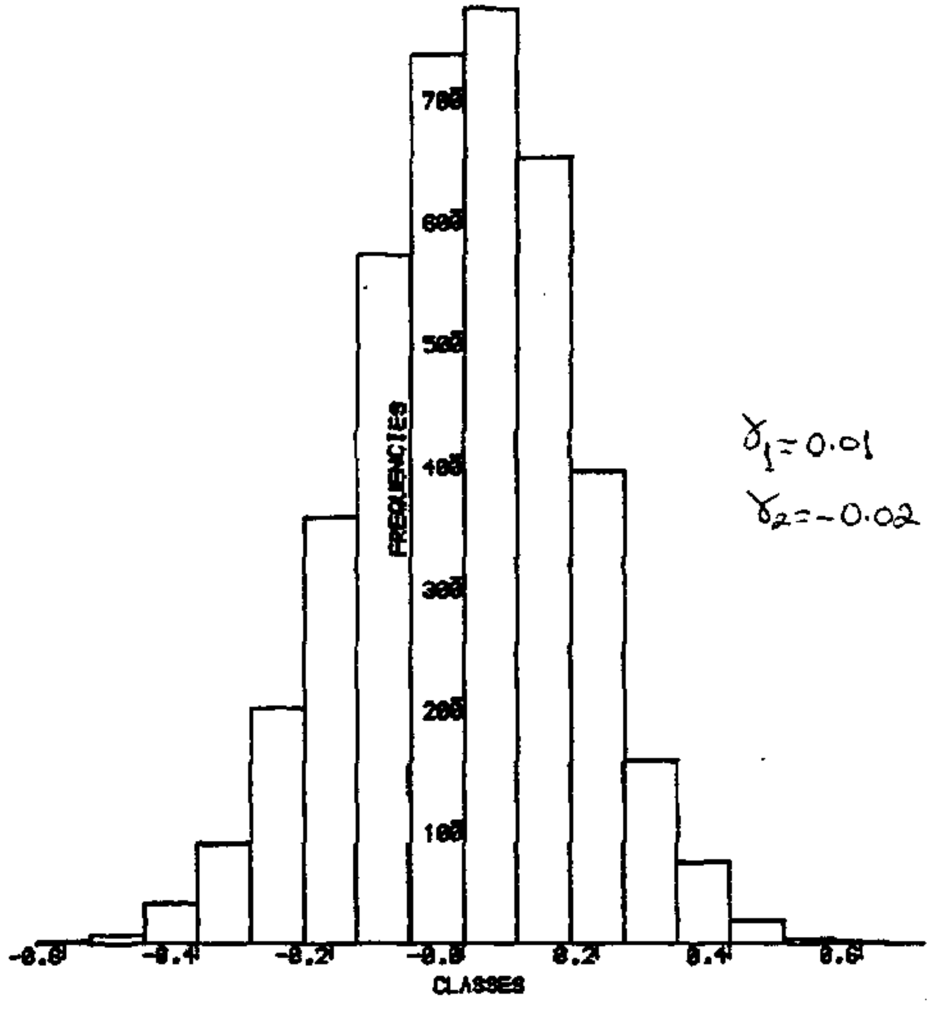

FIGURE (1.d) : HISTOGRAM PLOT FOR ML ESTIMATE $\hat{\beta}0$, m=10,g=5

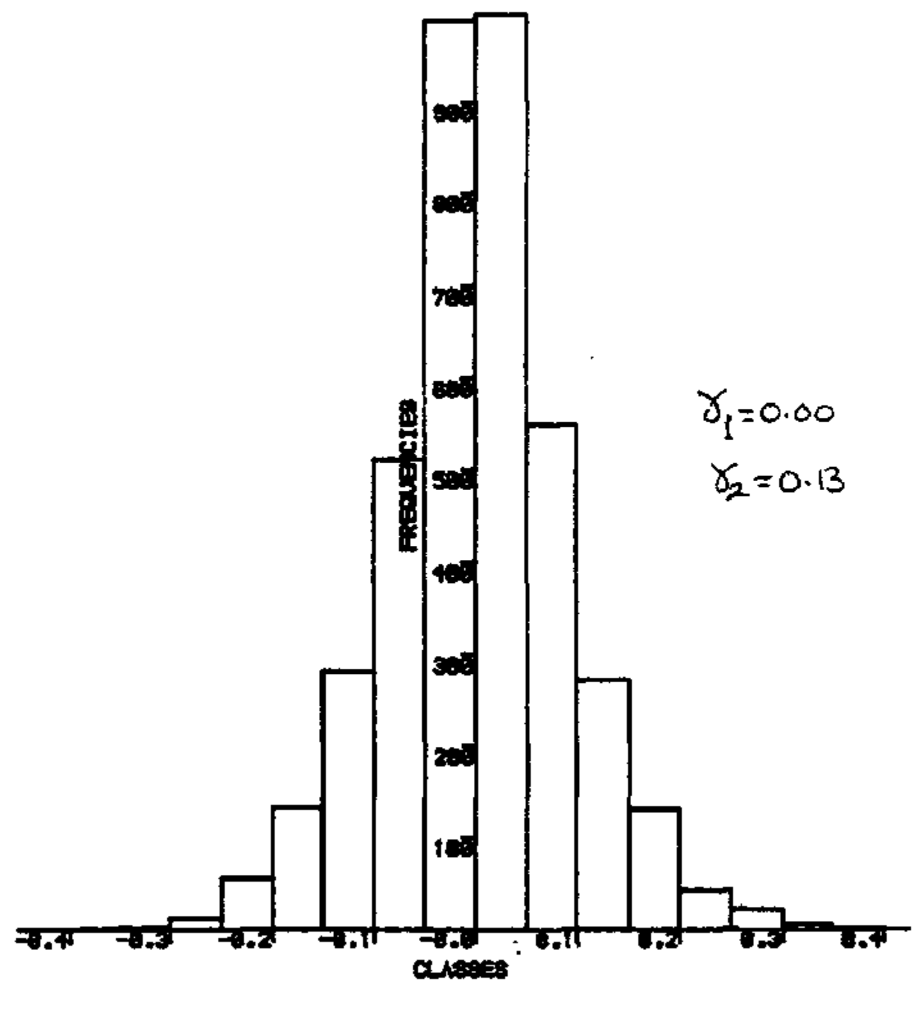

FIGURE (1.,e) : HISTOGRAM PLOT FOR ML ESTIMATE $\hat{\beta}_1$, m=10,g=5

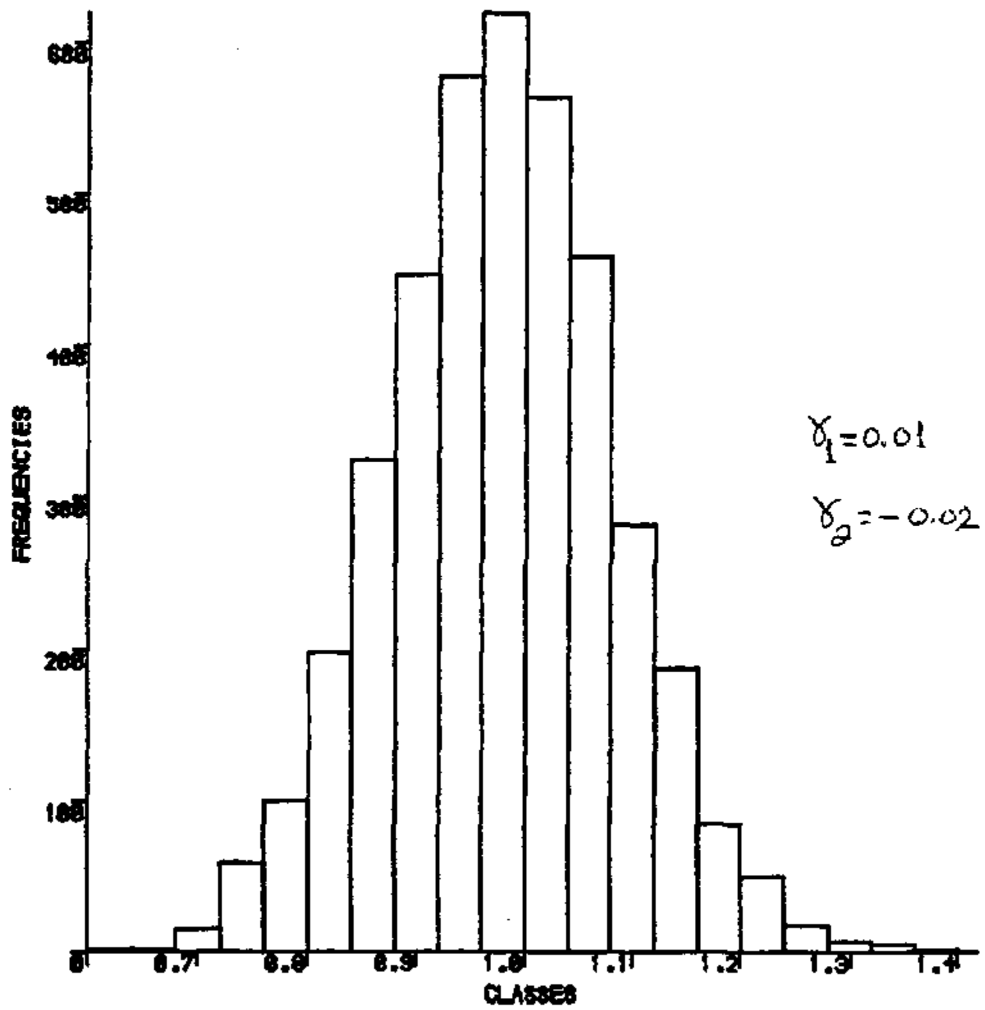

FIGURE (1.f) : HISTOGRAM PLOT FOR ML ESTIMATE $\hat{\theta}$, m=1θ,g=5

We have noted that the results in table 4 show that the bias and variance approximations for $\hat{\theta}$ work well even for moderately small values of n and that the bias is considerably larger than the standard deviation. Further the biases of the estimator $\hat{\theta}$ are much larger than those of the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, This suggests that it may be worthwhile considering estimators of θ with reduced bias and possibly an improved mean square error performance. Two such estimators are now developed.

