ON LEAST SQUARES ESTIMATION WHEN THE DEPENDENT VARIABLE IS GROUPED

Mark B. Stewart

NUMBER 207

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

ON LEAST SQUARES ESTIMATION WHEN THE DEPENDENT VARIABLE IS GROUPED

Mark B. Stewart

NUMBER 207

September 1981
Revised: April 1982

I : Introduction

Models estimated from censored samples are now familiar in the econometrics literature. For many cases Least Squares approximations to the Maximum Likelihood estimators are nowwell established. This paper is concerned with a more general problem; that of estimating an equation on the basis of data in which the dependent variably is only observed to fall in a certain range on a continuous scale, its actual value remaining unobserved. The data arealso censored in the usual sense in that both end ranges are assumed to be open-ended. A number of Least Squares approximations to the Maximum Likelihood estimator are derived and compared. The results of Greene (1981) on the asymptotic bias of OLS are extended to this case. The question of information loss as a result of the grouping is also considered.

The latent structure of the model to be considered is assumed to be given by

$$y_i = \underset{\sim}{x}_i' \underset{\sim}{\beta} + u_i \qquad (i = 1, \ldots, N),$$

where y_i is the unobserved dependent variable, $\underset{\sim}{x}_i$ and $\underset{\sim}{\beta}$ are both $J \times 1$ vectors, the former being regressors and the latter unknown parameters. The u_i are assumed to be independent identically normally distributed random variables with zero mean and variance σ^2 and to be independent of $\underset{\sim}{x}_i$. The conditional distribution of the unobserved dependent variable is given by

$$y_i | \underset{\sim}{x}_i \sim N(\underset{\sim}{x}_i' \underset{\sim}{\beta}, \sigma^2) \qquad i = 1, \ldots, N.$$

The observed information concerning the dependent variable is that it falls into a certain range of the real line. The real line is divided into K ranges, the k-th being given by $(A_{k-1}, A_k]$. It is further allowed (although this need not be the case) that these K ranges exhaust the real line. Thus $A_0 = -\infty$ and $A_K = +\infty$, i.e. the first and K-th ranges are "open-ended". The observed information concerning the dependent variable is which of these K ranges it falls into, i.e. an indicator variable k_i is observed for each $i (1 \leq k_i \leq K)$.

This type of problem is encountered in the analysis of certain variables on a number of data sets. The one which prompted the investigation on which this paper is based is the earnings variable in the National Training Survey. (See Manpower Services Commission (1978) for details.) This survey, with its detailed employment, occupational and training histories, is fast becoming a major source for U.K. economists and its use will no doubt increase in the future as it becomes even more widely available. Simple techniques for the analysis of its earnings variable are thus urgently needed. The earnings variable in the Oxford Mobility Survey is also grouped in a similar way and the analysis of a number of variables in the General Household Survey (see Office of Population Censuses and Surveys (1978) for details) give rise to this type of problem. In particular housing expenses, the length of time with the present employer and duration of unemployment are all grouped in that survey.

Analysis of these large survey data sets is usually undetaken on one of the commonly available general statistical packages. The sample sizes involved severely restrict the range of software that is available. The National Training Survey, for example, contains approximately 54,000 observations.

When Maximum Likelihood or other iterative routines are available on these packages the sample size restrictions are usually such as to rule out their use with these data sets. Even when not so ruled out or when special programs are available the cost can often be prohibitive. Thus fast and easy one- or two-step Least Squares techniques are very useful in this work. This paper provides and illustrates such techniques for the problem under consideration. Further, since it might be thought that the usefulness of these variables is severely limited by the grouping, the paper seeks to investigate the extent of information loss as a result of the grouping in the context of the use to which the variable is to be put.

Ad-hoc Least Squares estimation might involve assignment of observations in any given group the midpoint (possibly after transformation of the variable), with the open-ended groups being assigned values on some even more ad-hoc basis. However such methods do not in general result in consistent estimates. Consistent estimates would be obtained by assigning each observation its conditional expectation,

$$E(y_i \mid A_{k-1} < y \leq A_k, \underset{\sim}{x}_i) = \underset{\sim}{x}_i' \underset{\sim}{\beta} + \sigma \left[\frac{f(Z_{k-1}) - f(Z_k)}{F(Z_k) - F(Z_{k-1})}\right]$$

where $Z_k = (A_k - \underset{\sim}{x}_i' \underset{\sim}{\beta})/\sigma$, f is the standard normal density function and F is its cumulative distribution. Hence the requisite estimation of the conditional expectations requires estimates of β and σ. However, as will be seen in the next section this approach provides a convergent maximum likelihood algorithm and hence possibilities for least squares approximations.

The remainder of this paper is laid out as follows. Section II defines the Maximum Likelihood estimates of the parameters in the model under consideration and demonstrates an algorithm based on Least Squares that will attain these Maximum Likelihood estimates and converge monotonically. Section III dervies a "moment" estimator for the normal regressors case, extending the recent work of Olsen (1980) and Greene (1981). Section IV then considers a number of Least Squares approximations involving the moment estimator in conjunction with early termination of the convergent algorithm. These Least Squares approximations and the full Maximum Likelihood are then illustrated and compared in Section V by the estimation of earnings equations using NTS data. In addition the extent of information loss as a consequence of the grouping is examined by comparing earnings equations based on GHS data grouped for the purpose with those estimated from the original data. In Section VI the results of a number of simulation experiments on these methods are presented in an attempt to assess the sensitivity of the estimators to the properties of the sample data and the underlying model. Section VII presents some conclusions.

II : Maximum Likelihood Estimation

The log likelihood function of the problem outlined in the previous section is given by

$$\log L = \sum_{k=1}^{K} \sum_{i \epsilon k} \log \left[F\left(\frac{A_k - x_i'\beta}{\sigma}\right) - F\left(\frac{A_{k-1} - x_i'\beta}{\sigma}\right) \right] .$$

We omit the i subscripts and further simplify the notation in an obvious way to give

$$\log L = \sum_{k=1}^{K} \sum_{i \epsilon k} \log(F_k - F_{k-1}) .$$

The first-order partials with respect to the parameters are given by:

$$\frac{\partial \log L}{\partial \beta_j} = \sum_i \frac{x_{ij}}{\sigma} \left[\frac{f_{k-1} - f_k}{F_k - F_{k-1}} \right] \qquad j = 1, \ldots, J$$

$$\frac{\partial \log L}{\partial \sigma} = \sum_i \frac{1}{\sigma} \left[\frac{Z_{k-1} f_{k-1} - Z_k f_k}{F_k - F_{k-1}} \right]$$

where in simplified notation compatible with that used above $Z_k = \left(\frac{A_k - x_i'\beta}{\sigma}\right)$, $f_k = f(Z_k)$ and f is the p.d.f. of the standard normal.