Setting c = 1.3794 and a = 0.6079 we have to order n

$$b_{\hat{\theta}} = -c\,\theta\, n^{-1}, \qquad \mathrm{var}(\hat{\theta}) = a\theta^2 n^{-1}$$

which gives the approximation

$$\mathrm{mse}(\hat{\theta}) \approx \theta^2 (an^{-1} + C^2 n^{-2}). \tag{3.41}$$

An estimator having bias of order n is given by

$$\hat{\theta}^* = (1 + cn^{-1})\,\hat{\theta} \tag{3.42}$$

with $b_{\hat{\theta}^*} \approx -c^2\theta/n^2$. Ignoring terms of order n^{-3} and smaller terms, we have the approximation

$$\mathrm{mse}(\hat{\theta}^*) \approx \theta^2 (an^{-1} + 2acn^{-2}) \tag{3.43}$$

so the proportionate reduction in the mse" is

$$\frac{\mathrm{mse}(\hat{\theta}) - \mathrm{mse}(\hat{\theta}^*)}{\mathrm{mse}(\hat{\theta})} \approx \frac{C^2 + 2ac}{an + c^2} = \frac{0.226}{0.6.8n + 1.902}. \tag{3.44}$$

An alternative estimator of the form $\hat{\theta}^{**} \approx (1 + kn^{-1})\,\hat{\theta}$ can be considered, where k is selected to minimise the approximating mse of the estimator. The bias and variance of $\hat{\theta}^{**}$ to order n^{-1} are

$$b_{\hat{\theta}^{**}} = \theta\{(k - c)n^{-1} - kcn^{-2}\}, \qquad \mathrm{var}(\hat{\theta}^{**}) = k^2 a\theta^2 n^{-1} \tag{3.45}$$

giving the approximation

$$\mathrm{mse}(\hat{\theta}^{**}) \approx \theta^2 \left[an^{-1} + n^{-2}\{2ak + (k-c)^2\}\right]. \tag{3.46}$$

Setting $\partial\, \mathrm{mse}(\hat{\theta}^{**}) / \partial k = 0$ gives k = c - a, so the estimator is

$$\hat{\theta}^{**} = (1 + 0.7715n^{-1})\,\hat{\theta} \tag{3.47}$$

with an associated approximate mse

$$\mathrm{mse}(\hat{\theta}^{**}) \approx \theta^2 (an^{-1} + \{2ac - a^2) n^{-2}) \;. \qquad (3.48)$$

The proportionate reduction in the mse compared with the ML estimator $\hat{\theta}$ is

$$\frac{\mathrm{mse}(\hat{\theta}) - \mathrm{mse}(\hat{\theta}^*)}{\mathrm{mse}(\hat{\theta})} \approx \frac{(C - a)^2}{an + c^2} = \frac{0.595}{0.6.8n + 1.903} \;. \qquad (3.49)$$

Values of the biases and mse's of the estimates $\hat{\theta}$, $\hat{\theta}^*$ and $\hat{\theta}^{**}$ as estimated by simulation are shown in table 5.

The results show that the bias reduction estimator $\hat{\theta}^*$ has a much better bias performance than the ML estimator $\hat{\theta}$. Its mean square error performance is also much better when m= 1, but for m> 1 the differences in mean square error are small. The estimator $\hat{\theta}^{**}$ generally has a slightly improved mean square error performance compared with $\hat{\theta}^*$ but it has a poorer bias performance.

<u>Table 5</u>

Biases × $10^2\ \theta^{-1}$ and mse's × $10^2\ \theta^{-2}$ of the estimators $\hat{\theta}$ $\hat{\theta}^*$ $\hat{\theta}^{**}$

		Bias			mse		
	m	$\hat{\theta}$	$\hat{\theta}^*$	$\hat{\theta}^{**}$	$\hat{\theta}$	$\hat{\theta}^*$	$\hat{\theta}^{**}$
g=5	1	-29.87	-10.52	-19.05	19.280	17.967	17.429
	2	-14.51	-2.71	-7.91	7.847	7.510	7.288
	5	-5.18	-0.58	-2.87	2.755	2.694	2.651
	10	-2.61	0.08	-1.11	1.310	1.310	1.292
	15	-1 .89	-0.08	-0.88	0.862	0.857	0.851
	20	-1 .37	-0.01	-0.61	0.617	0.615	0.61 1
g=10	1	-14.94	-3.21	.-8.38	8.179	7,800	7.598
	2	-7.05	-0.64	-3.46	3.428	3.353	3.281
	5	-2.93	-0.25	-1.43	1.326	1.308	1 .296
	10	-1.31	0.05	-0.55	0.625	0.625	0.620
	15	-0.87	0.05	-0.36	0.419	0.419	0.417
	20	-0.67	0.02	-0.29	0.319	0.318	0.318

4. Moment Properties Of The Ordinary Least Squares Estimators

The OLS estimator of $\underset{\sim}{\beta}$ is $\tilde{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}$ showing that the individual parameter estimates $\tilde{\beta}_r$, r = 0,1,...,k, are linear functions of the $\{Y_i\}$, whose moments are known exactly. It follows that the exact moments of the $(\tilde{\beta}_r\}$ can be found. The estimators are unbiased and we have

$$\text{cov}\ (\tilde{\underset{\sim}{\beta}}) = \frac{1}{6}\pi^2\theta^2\ (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\ . \tag{4.1}$$

Assuming that $\sum_{i=1}^{n} X_{ir} = 0$ for r = 1 k we have

$$\underset{\sim}{X}'\underset{\sim}{X} = \begin{bmatrix} n & 0 \\ 0 & \underset{\sim 2}{I} \end{bmatrix},\quad \text{where}\quad \underset{\sim 2}{I} = ((\sum_i x_{ir}x_{is}))\ .$$

Hence we have

$$\text{var}\ (\tilde{\beta}_0) = \frac{1}{6}\pi^2\ \theta^2/n\ = 1.6449\ \theta^2\ /n \tag{4.2}$$

and the covariance matrix of $\tilde{\underset{\sim 1}{\beta}} = (\tilde{\beta}_0\ \tilde{\beta}\ 1,\ \ldots\ \beta k)^{\sim}$ is

$$\text{cov}(\tilde{\underset{\sim 1}{\beta}}) = \frac{1}{6}\pi^2\theta^2\ \underset{\sim 2}{I}^{-1} = 1.6449\ \theta^2((\sum_i\ x_{ir}x_{is}))^{-1}\ . \tag{4.3}$$

The exact skewness and kurtosis coefficients of $\tilde{\beta}_r$, can be found using the following results due to Scheffé (1959). Let $\xi = \sum_{i=1}^{n} C_i Y_i$ be a linear combination of n independent random variables $\{Y_i\}$ where Y. has variance σ_1^2, skewness coefficient $\gamma_{i,i}$. and kurtosis coefficient $\gamma_{2,}$ *i* . .. Then the skewness and kurtosis coefficients for ξ are given by

$$\gamma_{1,\xi} = \sum_{i=1}^{n} \alpha_i^{3/2}\gamma_{1,i} \qquad \gamma_{2,\xi} = \sum_{i=1}^{n} \alpha_i^2\gamma_{2,i} \tag{4.4}$$

where

$$\alpha_i = C_i^2\sigma_i^2 / \sum_i C_i^2\sigma_i^2\ .$$

We let c_{ir} denote the i th element in the rth row of $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$ so

$$\tilde{\beta}_r = \sum_{i=1}^{n} C_{ir}Y_i\ ,\quad \text{with}\quad \sigma_i^2 = \frac{1}{6}\pi^2\theta^2\ ,\ \gamma_{1,i} = 1.2986\ ,\ \gamma_{2,i} = 2.4000\ ,$$

i - 1,...,n. Hence use of (4.4)gives

$$\gamma_1\ (\tilde{\beta}_r) = 1.2986\ \sum_{i=1}^{n} C_{ir}^3/(\sum_i C_{ir}^2)^{3/2} \tag{4.5}$$

$$\gamma_2\ (\tilde{\beta}_r) = 2.4\ \sum_{i=1}^{n} C_{ir}^4/\sum_i C_{ir}^2 \tag{4.6}$$

for r = 0,1,...,k.

For the OLS estimator $\tilde{\theta}$ of 9, approximate bias and variance results can be obtained using the exact results available for the mean and variance of $\tilde{\theta}^2$ The LS estimator is $\tilde{\sigma}^2$ unbiased for σ^2 and from Atiqullah (1962), its exact variance is

$$\operatorname{var}(\tilde{\sigma}^2) = \frac{2\sigma^4}{n-k-1}\left\{1+\frac{1}{2}\,\frac{\gamma^2}{n-k-1}\sum_i (1-h_{ii})^2\right\} \tag{4.7}$$

where γ_2 is the coefficient of kurtosis of the distribution of the response variable Y and h_{ii}.- is the ith diagonal element in the 'hat' matrix $\tilde{H} = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ For the type 1 EV distribution we have $\gamma_2 = 2.4$ and hence the unbiased estimator of θ^2 given by

$$\tilde{\theta}^2 = 6\sum_i (Y_i - \tilde{X}_i'\tilde{\beta})^2 / \{(n-k-1)\pi^2\} \tag{4.8}$$

has exact variance

$$\operatorname{var}(\tilde{\theta}^2) = \frac{2\theta^4}{n-k-1}\left\{1+ \frac{1.2}{n-k-1}\sum_{i=1}^{n} (1-h_{ii})^2\right\}. \tag{4.9}$$

Writing

$$\tilde{\theta} = \theta\{1 + (\tilde{\theta}^2 - \theta^2)/\theta^2\}^{\frac{1}{2}} = \theta + (\tilde{\theta}^2 - \theta^2)/(2\theta) - (\tilde{\theta}^2 - \theta^2)^2/(8\theta^2)\ldots$$

and using the result that $\operatorname{tr}(\tilde{H}) = k+1$ we obtain to order n^{-1},

$$E(\tilde{\theta}) = \theta(1 - 0.55n^{-1}) \tag{4.10}$$

$$\operatorname{var}(\tilde{\theta}) = 1.1n^{-1}\theta^2 . \tag{4.11}$$

Table 6 gives values of the exact variances $(\times\theta^{-2})$ skewness and kurtosis coefficients for the OLS estimators $\tilde{\beta}_0$ and $\tilde{\beta}1$ for the simple linear regression model considered in the presvious section. Table 7 shows values of the approximate biases and variances of the OLS estimator $\tilde{\theta}$ given by (4.10) and (4.11) for the same model. The biases and variances of $\tilde{\theta}$ as obtained by simulation are also shown, the simulation again using a run-size of 4000.