Hence the maximum likelihood estimates are defined by the set of equations

$$\sum_i x_{ij} \left[\frac{f_{k-1} - f_k}{F_k - F_{k-1}}\right] = 0 \qquad j = 1, \ldots, J$$

$$\sum_i \left[\frac{Z_{k-1} f_{k-1} - Z_k f_k}{F_k - F_{k-1}}\right] = 0$$

A number of different algorithms may be used to obtain these Maximum Likelihood estimates. This section concentrates on the derivation of one that requires only OLS at each iteration.

The conditional means of the unobserved y_i are given by

$$m_i = \left[E\, y_i \middle| k_i; \underset{\sim}{\beta}, \sigma^2\right] = \underset{\sim}{x}_i' \underset{\sim}{\beta} + \sigma\left[\frac{f_{k-1} - f_k}{F_k - F_{k-1}}\right]$$

(All expectations in this section are also conditional on $\underset{\sim}{x}_i$, although it is omitted from the notation for the sake of simplicity).

Hence the first J of the first-order conditions can be written as

$$\sum_i x_{ij} \left(E\left[y_i \middle| k_i; \hat{\underset{\sim}{\beta}}, \hat{\sigma}^2\right] - \underset{\sim}{x}_i' \hat{\underset{\sim}{\beta}}\right) = 0 \qquad j = 1, \ldots, J$$

or as

$$\sum_i (x_{ij} \hat{m}_i - x_{ij} \underset{\sim}{x}_i' \hat{\underset{\sim}{\beta}}) = 0 \qquad j = 1, \ldots, J$$

or in terms of obviously defined matrices and vectors as

$$\underset{\sim}{X}'\hat{\underset{\sim}{m}} - \underset{\sim}{X}'\underset{\sim}{X}\,\hat{\underset{\sim}{\beta}} = 0$$

Hence given estimates of the conditional expectations, an estimate of the β-vector is given by:

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'\,\hat{\underset{\sim}{m}}$$

Turning our attention to the final first-order condition, this can be rearranged to give an estimate of σ^2 in terms of the "residual"-sum-of-squares from this least-squares regression as follows.

The conditional expectation of y_i^2 is given by

$$E(y_i^2|k_i\,;\,\beta,\,\sigma^2) = \sigma^2\left[\frac{Z_{k-1}f(Z_{k-1}) - Z_k f(Z_k)}{F(Z_k) - F(Z_{k-1})}\right] + \sigma^2 + (\underset{\sim i}{x}'\,\underset{\sim}{\beta})^2 + 2\sigma(\underset{\sim i}{x}'\,\underset{\sim}{\beta})\left[\frac{f(Z_{k-1}) - f(Z_k)}{F(Z_k) - F(Z_{k-1})}\right]$$

and so the conditional variances of the y_i are given by

$$E(y_i^2|k_i\,;\,\beta,\,\sigma^2) - [E(y_i|k_i\,;\,\beta,\,\sigma^2)]^2$$

$$= \sigma^2\left\{\left[\frac{Z_{k-1}f(Z_{k-1}) - Z_k f(Z_k)}{F(Z_k) - F(Z_{k-1})}\right] - \left[\frac{f(Z_{k-1}) - f(Z_k)}{F(Z_k) - F(Z_{k-1})}\right]^2 + 1\right\}$$

$$= \sigma^2 v_i \text{ , say.}$$

Hence the final first-order condition can be written as

$$\sum_i(\hat{\sigma}^2\,\hat{v}_i + (\hat{m}_i - \underset{\sim i}{x}'\,\hat{\underset{\sim}{\beta}})^2 - \hat{\sigma}^2) = 0$$

Thus given estimates of the conditional expectations and hence the β-vector an estimate of σ^2 is given by :

$$\hat{\sigma}^2 = \frac{1}{\hat{d}} \Sigma_i (\hat{m}_i - \underset{\sim}{x}_i' \hat{\underset{\sim}{\beta}})^2$$

where $d = \Sigma_i (1 - v_i)$

$$= \Sigma_i \left\{ \left[\frac{f(Z_{k-1}) - f(Z_k)}{F(Z_k) - F(Z_{k-1})} \right]^2 - \left[\frac{Z_{k-1}(Z_{k-1}) - Z_k f(Z_k)}{F(Z_k) - F(Z_{k-1})} \right] \right\}$$

Hence the likelihood conditions can be solved by iterating between $\hat{\underset{\sim}{m}}$ and $(\hat{\underset{\sim}{\beta}}, \hat{\sigma}^2)$.

Since d can clearly be expressed in terms of conditional expectations of sufficient statistics for the y_i , this iterative method for solving the likelihood conditions can be seen to be an application of the EM Algorithm discussed by Dempster et al. (1977). Hence convergence is guaranteed, and the likelihood is increased at <u>each</u> iteration.

The main advantage of the method, over for example Newton-Raphson, lies in its simplicity. It is purely a series of OLS estimations. In addition, since the cross-product matrix $X'X$ does not change from one iteration to the next, only one matrix inversion, or equivalent, is required, in contrast to Newton-Raphson where evaluation and inversion of the matrix of second derivatives is required at each iteration.