Table 6

Exact variance (x $10^2 \theta^{-2}$), skewness and kurtosis coefficients of the OLS estimators $\tilde{\beta}_0$ $\tilde{\beta}_1$

		$\tilde{\beta}_0$			$\tilde{\beta}_1$		
	m	var.	skew.	kurt.	var.	skew.	kurt.
g=5	1	32.899	0.58	0.10	16.449	0.00	0.08
	2	16.449	0.41	0.02	8.225	0.00	0.02
	5	6.580	0.26	0.00	3.290	0.00	0.00
	10	3.290	0.18	0.00	1.645	0.00	0.00
	15	2. 193	0. 15	0.00	1.097	0.00	0.00
	20	1 .645	0. 13	0.00	0.823	0.00	0.00
g=10	1	16.449	0.41	0.02	1 .994	0.00	0.01
	2	8.225	0.29	0.01	0.997	0.00	0.00
	5	3.290	0.18	0.00	0.399	0.00	0.00
	10	1.645	0.13	0.00	0. 199	0.00	0.00
	15	1.097	0.11	0.00	0. 133	0.00	0.00
	20	0.823	0.09	0.00	0. 100	0.00	0.00

Table 7

Biases and variances (x $10^2 \theta^{-2}$)

		Bias		Variance	
	m	App. (4.10)	Simul	App. (4.11)	Simul
g=5	1	-11.00	-10.40	22.000	18.921
	2	-5.50	-5.21	11.000	9.413
	5	-2.20	-2.43	4.400	3.929
	10	-1.10	-1 .03	2.200	2.060
	15	-0.73	-0.88	1.467	1.461
	20	-0.55	-0.66	1.110	1.074
g=10	1	-5.50	-5.60	11.000	9.623
	2	-2.65	-2.68	5.500	4.916
	5	-1.10	-1.16	2.200	2.126
	10	-0.55	-0.63	1.100	1.058
	15	-0.37	-0.35	0.738	0.759
	20	-0.27	-0.32	0.550	0.540

Histogram plots of the distributions of $\tilde{\beta}_0$, $\tilde{\beta}_1$, and $\tilde{\theta}$ as generated by simulation are shown in figures 2a,...,2f for g = 5 and m = 5,10. The values of the skewness and kurtosis coefficients for $\tilde{\theta}$ are also shown. The plots show that the distribution of the estimators are close to the normal.

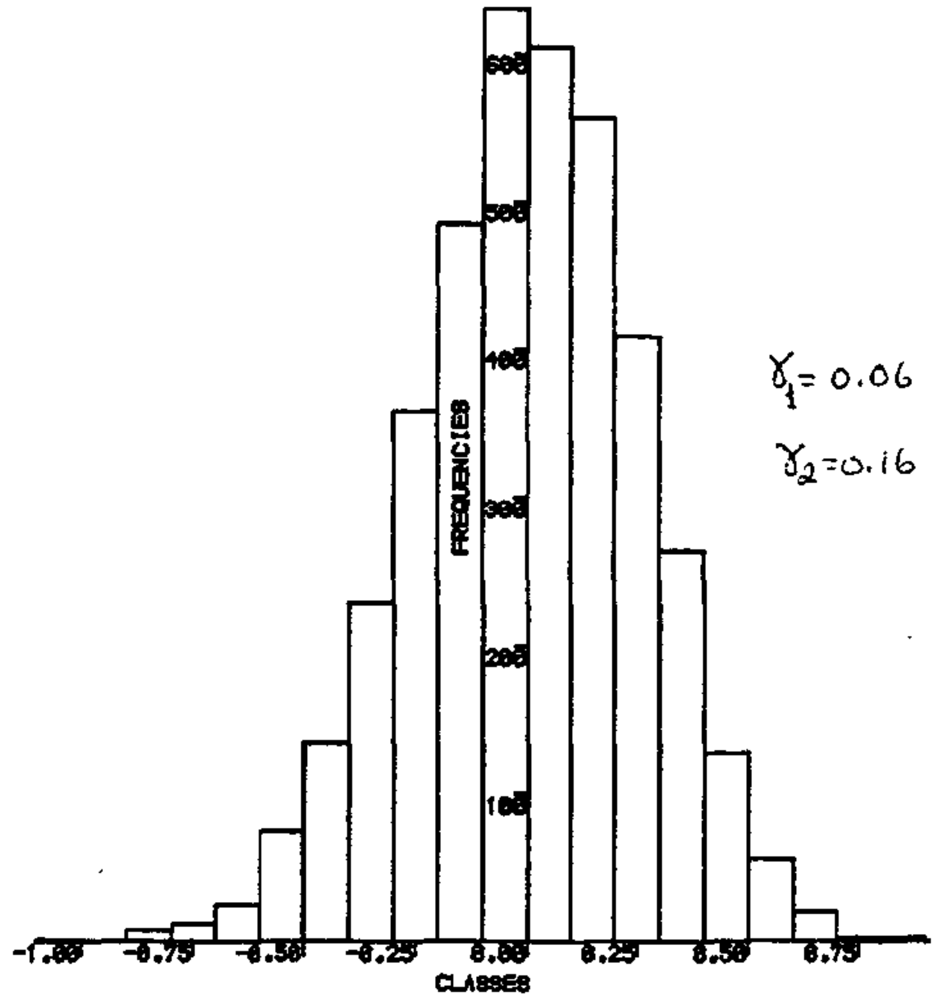

FIGURE (2.α) : HISTOGRAM PLOT FOR OLS ESTWATE $\widetilde{\beta 0}$, m=5, g=5

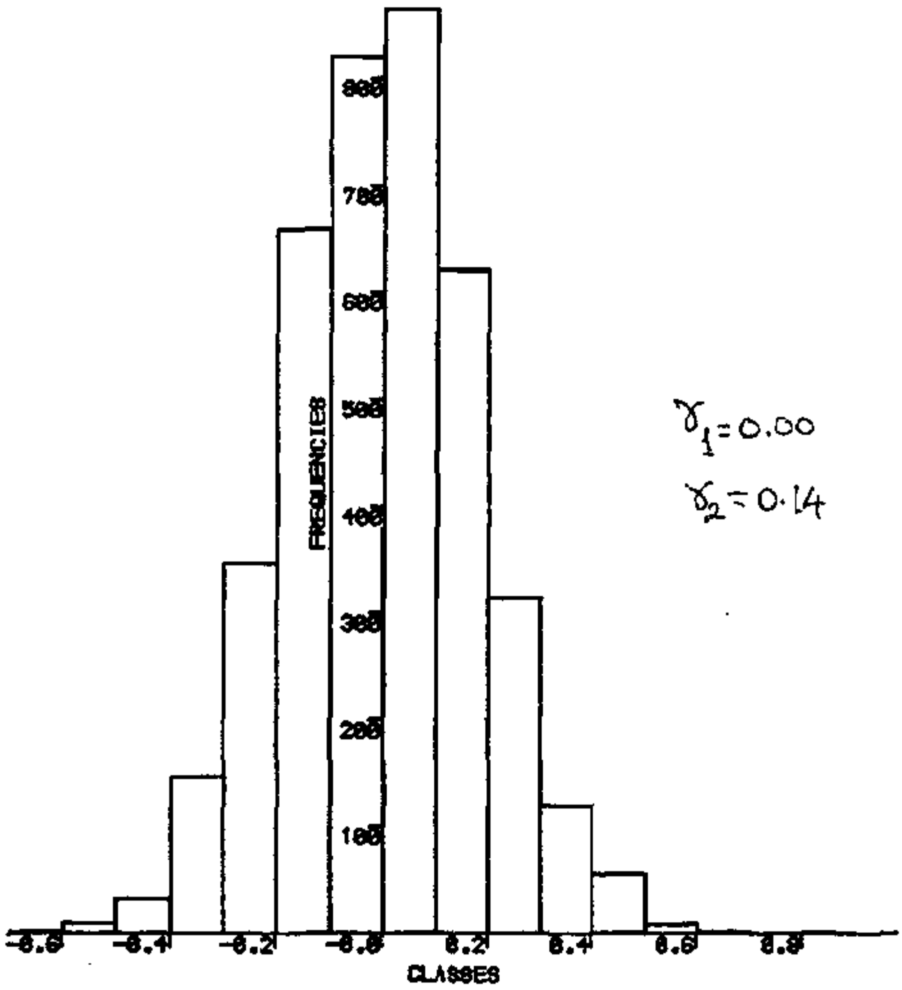

FIGURE (2.b) : HISTOGRAM PLOT FOR OLS ESTIMATE $\widetilde{\beta 1}$, m=5,g=5

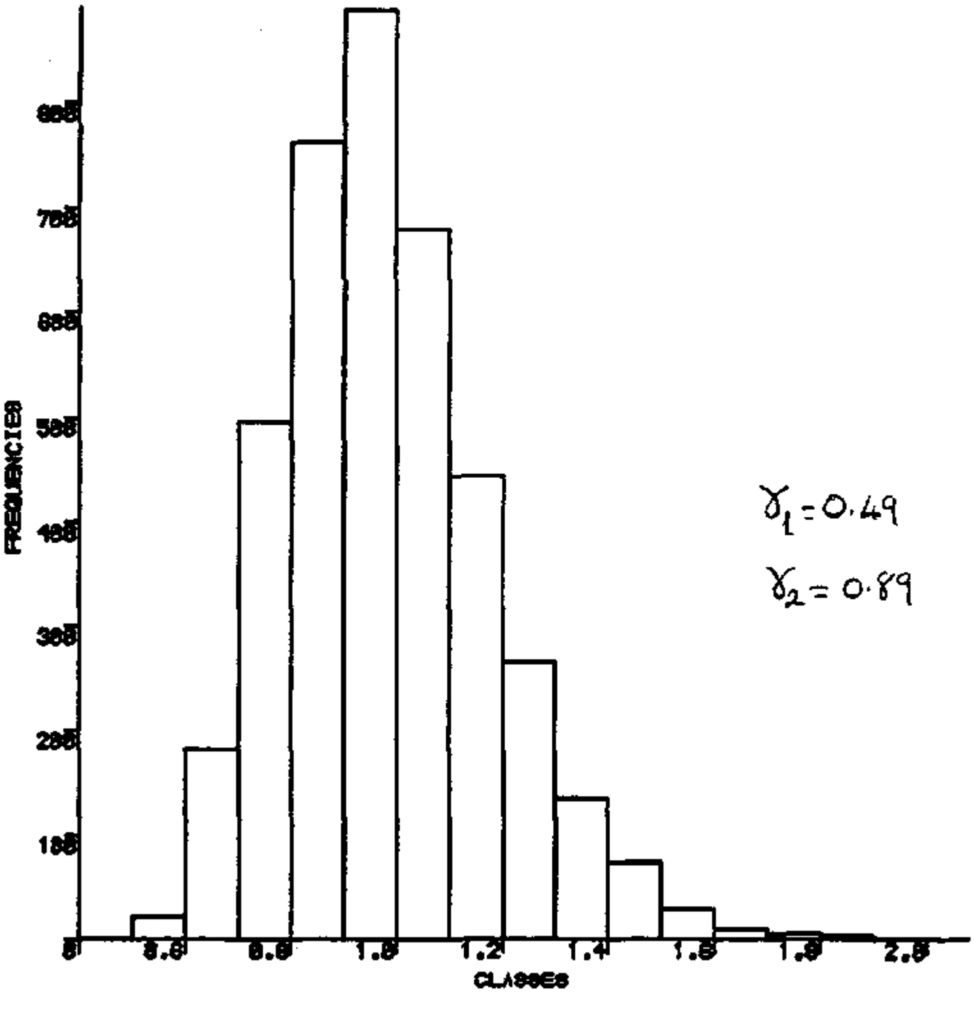

FIGURE (2.c) : HISTOGRAM PLOT FOR OLS ESTIMATE $\tilde{\theta}$, m=5,g=5

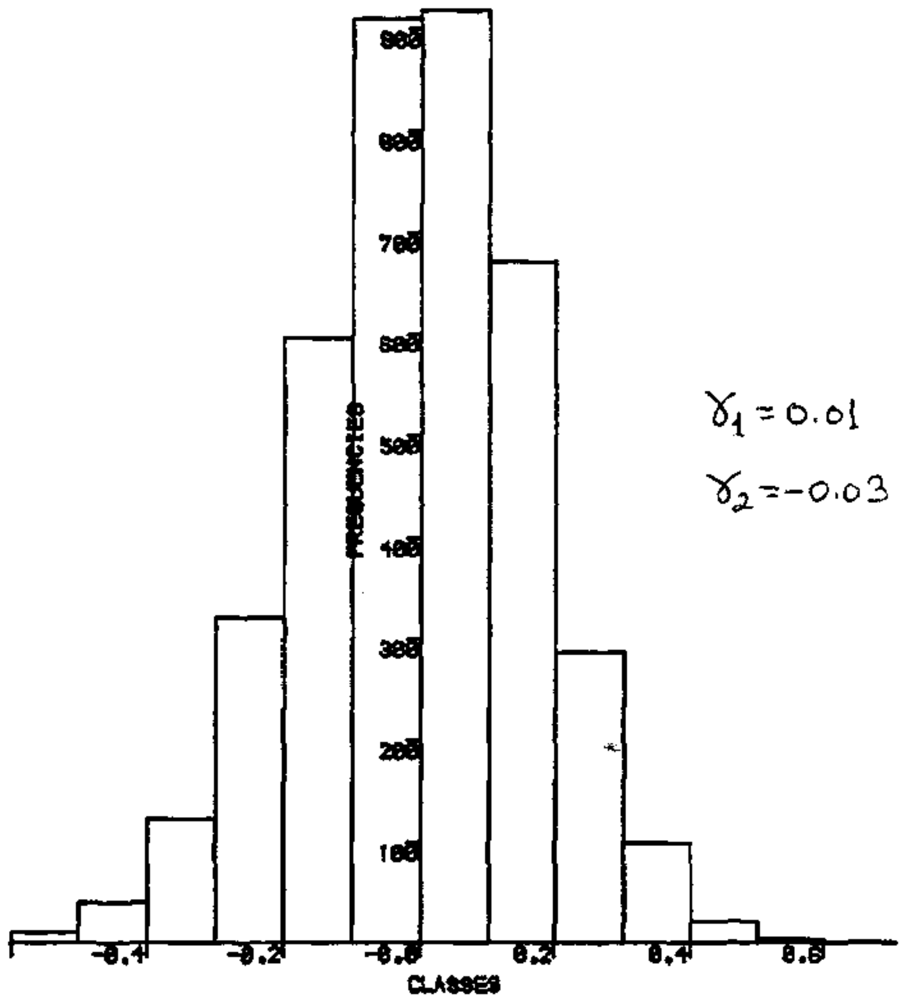

FIGURE (2.d) : HISTOGRAM PLOT FOR OLS ESTIMATE $\tilde{\beta}_0$, m=10,g=5

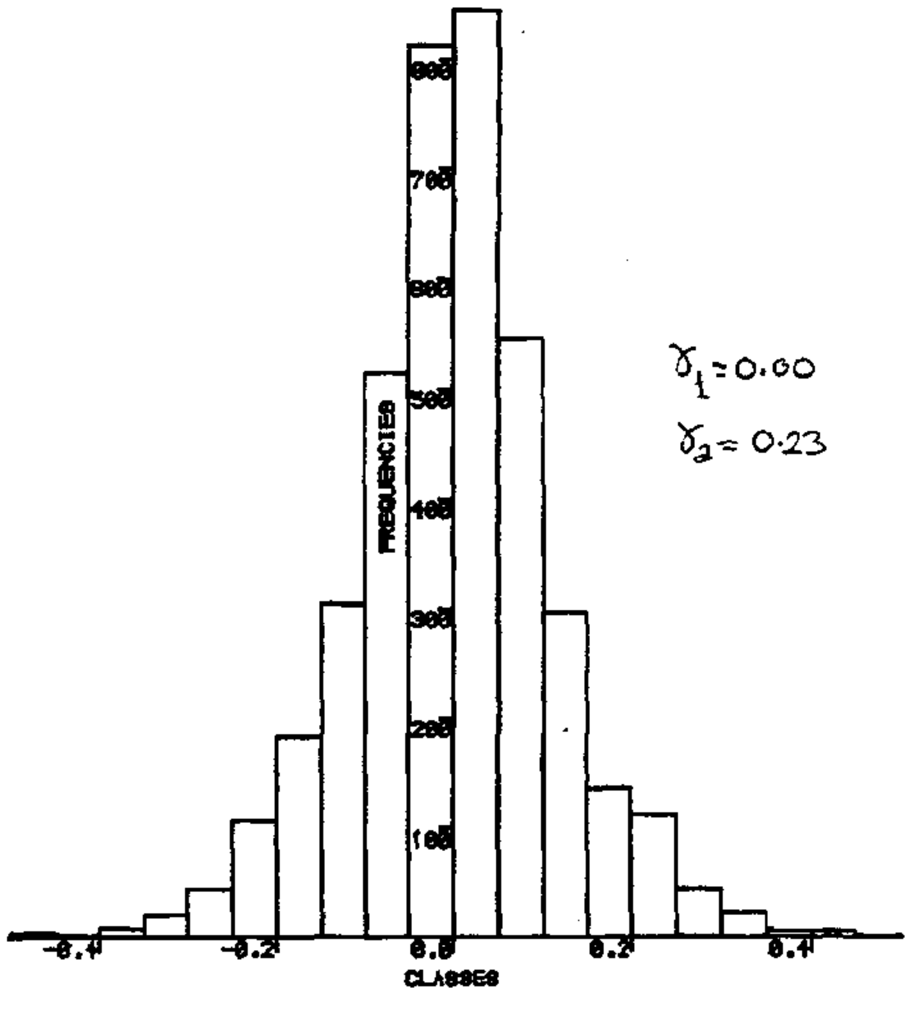

FIGURE (2.e) : HISTOGRAM PLOT FOR OLS ESTIMATE $\tilde{\beta}_1$, m=1θ,g=5

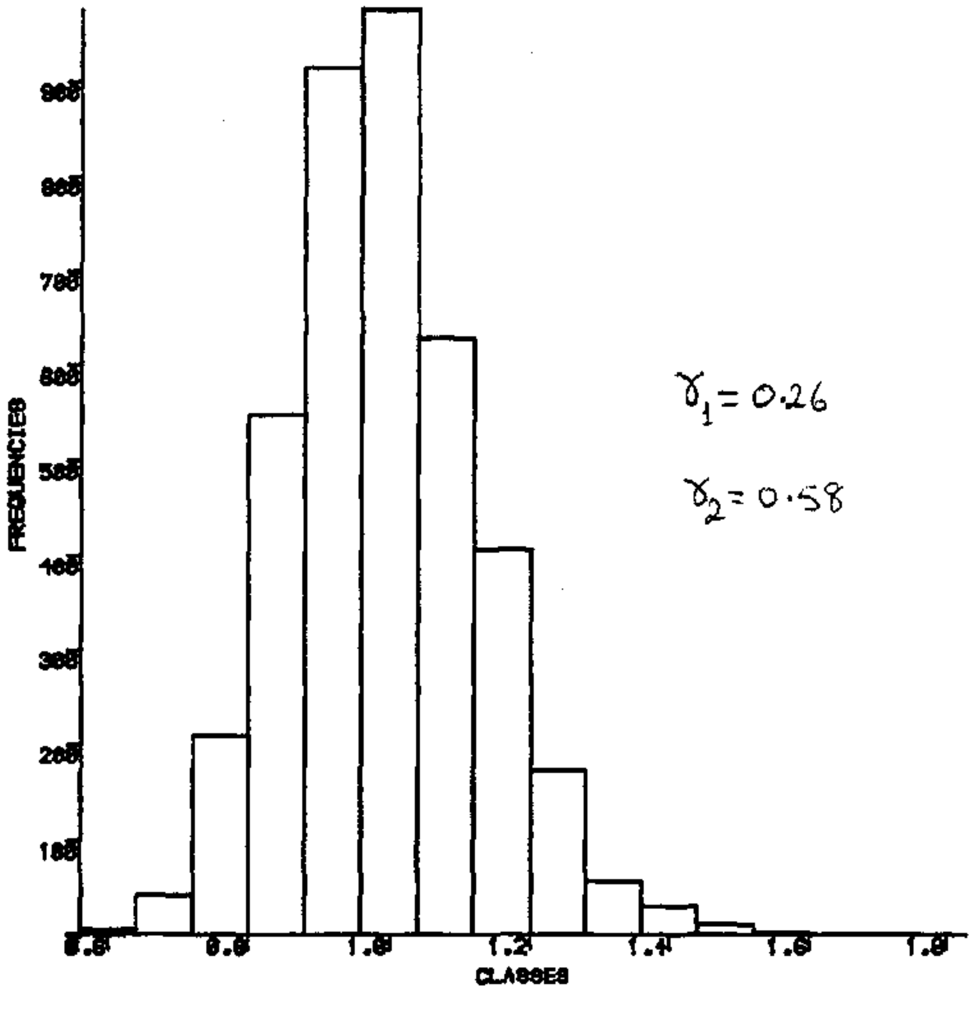

FIGURE (2.f) : HISTOGRAM PLOT FOR OLS ESTIMATE 8, m=10,g=5

5. Best Linear Unbiased Estimators Based On Order Statistics For Grouped Data

In this section, we shall assume that we have g distinct sets of values $x_{i1}, \ldots, x_{ik}$, i = 1,...,g for the explanatory variables and suppose that m_i observations on Y are made at the point $(x_{i1}, \ldots, x_{ik})$.