The Maximum Likelihood estimates are consistent and asymptotically efficient and asymptotic standard errors can be obtained

by inverting the matrix of second derivatives after convergence has been attained. The second derivatives of the log-likelihood function are given by

$$\frac{\partial^2 \log L}{\partial \beta_j \, \partial \beta_h} = \frac{1}{\sigma^2} \sum_i x_{ij} \, x_{ih} \left\{ \frac{Z_{k-1} f_{k-1} - Z_k f_k}{F_k - F_{k-1}} - \left[\frac{f_{k-1} - f_k}{F_k - F_{k-1}} \right]^2 \right\}$$

$$\frac{\partial^2 \log L}{\partial \beta_j \, \partial \sigma} = \frac{1}{\sigma^2} \sum_i x_{ij} \left\{ \frac{Z_{k-1}^2 f_{k-1} - Z_k^2 f_k}{F_k - F_{k-1}} - \frac{f_{k-1} - f_k}{F_k - F_{k-1}} - \left[\frac{Z_{k-1} f_{k-1} - Z_k f_k}{F_k - F_{k-1}} \right] \left[\frac{f_{k-1} - f_k}{F_k - K_{k-1}} \right] \right\}$$

$$\frac{\partial^2 \log L}{\partial \sigma^2} = \frac{1}{\sigma^2} \sum_i \left\{ \frac{Z_{k-1}^3 f_{k-1} - Z_k^3 f_k}{F_k - F_{k-1}} - 2 \left[\frac{Z_{k-1} f_{k-1} - Z_k f_k}{F_k - F_{k-1}} \right] - \left[\frac{Z_{k-1} f_{k-1} - Z_k f_k}{F_k - F_{k-1}} \right]^2 \right\}$$

These can be written more compactly by defining

$$M_q = \frac{Z_{k-1}^q f_{k-1} - Z_k^q f_k}{F_k - F_{k-1}}$$

The second derivatives are then given by

$$\frac{\partial^2 \log L}{\partial \beta_j \, \partial \beta_h} = \frac{1}{\sigma^2} \sum_i x_{ij} \, x_{ih} \left\{ M_1 - M_0^2 \right\}$$

$$\frac{\partial^2 \log L}{\partial \beta_j \, \partial \sigma} = \frac{1}{\sigma^2} \sum_i x_{ij} \left\{ M_2 - M_0 - M_1 M_0 \right\}$$

$$\frac{\partial^2 \log L}{\partial \sigma^2} = \frac{1}{\sigma^2} \sum_i \left\{ M_3 - 2M_1 - M_1^2 \right\}$$

Since $\sum_i x_{ij} M_0 = 0$ $(j = 1, \ldots, J)$ and $\sum_i M_1 = 0$ at the maximum of the likelihood function, the middle terms in these last two can be omitted when they are being evaluated at the Maximum Likelihood solution. Hence estimated asymptotic variances and covariances are given by inverting the $(J + 1) \times (J + 1)$ matrix defined by

$$Q_{jh} = \frac{1}{\hat{\sigma}^2} \sum_i x_{ij} x_{ih} \left\{ \hat{M}_0^2 - \hat{M}_1 \right\} \qquad j, h = 1, \ldots, J$$

$$Q_{jJ+1} = \frac{1}{\hat{\sigma}^2} \sum_i x_{ij} \left\{ \hat{M}_1 \hat{M}_0 - \hat{M}_2 \right\} \qquad j = 1, \ldots, J$$

$$Q_{J+1J+1} = \frac{1}{\hat{\sigma}^2} \sum_i \left\{ \hat{M}_1^2 - \hat{M}_3 \right\}$$

where $\hat{}$'s indicate evaluation at the Maximum Likelihood estimates.

III : A Moment Estimator for the Normal Regressors Case

"Moment" estimators have been proposed recently for both the truncated regression model (Olsen (1980)) and the Tobit model (Green (1981)). This section derives such an estimator for the grouped dependent variable model. It is consistent in the case when the regressors are normally distributed. In passing, the results of Greene (1981) on the asymptotic bias of OLS are extended to the grouped dependent variable model.

The latent structure of the model under consideration is rewritten as

$$y_i = \alpha + \underset{\sim}{x}_i' \underset{\sim}{\gamma} + u_i \qquad (i = 1,\ldots,N)$$

where $\underset{\sim}{x}_i$ now excludes the constant term. It is assumed that $\underset{\sim}{x}_i$ is normally distributed. Thus

$$\begin{bmatrix} y_i \\ \underset{\sim}{x}_i \end{bmatrix} \sim N\left[\begin{pmatrix} \mu_y \\ \underset{\sim}{\mu}_x \end{pmatrix}, \begin{pmatrix} \sigma_y^2 & \underset{\sim}{\sigma}_{xy}' \\ \underset{\sim}{\sigma}_{xy} & \Sigma_{xx} \end{pmatrix}\right], \quad (i=1,\ldots,N)$$

Estimation of this equation by one Least Squares step would involve assigning a value for the "dependent variable" for all observations in a given group. Let the assigned values be q_k $(k = 1,\ldots,K)$ and let g be the "dependent variable" defined in this way. Thus

$$g_i = q_k \quad \text{if} \quad A_{k-1} < y_i \leq A_k \qquad (k = 1,\ldots,K)(i = 1,\ldots,N) .$$

The OLS regression of g on $\underset{\sim}{x}$ produces the following estimates:

$$\underset{\sim}{c} = \underset{\sim}{S}_{xx}^{-1} \underset{\sim}{S}_{xg}$$

$$a = \bar{g} - \underset{\sim}{\bar{x}}' \underset{\sim}{c}$$

$$s^2 = S_{gg} - \underset{\sim}{S}'_{xg} \underset{\sim}{c}$$

where S_{xx}, S_{xg} and S_{gg} are the appropriate sample moments, which tend in probability to their population equivalents.

To examine the inconsistency in the OLS estimates, some moments of the observed random variables must first be derived.

$$E(g) = \sum_{k=1}^{K} q_k P(A_{k-1} < y \leq A_k)$$

$$= \sum_{k=1}^{K} q_k \left\{ F\left(\frac{A_k - \mu_y}{\sigma_y}\right) - F\left(\frac{A_{k-1} - \mu_y}{\sigma_y}\right) \right\} .$$

$$E(g^2) = \sum_{k=1}^{K} q_k^2 \left\{ F\left(\frac{A_k - \mu_y}{\sigma_y}\right) - F\left(\frac{A_{k-1} - \mu_y}{\sigma_y}\right) \right\} .$$

$$E(\underset{\sim}{x}g) = \sum_{k=1}^{K} q_k E(\underset{\sim}{x}|g = q_k)P(g = q_k)$$

$$= \sum_{k=1}^{K} q_k E(\underset{\sim}{x}|A_{k-1} < y \leq A_k)P(A_{k-1} < y \leq A_k).$$

The conditional expectation here is given by

$$E(\underset{\sim}{x}|A_{k-1} < y \leq A_k) = \underset{\sim}{\mu}_x + (\underset{\sim}{\sigma}_{xy}/\sigma_y^2)\left\{E(y|A_{k-1} < y \leq A_k) - \mu_y\right\}$$

$$= \underset{\sim}{\mu}_x + \frac{\underset{\sim}{\sigma}_{xy}}{\sigma_y}\left(\frac{f(B_{k-1}) - f(B_k)}{F(B_k) - F(B_{k-1})}\right)$$

where $B_k = \left(\frac{A_k - \mu_y}{\sigma_y}\right)$

whilst $P(A_{k-1} < y \leq A_k) = F(B_k) - F(B_{k-1})$.

Thus

$$E(\underset{\sim}{x}g) = \sum_{k=1}^{K} q_k \left\{\underset{\sim}{\mu}_x(F(B_k) - F(B_{k-1})) + (\underset{\sim}{\sigma}_{xy}|\sigma_y)(f(B_{k-1}) - f(B_k))\right\}.$$

Given these moments, the probability limits of the OLS estimates can be found as follows

$$\text{plim } \underset{\sim}{S}_{xg} = \text{cov}(\underset{\sim}{x}, g)$$

$$= E(\underset{\sim}{x}g) - \underset{\sim}{\mu}_x E(g)$$

$$= \sum_{k=1}^{K} q_k \frac{\underset{\sim}{\sigma}_{xy}}{\sigma_y} \left\{ f(B_{k-1}) - f(B_k) \right\}$$

$$= \underset{\sim}{\sigma}_{xy} \sum_{k=1}^{K} q_k \frac{1}{\sigma_y} \left\{ f(B_{k-1}) - f(B_k) \right\}$$

$$\text{plim } \underset{\sim}{S}_{xx} = \underset{\sim}{\Sigma}_{xx}$$

and $\underset{\sim}{\gamma} = \underset{\sim}{\Sigma}_{xx}^{-1} \underset{\sim}{\sigma}_{xy}$.

Thus

$$\text{plim } \underset{\sim}{c} = \text{plim } (\underset{\sim}{S}_{xx}^{-1} \underset{\sim}{S}_{xg})$$

$$= \underset{\sim}{\gamma} \sum_{k=1}^{K} q_k \frac{1}{\sigma_y} \left\{ f(B_{k-1}) - f(B_k) \right\} \neq \underset{\sim}{\gamma} \text{ , in general.}$$

Thus all the OLS slope coefficient estimates are inconsistent by the same proportion.

Turning to a,

$$\text{plim } a = E(g) - \underset{\sim}{\mu}_x' \text{ plim } \underset{\sim}{c}$$

$$= \sum_{k=1}^{K} q_k \left\{ F(B_k) - F(B_{k-1}) \right\} - \underset{\sim}{\mu}_x' \underset{\sim}{\gamma} \sum_{k=1}^{K} q_k \frac{1}{\sigma_y} \left\{ f(B_{k-1}) - f(B_k) \right\}$$

$$= \sum_{k=1}^{K} q_k \left\{ F(B_k) - F(B_{k-1}) \right\} + (\alpha - \mu_y) \sum_{k=1}^{K} q_k \frac{1}{\sigma_y} \left\{ f(B_{k-1}) - f(B_k) \right\} .$$

At this point it is convenient to define the following scalars

$$\lambda = \sum_{k=1}^{K} q_k \left\{ F(B_k) - F(B_{k-1}) \right\}$$

$$\psi = \sum_{k=1}^{K} q_k^2 \left\{ F(B_k) - F(B_{k-1}) \right\}$$

$$\theta = \sum_{k=1}^{K} q_k \frac{1}{\sigma_y} \left\{ f(B_{k-1}) - f(B_k) \right\}.$$

Note that the unknown parameters involved in each case are μ_y and σ_y.

Then

$$\text{plim } \underset{\sim}{c} = \underset{\sim}{\gamma} \theta$$

and

$$\text{plim } a = \lambda + (\alpha - \mu_y)\, \theta.$$

Finally turning to s^2,

$$\text{plim } S_{gg} = \text{Var}(g)$$

$$= \psi - \lambda^2$$

and

$$\text{plim } \underset{\sim}{S}_{xg} = \underset{\sim}{\sigma}_{xy}\, \theta.$$

Thus

$$\text{plim } s^2 = \psi - \lambda^2 - \theta \underset{\sim}{\sigma}'_{xy} \text{ plim } \underset{\sim}{c}$$

$$= \psi - \lambda^2 - \theta^2 \underset{\sim}{\sigma}'_{xy} \underset{\sim}{\gamma}$$

$$= \psi - \lambda^2 - \theta^2 \sigma_y^2 \rho^2$$

where ρ^2 is the multiple correlation between y and $\underset{\sim}{x}$.

Given that $\rho^2 = 1 - \frac{\sigma^2}{\sigma_y^2}$,

$$\sigma_y^2 \rho^2 = \sigma_y^2 - \sigma^2.$$

Thus

$$\text{plim } s^2 = \psi - \lambda^2 - \theta^2(\sigma_y^2 - \sigma^2) \neq \sigma^2 \text{, in general.}$$

Clearly the OLS estimates are in general inconsistent. However given consistent estimates of μ_y, σ_y consistent estimates of $\underset{\sim}{\gamma}$, α and σ^2 can easily be derived from them using the following simple adjustments.

Define

$$\hat{\underset{\sim}{\gamma}} = \underset{\sim}{c}/\hat{\theta}$$

$$\hat{\alpha} = \hat{\mu}_y + \frac{a - \hat{\lambda}}{\hat{\theta}}$$

$$\hat{\sigma}^2 = \frac{s^2 - \hat{\psi} + \hat{\lambda}^2}{\hat{\theta}^2} + \hat{\sigma}_y^2$$

where $\hat{\lambda}$, $\hat{\psi}$, $\hat{\theta}$ are λ, ψ, θ evaluated at $\hat{\mu}_y$ and $\hat{\sigma}_y$, themselves consistent estimates of μ_y and σ_y. Then $\hat{\underset{\sim}{\gamma}}$, $\hat{\alpha}$, $\hat{\sigma}^2$ are consistent estimates of $\underset{\sim}{\gamma}$, α, σ^2 respectively. Thus for any relevant choice of $q_k (k = 1,\ldots,K)$ consistent estimation of $(\underset{\sim}{\gamma}, \alpha, \sigma^2)$ requires only consistent estimation of μ_y and σ_y in addition to the OLS estimates.