We let $n = \sum_{i=1}^{g} m_i$. denote the total number of observations. With grouped data, it is possible to find the best linear unbiased estimates (BLUE's) of the $\{\beta_r\}$ and θ based on the order statistics within the groups. These estimators are asymptotically as efficient as the ML estimators and have appreciably smaller variances than the OLS estimators.

We first need some results for the case when a single random sample of m observations is drawn from a population with the type 1 EV c.d.f.

$$F(y) = 1 - \exp\left\{- e^{(y-\xi)/\theta}\right\}, \quad -\infty < y < \infty . \tag{5.1}$$

It is convenient to allow for censoring. Thus if $Y_{(1)} < Y_{(2)} , < \ldots < Y_{(m)}$ denote the order statistics in the sample, we shall assume that only The first r order statistics observed. We let

$$\hat{\xi}_* = \sum_{i=1}^{r} a_i Y_{(i)} , \quad \hat{\theta}_* = \sum_{i=1}^{r} b_i Y_{(i)} \tag{5.2}$$

denote the BLUE's for ξ and θ respectively, that is we require $E(\hat{\xi}_*) = \xi$, $E(\hat{\theta}_*)$ = 0 and $\mathrm{var}(\hat{\xi}_*) \leq$ var, $(\hat{\xi}')$ var ≤. $\mathrm{var}(\hat{\theta}_*)$, where $(\hat{\theta}*)$,denote any other linear unbiased estimators of ξ and θ, respectively.

If we put $X_{(i)} = (Y_{(i)}-\xi)/\theta$, i = 1, ...r, then $X(1), X_{(2)}, \ldots, X(r)$ are distributed as the first r order statistics in a random sample of m observations from the standardised type 1 EV distribution with c.d.f. F(x) = 1 - exp(-expx), $-\infty < x < \infty$. Putting

$$E(X_{(i)}) = e_{i,m}, \qquad \mathrm{cov}(X_{(i)},X_{(j)}) = c_{ij,m} \tag{5.3}$$

we have

$$E(Y_{(i)}) = \xi + \theta e_i , \quad \mathrm{cov}(Y_{(i)},Y_{(j)}) - \theta^2{}_{ij}, m3. \tag{5.4}$$

In matrix form, we have the linear model representation

$$E(\underset{\sim}{Y}_{(.)}) = \underset{\sim}{A}\underset{\sim}{n} , \quad \mathrm{cov}(\underset{\sim}{Y}_{(.)}) = \theta^2 \underset{\sim}{C} \tag{5.5}$$

where $\underset{\sim(.)}{Y'} = (Y_{(1)}, Y_{(2)}, \ldots, Y_{(r)})$, $\underset{\sim}{n'} = (\xi, \theta)$

$$\underset{\sim}{A} = \begin{bmatrix} 1 & e_{1,m} \\ 1 & e_{2,m} \\ . & . \\ . & . \\ . & . \\ 1 & e_{r,m} \end{bmatrix}, \quad \underset{\sim}{C} = \begin{bmatrix} C_{11,m} & \cdots & C_{1r,m} \\ C_{21,m} & \cdots & C_{2r,m} \\ . & & . \\ . & & . \\ C_{r1,m} & & C_{rr,m} \end{bmatrix} \tag{5.6}$$

From generalised least squares theory, the BLUE for $\underset{\sim}{n}$ is $\hat{\underset{\sim}{n}} = (\underset{\sim}{A}'\underset{\sim}{C}^{-1}\underset{\sim}{A})^{-1}\underset{\sim}{A}'\underset{\sim}{C}^{-1}\underset{\sim(.)}{Y}$ with $\text{cov}(\hat{\underset{\sim}{n}}) = (\underset{\sim}{A}'\underset{\sim}{C}^{-1}\underset{\sim}{A})^{-1}\theta^2$. The linear co-efficients $a_i = a_i(r,m)$ and $b_i = b_i(r,m)$ giving the BLUES are obtained as the elements in the first and second rows of the matrix $(\underset{\sim}{A}'\underset{\sim}{C}^{-1}\underset{\sim}{A})^{-1}\underset{\sim}{A}'\underset{\sim}{C}^{-1}$ which -is of order $2 \times r$. Further, writing

$$(\underset{\sim}{A}'\underset{\sim}{C}^{-1}\underset{\sim}{A})^{-1} = \underset{\sim}{V} = \begin{bmatrix} V^{(1)}_{r,m} & V^{(2)}_{r,m} \\ V^{(2)}_{r,m} & V^{(3)}_{r,m} \end{bmatrix} \tag{5.7}$$

we have

$$\text{va}(\xi) = \theta^2 V^{(1)}_{r,m}, \quad \text{var}(\hat{\theta}) = \theta^2 V^{(3)}_{r,m}, \quad \text{cov}(\hat{\xi},\hat{\theta}), = \theta^2 V^{(2)}_{r,m}. \tag{5.8}$$

White (1964) gives tables of the linear coefficients $\{a_i.\}$ and $\{b_i\}$ and the elements $V^{(1)}_{r,m}$, $V^{(2)}_{r,m}$, $V^{(3)}_{r,m}$ required for computation of the covariance matrix for $2 \le n \le. 20$ and $r = 2(1\ )n$. The mean of the distribution given by (5.1) is $\mu = \xi - \gamma\theta$ so the BLUE for u is

$$\hat{\mu}_* = \sum_{i=1}^{r} (a_i - \gamma b_i) Y_{(i)} \tag{5.9}$$

with

$$\text{var}(\hat{\mu}_*) = \theta^2 (V^{(1)}_{r,m} - 2\ \gamma V^{(2)}_{r,m} + \gamma^2 V^{(3)}_{r,m}) = \theta^2 V r,m \ \text{ say} \tag{5.10}$$

Suppose now that we have g populations, the c.d.f. for the ith population being

$$F_i(y) = 1 - \exp[-\exp\{(y - \xi_i)/\theta\}], \qquad -\infty < y < \infty \tag{5.11}$$

$i = 1,\ldots,g$. Let $Y_{i(1)}, .. < Y_{i(2)}., < \ldots < Y_{i(m_i)}$. denote the order statistics in a random sample of m_i. observations drawn from the ith population. If only the first r_i order statistics $Y_{i(1)}, ., Y_{i(2)} \ldots, Y_{i(r_i)}, .$ are observed for the ith group, the BLUE's are

$$\hat{\xi}_i = \sum_{j=1}^{r_i} a_{ij} Y_{i(j)} , \qquad \hat{\theta}_i = \sum_{j=1}^{r_i} b_{ij} Y_{i(j)} , \tag{5.12}$$

where $a_{ij} = a_j(r_i, m_i)$ and $b_{ij} = b_j(r_i, m_i.)$ are the linear coefficients for best linear unbiased estimation. Since the $\{\hat{\theta}_i\}$ are independent estimators of the assumed common θ with $\mathrm{var}(\hat{\theta}_i) = \theta^2 V^{(3)}_{r_i,m_i}$, the minimum variance linear unbiased estimator of θ is

$$\hat{\theta}_* = \frac{\sum_{i=1}^{g} \hat{\theta}_i / V^{(3)}_{r_i,m_i}}{\sum_{i=1}^{g} \left(1/V^{(3)}_{r_i,m_i}\right)} = \sum_{i=1}^{g} w_i \hat{\theta}_i \text{ say} \tag{5.13}$$

where

$$w_i = \left\{ V^{(3)}_{r_i,m_i} \sum_{i=1}^{g} (1/V^{(3)}_{r_i,m_i}) \right\}^{-1} , \quad i = 1,\ldots\ldots,g . \tag{5.14}$$

We have

$$E(\hat{\theta}_*) = \theta , \qquad \mathrm{var}(\hat{\theta}_*) = \theta^2 \left\{ \sum_i (1/V^{(3)}_{r_i,m_i}) \right\}^{-1} . \tag{5.15}$$

The best linear unbiased estimator of the mean $\mu_i. = \xi_i . - \gamma\theta$ for the ith group is

$$\hat{\mu}_{*i} = \hat{\xi}_i - \gamma \hat{\theta}_* . \tag{5.16}$$

Using the result that $\mathrm{cov}(\xi_i, \hat{\theta}_*) = W_i \, \mathrm{Cov}(\hat{\xi}_i, \hat{\theta}_i) = W_i \, V^{(2)}_{r_i, m_i} \theta^2$, we obtain

$$\mathrm{var}(\hat{\mu}_{*i}) = \theta^2 \left[V^{(1)}_{r_i,m_i} + \left\{ -2\gamma\gamma({}^{(2)}_{r_i,m_i} / V^{(3)}_{r_i,m_i}) + \gamma^2 \right\} \left\{ \sum_i (1/V^{(3)}_{r_i,m_i}) \right\}^{-1} \right]$$

$$= \theta^2 w_{1i} \text{, say.} \tag{5.17}$$

Also

$$\mathrm{cov}(\hat{\mu}_{*i}, \hat{\mu}_{*j}) = -\gamma \left\{ \mathrm{cov}(\hat{\theta}_*, \hat{\xi}_i) + \mathrm{cov}(\hat{\theta}_*, \hat{\xi}_i) + \gamma^2 \, \mathrm{var}(\hat{\theta}_*) \right\}$$

$$= \theta^2 \frac{\left\{ -\gamma\left(V^{(2)}_{r_i,m_i} / V^{(3)}_{r_i,m_i} + V^{(2)}_{r_j,m_j} / V^{(3)}_{r_j,m_j} \right) + \gamma^2 \right\}}{\sum_i (1/V^{(3)}_{r_i,m_i})}$$