IV : Least Squares Approximations

Maximum Likelihood estimation of the model under consideration can be extremely expensive on computer time, particularly when large samples are involved. Hence convenient Least Squares approximations to the full Maximum Likelihood solutions are desirable. A consistent estimate needing only simple adjustments to any OLS estimates was presented in Section III. However the result demonstrated there, on which the estimator is based, that the OLS slope estimates are all inconsistent by the same proportion, assumes normally distributed regressors. In the absence of such normality the fact that the moment estimator adjusts all the slope coefficient estimates by the same proportion is likely to be a weakness, since the proportional inconsistencies will not in general be equal.

The monotonic convergence property of the algorithm outlined in Section II means that Least Squares approximations to the full Maximum Likelihool solutions can also be obtained simply by early termination of this algorithm. However this places great emphasis on the starting point of the algorithm, particularly if only one or two iterations are then performed to give the approximation.

Hence in both cases a combination of the methods from Sections II and III will be beneficial. An interation of the monotonically convergent algorithm will improve on the moment estimator (in the sense of increasing the value of the likelihood) and is likely to be particularly useful when the required adjustments to the OLS slope coefficients are not proportional. On the other hand the moment estimator can provide the necessary starting values for the iterative method.

The moment estimator adjustments can be applied to the OLS estimates of an equation based on any appropriate $(q_k;\ k = 1,\ldots,K)$. The adjustment factors require only consistent estimates of μ_y and σ_y.

The q_k could be chosen in a number of ad hoc ways. Any set of values satisfying

$$A_{k-1} < q_k < A_k \qquad (k = 1,\ldots,K)$$

would suffice. However given consistent estimates of μ_y and σ_y a more systematic choice of the q_k can be made based on conditional expectations of the marginal distribution. These are given by

$$E(y_i | A_{k-1} < y_i \leq A_k) = \mu_y + \sigma_y \frac{f(Z^O_{k-1}) - f(Z^O_k)}{F(Z^O_k) - f(Z^O_{k-1})}$$

where $Z^O_k = (A_k - \mu_y)/\sigma_y$. Consistent estimates of these conditional expectations can be obtained and used for the q_k,

$$q_k = \hat{\mu}_y + \hat{\sigma}_y \frac{f(\hat{Z}^O_{k-1}) - f(\hat{Z}^O_k)}{F(\hat{Z}^O_k) - F(\hat{Z}^O_{k-1})} \qquad (k = 1,\ldots,K)$$

where $\hat{Z}^O_k = (A_k - \hat{\mu}_y)/\hat{\sigma}_y$. OLS estimation of $\underset{\sim}{\beta}$ using these q_k is then equivalent to one iteration of the algorithm described in Section II except that the m_i are evaluated on the basis of consistent estimates of the parameters of the marginal distribution rather than those of the conditional distribution, the latter not being available at this stage of the procedure.

The adjustment factors for the moment estimator described in Section III are applied direct to the OLS estimates. Hence the OLS estimate of σ should be used rather than the iterative estimate using d derived in Section II. If however the initial iteration estimates are to be used on their own, or with additional iterations without the moment adjustments then the adjustment (using d) of Section II should be made.

This moment estimator is extremely convenient and simple to construct and is consistent in the case of normal regressors. However a weakness with the initial OLS estimates is that the information contained in the explanatory variables for any given observation is not utilised in the construction of the estimated conditional expectations. This is inevitable since no estimate of $\underset{\sim}{\beta}$ is available at that stage. For this reason one iteration of the Maximum Likelihood algorithm of Section II may produce considerable improvements in the approximation to the Maximum Likelihood estimates.

Hence the proposed two-step approximation involves applying one iteration of the Maximum Likelihood algorithm to the moment estimator based on the initial iteration described above. The m_i in this second iteration can be evaluated on the basis of the parameters of the conditional distribution as described in Section II. This estimator will be referred to as the "two-step estimator" and is compared with the Maximum Likelihood estimator and a number of alternative approximations in the next section.

The required consistent estimates of the parameters of the marginal distribution, μ_y and σ_y, can be obtained by fitting a normal distribution to the sample distribution of the partially observed dependent variable. One simple and convenient way of doing this, a Least Squares variant of the graphical method of Aitchison and Brown (1966); is as follows. If C_k is the sample cumulative frequency, i.e. the proportion of the sample with values of the dependent variable less than A_k, then the distribution is fitted by regressing $F^{-1}(C_k)$ on A_k. This provides consistent estimates of μ_y and σ_y. Other methods could be substituted.

V : An Illustration - The Estimation of Earnings Equations

This section illustrates the methods presented above in the context of the estimation of earnings equations. In the first illustration the "two-step estimator" and some of the others outlined in the previous section together with two ad hoc Least Squares estimators are compared both with one another and with the full Maximum Likelihood estimates on a typical earnings equation. The data source is the National Training Survey (NTS) conducted on behalf of the Manpower Services Commission in late 1975. (For details see Manpower Services Commission (1978)). The sample used here is restricted to full-time manual male employees in manufacturing, giving a sample size of 5352. The dependent variable is the logarithm of weekly earnings and the explanatory variables are as listed at the foot of Table 1. The NTS earnings variable is in ten groups each of width £10. The open-ended groups are <£25 and >£105.

The first ad hoc method used for comparison involves allocating to all individuals in a given group the mean of the logarithm of weekly earnings of the comparable sample of male workers in that range in the 1975 General Household Survey (see Office of Population Censuses and Surveys (1978) for details). The second ad hoc method used involves allocating arithmetic midpoints to the internal groups and arbitrarily taking £15 p.w. for the open-ended group with weekly earnings <£25 and £130 p.w. for the group at the other end with weekly earnings >£105.

The results of this comparative exercise are presented in Table 1. Table 1(a) presents the results for the initial iteration estimator and

Table 1 : Comparison of Approximations with Maximum Likelihood Estimates.

Table 1(a) : 1-step estimators without moment adjustment.

Method	Maximum Likelihood	initial iteration only	% difference from MI	1st ad hoc method (see text)	% difference from MI	2nd ad hoc method (see text)	% difference from ML
Const.	3.0720 (.0795)	3.1078 (.0977)	1.2	2.9074 (.1058)	-5.4	2.9895 (.0913)	-2.7
X	.0239 (.0014)	.0226 (.0017)	-5.3	.0348 (.0019)	45.6	.0289 (.0016)	21.2
X^2	-.00045 (.00002)	-.00042 (.00003)	-5.0	-.00063 (.00003)	41.9	-.00054 (.00003)	19.7
S	.0252 (.0047)	.0242 (.0058)	-4.2	.0250 (.0063)	-0.8	.0253 (.0054)	0.2
F	.0543 (.0089)	.0524 (.0109)	-3.5	.0491 (.0118)	-9.6	.0531 (.0102)	-2.4
A	.0214 (.0085)	.0212 (.0104)	-0.8	.0038 (.0113)	-82.3	.0143 (.0097)	-33.2
M	.1024 (.0125)	.0985 (.0153)	-3.8	.1256 (.0166)	22.7	.1149 (.0143)	12.3
SW	-.1235 (.0162)	-.1161 (.0199)	-6.0	-.1613 (.0216)	30.7	-.1410 (.0186)	14.2
R	.1244 (.0095)	.1211 (.0117)	-2.6	.1285 (.0126)	3.3	.1283 (.0109)	3.2
T	.0796 (.0085)	.0768 (.0105)	-3.4	.1016 (.0113)	27.8	.0910 (.0098)	14.4
U	.1054 (.0086)	.0999 (.0106)	-5.3	.1261 (.0115)	19.6	.1157 (.0099)	9.7
$\hat{\sigma}$	.2601	.2622		.3490		.3003	
Log L	-8966.2	-8969.6		-9384.5		-9078.5	
R^2	.364	.362		.335		.357	
mean absolute percentage difference from ML			3.7		26.3		12.1

Variables: X = Experience, S = Age completed full-time education, F = Any further education since initial finishing, A = Taken apprenticeship, M = married, SW = secondary worker, R = job involves responsibility for the work of others, T = Training need to get a job of this type, U = Member of Trade Union.

Sample: Male manual workers in manufacturing. Sample size = 5352

Standard errors are given in parentheses.

Table 1 (b) : 1-step estimators with moment adjustment

Method	Maximum Likelihood	initial iteration + moment adjustment	% difference from MI	1st ad hoc method + moment adjustment	% difference from MI	2nd ad hoc method + moment adjustment	% difference from ML
Const.	3.0720 (.0795)	3.0743 (.0794)	0.1	3.0027 (.0885)	-2.3	3.0450 (.0827)	-0.9
X	.0239 (.0014)	.0236 (.0014)	-1.0	.0316 (.0016)	32.3	.0273 (.0014)	14.4
X^2	-.00045 (.00002)	-.00044 (.00002)	-0.7	-.00058 (.00003)	28.9	-.00051 (.00003)	13.1
S	.0252 (.0047)	.0253 (.0047)	0.1	.0227 (.0053)	-9.9	.0239 (.0049)	-5.4
F	.0543 (.0089)	.0548 (.0089)	0.9	.0447 (.0099)	-17.8	.0501 (.0092)	-7.8
A	.0214 (.0085)	.0222 (.0085)	3.7	.0034 (.0095)	-83.9	.0135 (.0088)	-36.9
M	.1024 (.0125)	.1030 (.0125)	0.6	.1141 (.0139)	11.5	.1085 (.0130)	6.0
SW	-.1235 (.0162)	-.1214 (.0162)	-1.7	-.1466 (.0180)	18.8	-.1332 (.0169)	7.9
R	.1244 (.0095)	.1266 (.0095)	1.8	.1167 (.0105)	-6.1	.1212 (.0099)	-2.5
T	.0796 (.0085)	.0803 (.0085)	0.9	.0924 (.0095)	16.1	.0860 (.0088)	8.1
U	.1054 (.0086)	.1044 (.0086)	-1.0	.1146 (.0096)	8.7	.1092 (.0090)	3.6
$\hat{\sigma}$	.2601	.2601		.2907		.2711	
Log L	-8966.2	-8966.4		-9084.6		-8988.7	
R^2	.364	.364		.350		.363	
mean absolute percentage difference from ML			1.1		21.5		9.7

Table 1(c): 2-step estimators without moment adjustment

Method	Maximum Likelihood	Two consecutive iterations estimator	% difference from ML	1st ad hoc method + one iteration	% difference from ML	2nd ad hoc method + one iteration	% difference from ML
Const.	3.0720 (.0795)	3.0748 (0.794)	0.1	3.0331 (.0830)	-1.3	3.0536 (.0811)	-0.6
X	.0239 (.0014)	.0238 (.0014)	-0.5	.0259 (.0014)	8.6	.0248 (.0014)	3.9
X^2	-.00045 (.00002)	-.00045 (.00002)	-0.4	-.00048 (.00003)	7.9	-.00046 (.00002)	3.6
S	.0252 (.0047)	.0252 (.0047)	-0.3	.0255 (.0049)	1.0	.0254 (.0048)	0.5
F	.0543 (.0089)	.0542 (.0089)	-0.2	.0547 (.0093)	0.6	.0546 (.0090)	0.4
A	.0214 (.0085)	.0214 (.0085)	0.3	.0186 (.0089)	-13.2	.0202 (.0086)	-5.4
M	.1024 (.0125)	.1021 (.0125)	-0.2	.1064 (.0130)	4.0	.1044 (.0127)	2.0
SW	-.1235 (.0162)	-.1228 (.0162)	-0.6	-.1308 (.0169)	5.9	-.1268 (.0165)	2.7
R	.1244 (.0095)	.1242 (.0095)	-0.1	.1261 (.0099)	1.4	.1252 (.0097)	0.7
T	.0796 (.0085)	.0794 (.0085)	-0.2	.0839 (.0089)	5.4	.0816 (.0087)	2.6
U	.1054 (.0086)	.1050 (.0086)	-0.4	.1106 (.0090)	4.9	.1078 (.0088)	2.3
$\hat{\sigma}$	.2601	.2601		.2723		.2656	
Log L	-8966.2	-8966.2		-8981.4		-8969.5	
R^2	.364	.364		.364		.364	
mean absolute percentage difference from ML			0.3		4.9		2.2

Table 1(d): 2-step estimators with moment adjustment.