$$= \theta^2 w_{ij} , \quad \text{say} \tag{5.18}$$

Since μ_{*i}. is an unbiased estimator of $\mu_i. = \underset{\sim}{x}_i' \underset{\sim}{\beta}$ we have the linear model

$$\hat{\mu}_{*i} = \underset{\sim}{x}_i' \underset{\sim}{\beta} + \varepsilon_{*i} \qquad i = 1,\ldots,g \tag{5.19}$$

where $E(\varepsilon_{*i}.) = 0$, $\mathrm{var}(\varepsilon_{*i}) = \theta^2 w_{ii}$ and $\mathrm{cov}(\varepsilon_{*i}, \varepsilon_{*j}) = \theta^2 w_{ij}$. Setting $\underset{\sim}{\hat{\mu}}_*' = (\hat{\mu}_{*1},\ldots, \hat{\mu}_{*g})$ and $\underset{\sim}{\varepsilon}_*' = (\varepsilon_{*1},\ldots, \varepsilon_{*g})$, the matrix form of the linear model repres entation is

$$\underset{\sim}{\hat{\mu}}_* = \underset{\sim}{X}_1 \underset{\sim}{\beta} + \underset{\sim}{\varepsilon}_* \tag{5.20}$$

with $E(\underset{\sim}{\varepsilon}_*) = \underset{\sim}{0}$ and $\mathrm{cov}(\underset{\sim}{\varepsilon}_*) = \theta^z \underset{\sim}{W}$, where

$$\underset{\sim}{X}_1 = \begin{bmatrix} 1 & x_{11} & x_{1k} \\ 1 & x_{21} & x_{2k} \\ . & & \\ . & & \\ 1 & x_{g1} \cdots & x_{gk} \end{bmatrix}, \qquad \underset{\sim}{W} = \begin{bmatrix} w_{11} & w_{12} \cdots & w_{1g} \\ w_{21} & w_{22} \cdots & w_{2g} \\ . & & \\ . & & \\ w_{g1} & w_{g2} \cdots & w_{gg} \end{bmatrix} . \tag{5.21}$$

Based on the $\{\hat{\mu}_{*i}\}$, the BLUE of $\underset{\sim}{\beta}$ is given by

$$\underset{\sim}{\hat{\beta}}_* = (\underset{\sim}{X}_1' \underset{\sim}{W}^{-1} \underset{\sim}{X}_1)^{-1} \underset{\sim}{X}_1' \underset{\sim}{W}^{-1} \underset{\sim}{\hat{\mu}}, \tag{5.22}$$

with covariance matrix

$$\mathrm{cov}\,(\underset{\sim}{\hat{\beta}}_*) = (\underset{\sim}{X}_1' \underset{\sim}{W}^{-1} \underset{\sim}{X}_1)^{-1} \theta^2 . \tag{5.23}$$

We now consider the important special case when the sample sizes are equal and the same degree of censoring occurs within each group, that is $m_i = m$, $r_i. = r$ for $i = 1,\ldots,g$. In this case we have

$$W_{ii} = V^{(1)}_{r,m} + g^{-1} \left\{ \gamma^2 V^{(3)}_{r,m} - 2\gamma V^{(2)}_{r,m} \right\} = V , \quad \text{say} \tag{5.24}$$

$$W_{ij} = g^{-1} \left\{ \gamma^2 V^{(3)}_{r,m} - 2\gamma V^{(2)}_{r,m} \right. = w , \quad \text{say} \quad . \tag{5.25}$$

Letting $w^{(i\,j)}$ denote the (i,j)th element in $\underset{\sim}{W}^{-1}$, we have

$$w^{(ii)} = \frac{v+(g-2)w}{(v-w)\{v+(g-1)w\}}\ ,\ w^{(ij)} = -\frac{w}{(v-w)\{v+(g-1)w\}}$$

for $i \neq j = 1,2, \ldots ,g$.

putting

$$\underset{\sim}{M}\begin{bmatrix} \sum_{i=1}^{g} x_{i1} & \sum_{i=1}^{g} x_{i1}x_{i2} \ldots & \sum_{i=1}^{g} x_{i1}x_{i1} \\ \sum_{i=1}^{g} x_{i1}x_{i2} \ldots & \sum_{i=1}^{g} x_{i2}^{2} \ \cdots & \sum_{i=1}^{g} x_{i2}x_{ik} \\ . & & \\ . & & \\ \sum_{i=1}^{g} x_{i1}x_{ik} \ldots & \sum_{i=1}^{g} x_{i2}x_{ik} \ldots & \sum_{i=1}^{g} x_{ik}^{2} \end{bmatrix}$$

we may write

$$(\underset{\sim 1}{X}'\underset{\sim}{W}^{-1}\underset{\sim 1}{X})^{-1} = \begin{bmatrix} \frac{v+(g-1)w}{g} & \underset{\sim}{0} \\ \underset{\sim}{0} & (v-w)\underset{\sim}{M}^{-1} \end{bmatrix},\quad \underset{\sim 1}{X}'\underset{\sim}{W}^{-1} = \begin{bmatrix} \{v+(g-1)w\}^{-1} & \ldots & \{v+(g-1)w\}^{-1} \\ (v-w)^{-1}x_{11} & \ldots & (v-w)^{-1}x_{g1} \\ . & & \\ . & & \\ (v-w)^{-1}x_{1k} & \ldots & (v-w)^{-1}x_{gk} \end{bmatrix} \tag{5.27}$$

Since

$$(\underset{\sim 1}{X}'\underset{\sim 1}{X})^{-1} = \begin{bmatrix} g^{-1} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{M}^{-1} \end{bmatrix},\quad \underset{\sim 1}{X}' = \begin{bmatrix} 1 & \ldots & 1 \\ x_{11} & \ldots & x_{g1} \\ . & & \\ . & & \\ x_{1k} & \ldots & x_{gk} \end{bmatrix} \tag{5.28}$$

it follows that

$$(\underset{\sim 1}{X}'\underset{\sim}{W}^{-1}\underset{\sim 1}{X})^{-1}\underset{\sim 1}{X}'\underset{\sim}{W}^{-1} = (\underset{\sim 1}{X}'\underset{\sim 1}{X})^{-1}\underset{\sim 1}{X}' \tag{5.29}$$

showing that the OLS esimate of $\underset{\sim}{\beta}$ based on the $\{\hat{\mu}_{*i}\}$ provides the BLUE. It should be stressed that this result only applies when the sample sizes are equal and the same degree of censoring occurs ia each sample. However, the result leads to a useful simplification for computation of the estimates.

Using (5,23) and (5.27) gives

$$\operatorname{var}(\hat{\beta}_{*0}) = g^{-1}\theta^2\{v + (g-1)w\}, \quad \operatorname{cov}(\hat{\beta}_{*1}) = \theta^2 (v-w)\underset{\sim}{M}^{-1} \tag{5.30}$$

where $\hat{\underset{\sim}{\beta}}_{*1} = (\hat{\beta}_{*1}, \ldots, \hat{\beta}_{*k})$.

Values of the exact variances ($\times\, \theta^{-2}$) of the BLUE's $\hat{\theta}*$, $\hat{\beta}_{*0}$ and $\hat{\beta}_{*i}$ have been computed using (5.15) and (5.30) for the simple linear regression model $Y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}$, $i = 1,\ldots,g$, $j - 1,\ldots,m_i$ with $x. = i - ½(g+1)$.The uncensored case with $r_i. = m_i = m$ was considered with $m = 5(5)20$ and $g = 5,10$. The values of $W^{(1)}_{m,m}$, $W^{(2)}_{m,m}$ and $w^{(3)}_{m,m}$ needed for the calculations were taken from White's tables. The results are shown in table 7.

Table 7

Exact variances (x $10^2\ 9^{-2}$) of the BLUE's $\hat{\beta}_{*0}$, $\hat{\beta}_{*1}$, $\hat{\theta}*$ for the simple linear regression model

	g = 5			g = 10		
m	$\hat{\beta}_{*0}$	$\hat{\beta}_{*1}$	$\hat{\theta}*$	$\hat{\beta}_{*0}$	$\hat{\beta}_{*1}$	$\hat{\theta}*$
2	16.449	6.596	14.237	8.225	0.800	7.119
5	6.523	2.314	3.333	3.262	0.281	1.667
10	3.244	1.133	1 .432	1.622	0.137	0.716
15	2.158	0.748	0.907	1.079	0.091	0.453
20	1.616	0.559	0.663	0.808	0.068	0.331

To examine the shape characteristics of the distributions of the BLUE's, their empirical distributions were obtained by simulation, using a run-size of 4000 in each case. Histogram plots of the distributions are shown in figures 7, 8 and 9. The values of the observed skewness and kurtosis coefficients are also given.