Method	Maximum Likelihood	Proposed 2-step estimator	% difference from ML	1st ad hoc method + both	% difference from ML	2nd ad hoc method + both	% difference from ML
Const.	3.0720 (.0795)	3.0725 (.0794)	0.02	3.0529 (.0810)	-0.6	3.0644 (.0800)	-0.2
X	.0239 (.0014)	.0239 (.0014)	-0.1	.0250 (.0014)	4.7	.0244 (.0014)	2.0
X^2	-.00045 (.00002)	-.00045 (.00002)	-0.1	-.00047 (.00002)	4.3	-.00046 (.00002)	1.8
S	.0252 (.0047)	.0252 (.0047)	-0.02	.0253 (.0048)	0.2	.0252 (.0048)	0.01
F	.0543 (.0089)	.0544 (.0089)	0.04	.0541 (.0090)	-0.4	.0542 (.0089)	-0.2
A	.0214 (.0085)	.0214 (.0085)	0.3	.0195 (.0086)	-8.9	.0206 (.0085)	-3.7
M	.1024 (.0125)	.1024 (.0125)	0.02	.1043 (.0127)	1.9	.1032 (.0126)	0.8
SW	-.1235 (.0162)	-.1232 (.0162)	-0.2	-.1277 (.0165)	3.4	-.1252 (.0163)	1.4
R	.1244 (.0095)	.1245 (.0095)	0.1	.1248 (.0097)	0.3	.1245 (.0095)	0.1
T	.0796 (.0085)	.0796 (.0085)	0.01	.0818 (.0087)	2.8	.0805 (.0086)	1.2
U	.1054 (.0086)	.1053 (.0086)	-0.1	.1080 (.0088)	2.4	.1064 (.0087)	1.0
$\hat{\sigma}$	.2601	.2600		.2654		.2621	
Log L	-8966.2	-8966.2		-8969.9		-8966.8	
R^2	.364	.364		.364		.364	
mean absolute percentage difference from ML			0.1		2.7		1.1

the two ad hoc methods. The moment estimators corresponding to each of these three are presented in Table 1(b). The results of applying one Maximum Likelihood iteration direct to each of the three are presented in Table 1(c) and Table 1(d) contains the results of applying this iteration to the three moment estimators. Hence the first of the three presented in Table 1(d) is the proposed "two-step estimator". In addition the fully iterated Maximum Likelihood estimates are given in each of the sub-tables for purposes of comparison. The percentage differences in the coefficient estimates from the corresponding Maximum Likelihood estimates are also presented for each of the estimators.

The single iteration on the basis of the estimated marginal distribution is clearly superior, in the sense of giving a better approximation to the Maximum Likelihood estimates, to both of the ad hoc methods. The mean absolute percentage difference in the coefficient estimates from the Maximum Likelihood estimates is 3.7% compared with 26.3% and 12.1% for the two ad hoc methods. The coefficient on which both ad hoc methods fall down most badly is that on the variable A, which has the lowest asymptotic t-ratio of those in the equation. This is obviously a serious drawback to the use of such ad hoc estimators. The single iteration estimator also provides a superior estimate of σ (differing from the Maximum Likelihood estimate by less than 1% compared with 34% and 15% for the two ad hoc methods) and attains a likelihood value much closer to the maximum (differing from the maximum by 3 as compared with 418 and 112).

Comparing the moment estimators in Table 1(b) with the corresponding columns of 1(a) there is a clear improvement in all three cases, despite the non-normality of the regressors. However the estimators based on the two ad hoc starts are still poor. The relative improvement in the percentage difference from M.L. is greatest for the moment estimator based on adjusting the single iteration estimator. The mean absolute percentage difference is now only 1.1%. In addition the estimate of σ differs by less than 0.1% and the log likelihood is only 0.2 away from its maximum. It would seem that in the non-normal regressors case the effectiveness of the moment adjustments is dependent on the initial choice of the q_k.

The improvement that results from an iteration of the Maximum Likelihood algorithm (Table 1(c)) is greater in each case than that from the moment adjustments. This is particularly true for the two based on ad hoc starts. These estimates are now reasonable approximations, but still considerably inferior to the estimator based on the iteration start. That gives a mean absolute percentage difference from the Maximum Likelihood coefficient estimates of 0.3% and an estimate of σ equal to 4 decimal places and is within 0.1 of the maximum of the log-likelihood function.

Finally interspersing the two iterations with the moment estimator adjustments to give the "two-step estimator" proposed in Section IV (Table 1(d)) gives a yet further improvement. The mean absolute percentage difference from the Maximum Likelihood coefficient estimates is now less than 0.1% and for no single coefficient does it exceed 0.3%. Thus in this illustration the proposed "two-step estimator" provides highly

satisfactory approximations to the Maximum Likelihood estimates. Whilst the convergence of the algorithm of Section II is monotonic, the improvements in estimates are much smaller in all cases for the remaining iterations. (Six to eight iterations are required for convergence when the largest parameter estimate change permitted is 10^{-5}.)

The second illustration of this section examines the consequences of such grouping again in the context of the estimation of earnings equations. Data from the General Household Survey are utilised to compare the Maximum Likelihood estimates on artificially grouped data (using the NTS grouping) with the estimates from using the original (ungrouped) data. The dependent variable is again the logarithm of weekly earnings and the explanatory variables are as listed at the foot of Table 2.

The results are presented in Table 2 and it can be seen that there is fairly close agreement between the Maximum Likelihood estimates and OLS estimates using the original ungrouped data. The mean absolute difference between the two is .0072. Since the dependent variable is the logarithm of weekly earnings this represents about three-quarters of a percentage point in the differential. The mean absolute percentage difference between the two sets of estimates is 5.7%. The correlation between the complete earnings data and the final Maximum Likelihood estimates of the conditional expectations is .9682, while the correlation between the predictions from the two sets of estimates (now not conditional on k in the case of the Maximum Likelihood estimates) is .9996. The consequences of grouping do not appear to be too severe in this case.