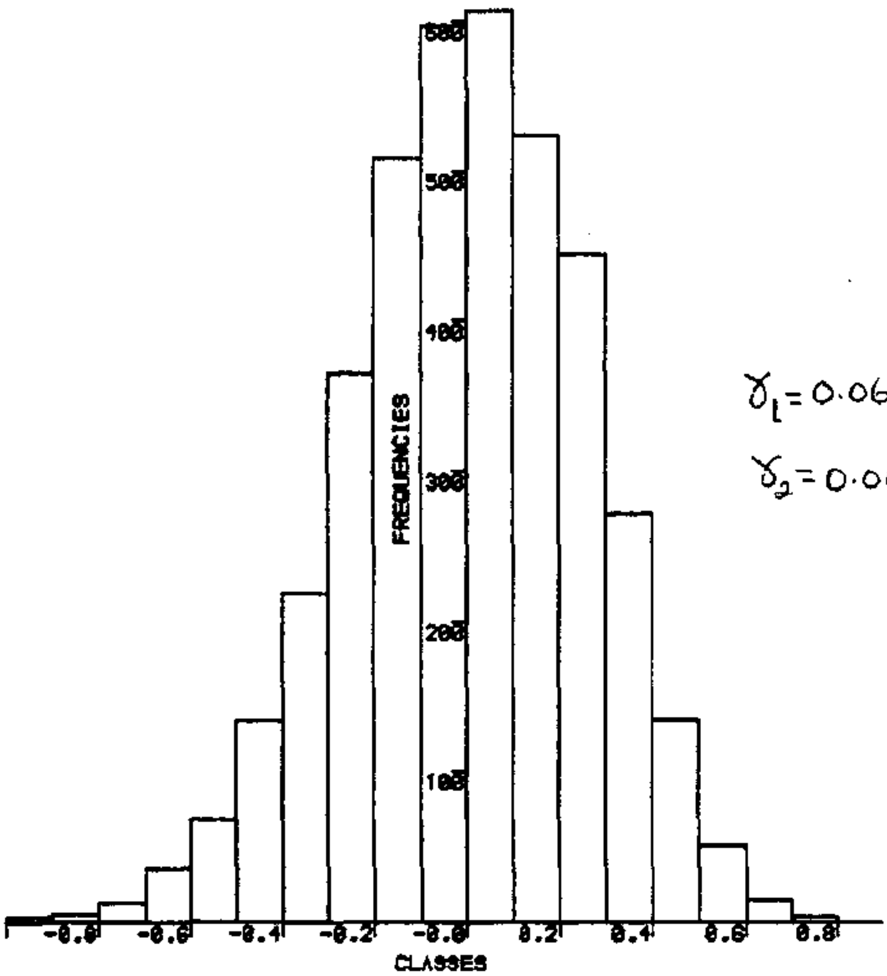

FIGURE (3.a) : HISTOGRAM PLOT FOR BLU ESTIMATE $\hat{\beta}0^{*}$, m=5, g=5

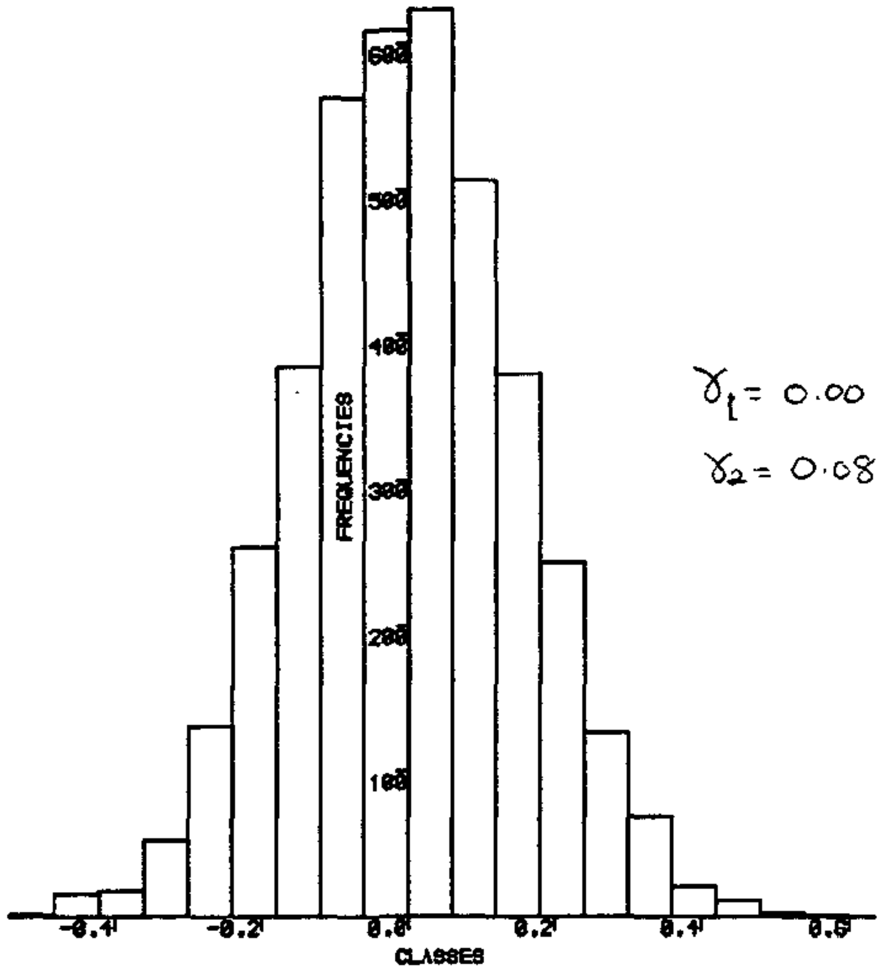

FIGURE (3.b) : HISTOGRAM PLOT FOR BLU ESTIMATE $\hat{\beta}1^*$, m=5, g=5

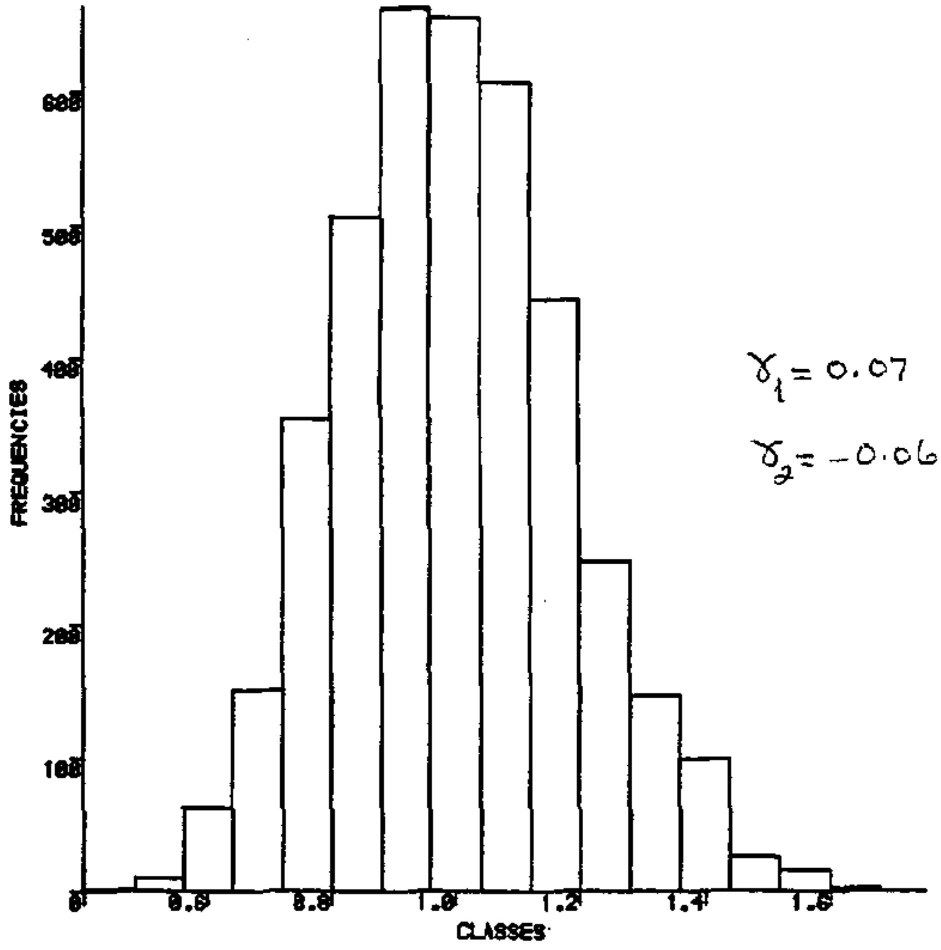

FIGURE (3.c) : HISTOGRAM PLOT FOR BLU ESTIMATE $\hat{\theta}^*$, m=5, g=5

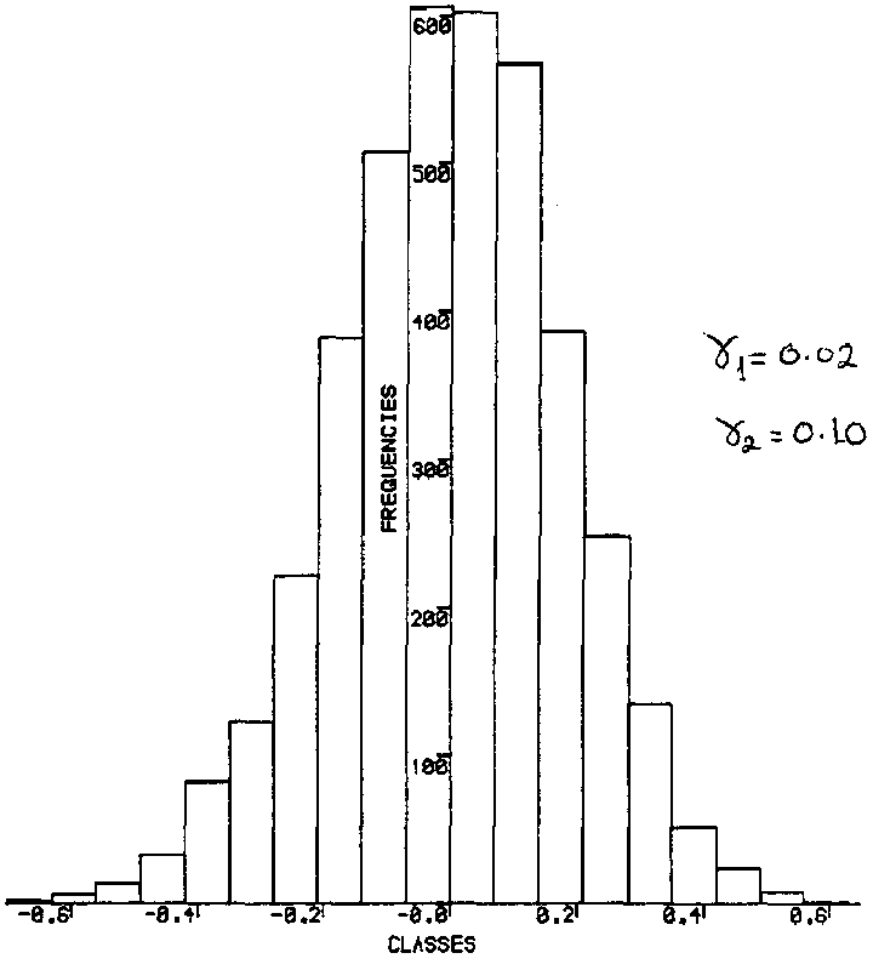

FIGURE (3.d) : HISTOGRAM PLOT FOR BLU ESTIMATE $\hat{\beta}0^*$, m=10, g=5

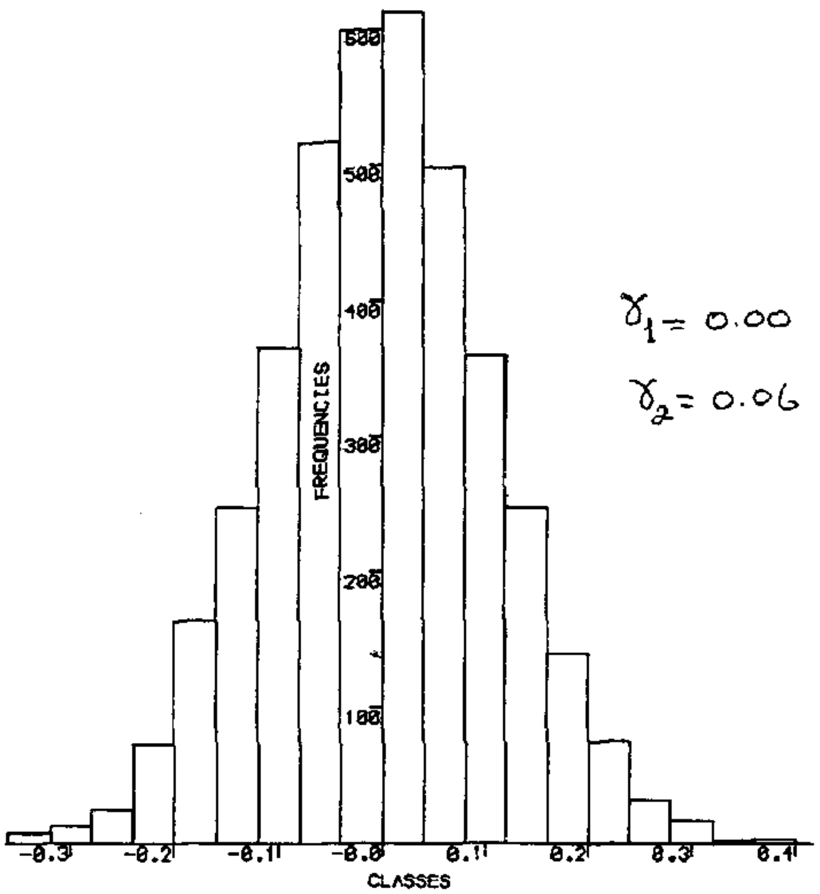

FIGURE (3.e) : HISTOGRAM PLOT FOR BLU ESTIMATE $\hat{\beta}_1^*$, m=10, g=5

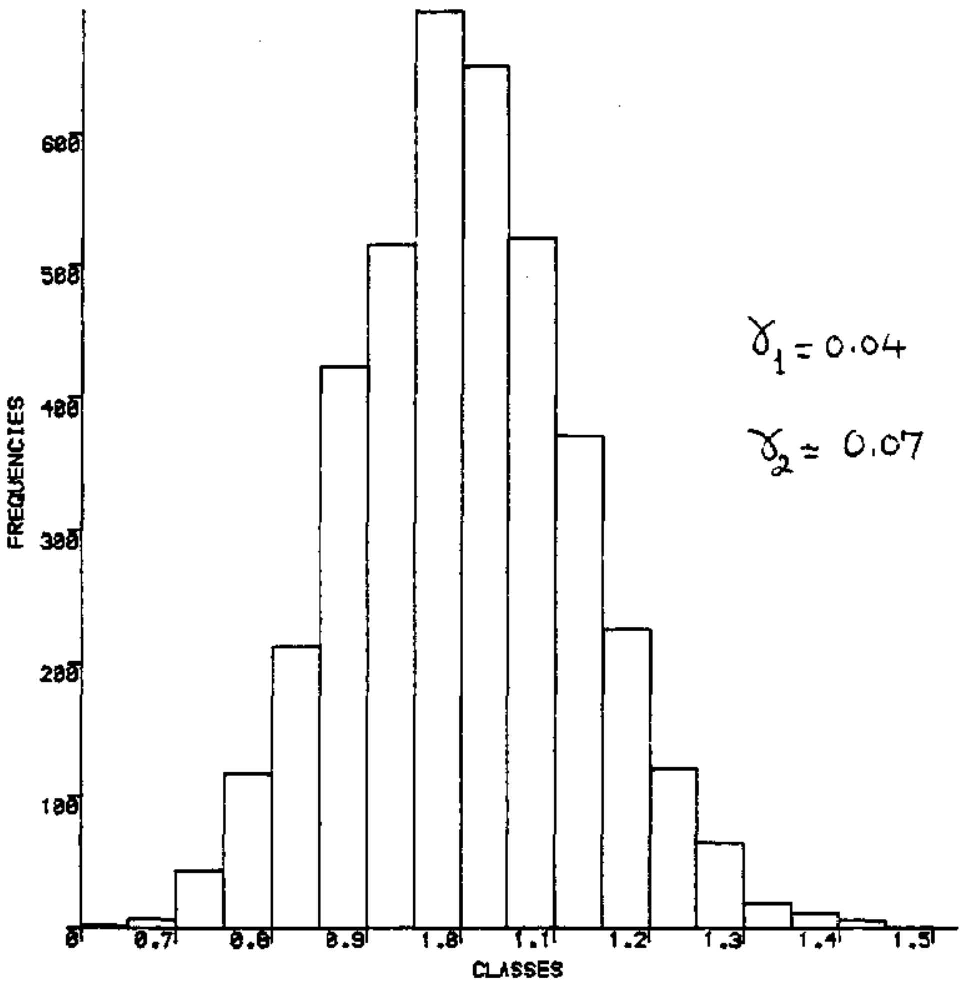

FIGURE (3.F) : HISTOGRAM PLOT FOR BLU ESTIMATE $\hat{\theta}^*$, m=10, g=5

6. Small Sample Variance Efficiency Results

The variance results given in the previous sections enable us to examine the small sample efficiencies of the OLS estimators and the BLUE's relative to the ML estimators and to assess how rapidly they approach the asymptotic efficiencies.

We let

$$E_{1\theta} = \mathrm{var}(\hat{\theta}) / \mathrm{var}(\tilde{\theta}) , \qquad E_{1r} = \mathrm{var}(\hat{\beta}_r) / \mathrm{var}(\tilde{\beta}_r) \qquad (6.1)$$

denote the efficiencies of the OLS estimators relative to the ML estimators, for θ and β_r respectively, $r = 0,1,...,k$. The corresponding efficiencies of the BLUE's based on order statistics for grouped data, relative to the ML estimators will be denoted by

$$E_{2\theta} = \mathrm{var}(\hat{\theta}) / \mathrm{var}(\hat{\theta}_*) , \qquad E_{2r} = \mathrm{var}(\hat{\beta}_r) / \mathrm{var}(\hat{\beta}_{*r}) , \qquad (6.2)$$

for $r = 0,1,...,k$.

Using the variance results given in (3.3), (3.4), (4.2) and (4.3), the asymptotic efficiences of the OLS estimators $\tilde{\beta}_r$ relative to $\hat{\beta}_r$ are

$$E^{(a)}_{1\theta} = 0.978 , \qquad E^{(a)}_{1r} = 6/\pi^2 = 0.608 , \; r = 0,1,,..,k . \qquad (6.3)$$

If we assume that $\lim \sum_{i=1}^{n} (1 - h_{ii})^2 / (n - k - 1) = 1$ as $n \to \infty$, use of (3.3) and (4.11) gives

$$E^{(a)}_{1\theta} = 0.553. \qquad (6.4)$$

Values of -the small sample variance efficiences have been computed for the simple linear regression model $Y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}$, $i = 1,...,g$, $j = 1,...,m$, for grouped data with equal sample sizes. The exact variances for $\tilde{\beta}_0$, $\tilde{\beta}_1$, $\tilde{\beta}_{*0}$ $\tilde{\beta}_{*1}$., and θ, which are known for all sample sizes were used, while estimated variances obtained by simulation were used for $\tilde{\theta}$, $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\theta}$. The results are shown in table 8.

Table 8

Variance efficiencies of OLS estimators and BLUE's relative to ML estimators for simple linear regression

	m	E_{10}	E_{20}	E_{11}	E_{21}	$E_{1\theta}$	$E_{2\theta}$
	1	0.99	-	0.87	-	0.55	-
	2	0.95	0.95	0.69	0,87	0.61	0.40
	5	1 .00	1 .01	0.65	0.92	0.62	0.73
g=5	10	0.96	0.97	0.62	0.90	0.60	0.87
	15	0.97	0.99	0.62	0,91	0.57	0.91
	20	0.96	0.98	0.62	0.91	0.56	0.90
	1	0.96	-	0.75	-	0.62	-
	2	0.96	0.96	0,67	0,84	0.60	0.43
g=10	10	0.96	0.97	0.65	0,93	0.58	0.74
		0.97	0.98	0.62	0.91	0.57	0.85
	15	0.98	1.00	0.62	0.91	0.54	0.91
	20	1.00	1.02	0.61	0.90	0.58	0.95

The following broad conclusions can be drawn from the results in table 8,

a) After allowance for the simulation errors, the efficiency values for OLS for β_0 and θ appear to converge very rapidly to the asymptotic values 0.978 and 0,553, respectively. For β_1 , the OLS efficiency is appreciably higher than the asymptotic value 0.608 when m = 1,2.

b) The efficiency of BLUE for β_1 is much higher than the OLS efficiency for all m and exceeds 90% for $m \geq 5$. For estimation of θ, BLUE is less efficient than OLS when m = 2, but for higher values of m its performance is much better than that of OLS.

References

Antle, C.L. and Bain, L.J. (1969). A property of maximum likelihood estimators of location and scale parameters. SIAM Review, 11, 251-253.

Atiqullah, M. (1962). Theestimation of residual variance in quadratica1ly balanced least-squares problems and the robustness of the F test. Bimetrika, 49, 83-91.

Cox, D.R. and Hinkley, D.V. (1968). The efficiency of least squares estimates. J.R.S.S., Series B, 30, 284-290.

Cox, D.R. and Snell, E.J. (1968). A general definition of residuals. J.R.S.S., Series B, 30, 248-265.

Lawless, J.F. (1982). Statistical Models and Methods for Lifetime Data. New York: Wiley.

Mann, N.R. (1968). Point and interval estimation procedures for the two-parameter Weibull and Extreme-value distributions. Technometrics, 10, 231-256.

Roger, J.H. and Peacock, S.D. (1982). Fitting the scale as a glim parameter for Weibull, Extreme-value, Logistic and Log-logistic regression models with censored data. GLIM Newsletter, 6, 30-37.

Scheffé, H. (1959). The Analysis of Variance, New York: Wiley.

White, J.S. (1964). Least squares unbiased censored linear estimation for the log Weibull (Extreme Value) distribution. J. Industrial Math. Soc., 14, 21-60.