TABLE 2 : Comparison Using GHS Data Grouped and Ungrouped

	Original Data (Ungrouped) (O.L.S. Coeff. & st. error)		Grouped Data (M.L. Coeff. & asym. st. error)		Difference	Percentage Difference
Const.	2.8360		2.8840		.0480	1.7
X1	.1345	(.0086)	.1269	(.0086)	-.0076	-5.7
X2	.0437	(.0050)	.0435	(.0047)	-.0002	-0.5
X3	.0125	(.0021)	.0113	(.0020)	-.0012	-9.1
X4	-.0020	(.0021)	-.0017	(.0020)	.0003	14.4
X5	-.0040	(.0022)	-.0044	(.0021)	-.0004	-10.2
X6	-.0147	(.0027)	-.0146	(.0026)	.0001	0.6
S16	.1615	(.0133)	.1521	(.0126)	-.0094	-5.9
S17	.2479	(.0211)	.2295	(.0200)	-.0184	-7.4
S18	.2992	(.0267)	.2655	(.0255)	-.0337	-11.3
S19+	.4358	(.0167)	.4076	(.0159)	-.0282	-6.5
F1	.1334	(.0171)	.1335	(.0162)	.0001	0.1
F2	.0754	(.0153)	.0786	(.0145)	.0032	4.2
F3	.0351	(.0172)	.0388	(.0162)	.0037	10.4
F4	.0486	(.0234)	.0508	(.0221)	.0022	4.5
ILL	-.0596	(.0108)	-.0582	(.0102)	.0014	2.3
MAR	.1560	(.0134)	.1501	(.0126)	-.0059	-3.8
COL	-.1699	(.0319)	-.1759	(.0301)	-.0060	-3.5
OLS $\hat{\sigma}$	.3203		-			
ML $\hat{\sigma}$	.3197		.2950			-7.7
Log L	-2668.8		-9771.6			
R^2	.3560		.3840			
Distribution of y:						
Mean	3.9910		3.9855			
S.D.	.3985		.3663			

Variables: X1 to X6 = Linear spline on years of experience (X1 and X2 are of width 5 years the remainder 10 years), S16 to S19+ = Age on completion of full-time education, F1 to F4 = Father's occupation was (1) non-manual (2) skilled manual (3) semi-skilled manual (4) farmer or similar (base group, is unskilled manual), ILL = Has long-standing illness or disability MAR = Married, COL = Non-white.

Sample: Full-time males Sample size = 5338 .

Standard errors in parentheses.

VI: Sensitivity to Sample Properties - A Simulation Exercise

In order to ascertain how dependent are the favourable results of the previous section on the particular samples involved a number of Monte Carlo experiments were conducted. Among the features to which the estimators might be expected to be sensitive are non-normality, and particularly skewness, in the underlying distribution, the proportion of observations in the open-ended groups (the degree of censoring), the multiple correlation and the extent of assymmetry in the grouping (relative to the underlying distribution of y).

The underlying model used in all the experiment is given by

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad (i = 1,\ldots,N)$$

$$\beta_1 = \beta_2 = 1.0$$

The grouping was performed with ten groups ($K = 10$) and $A_k = k$ ($k = 1,\ldots,9$). Hence the centre of the grouping is at 5.0. The characteristics of the experiments conducted are given in Table 3. In all cases samples of 1000 were generated, this being regarded as a typical medium-sized sample for the type of work and data sets that tbe estimation methods are likely to be employed upon. The values of x_1 were generated throughout from a standard normal distribution. The distributions generating x_2 and u were standardised in each experiment to have zero mean and prescribed variances (denoted σ_2^2 and σ^2 respectively). 50 replications were performed for each experiment.

In the base experiment (experiment 1) x_2 is generated by a standard normal and u by a normal distribution with $\sigma = 2$. Hence the multiple correlation equals .5.

The resultant marginal distribution of y is normal and has variance 4. α (and hence μ_y) is taken to be 5 making the grouping symmetric about μ_y and generating on average approximately 2½% of the observations in each of the open-ended groups.

The estimators are then examined in different situations by varying σ^2, σ_2^2, and the distributions used to generate x_2 and u (and hence y). ρ^2 and the proportion of observations in the open-ended groups can be varied by altering σ^2 and σ_2^2. Varying α results in assymmetry in the structure of the grouping (relative to the underlying distribution of y). Using different distributions to generate x_2 allows examination of the effects of non-normality in the marginal distribution of y on the estimators under consideration. Finally if distributions other than the normal are used to generate u the conditional distribution of y will also be non-normal and the Maximum Likelihood estimator itself may no longer be consistent.

Experiments 2 and 3 vary σ^2 and σ_2^2. In experiment 2 σ_2^2 is reduced to 0.2 and σ^2 raised to 2.8. This reduces ρ^2 to .3 while keeping the average size of the open-ended groups the same. Experiment 3 increases σ_2^2 to 3 and σ^2 to 4. This restores ρ^2 to .5 (as in experiment 1) and increases the average size of the open-ended groups to about 8% each .

In experiment 4, α and hence μ_y is taken to be 3 causing the grouping to become assymmetric relative to μ_y. This results in about 16% of the observations falling in the left-hand open-ended group with only one in a thousand on average in the right-hand one. Experiments

5 and 6 use two convenient skewed distributions to generate x_2. In experiment 5 x_2 is generated by the chi-square distribution with 2 degrees of freedom (coefficient of skewness = 2.0). While in experiment 6 x_2 is generated by the lognormal distribution with median 0.5 and shape parameter 1.0 (coefficient of skewness = 6.2). In each case the distribution is standardised to give mean and variance equal to that in experiment 1. Finally experiments 7 and 8 use these same two distributions to generate non-normal disturbances. In this case the values are standardised to have a mean of zero and a variance of 2 as in experiment 1.

The NAG function GO5DDF was used to generate normal pseudo-random variables. (See Numerical Algorithms Group (1981) for details). $\chi^2(d)$ variates were generated by summing d squared standard normal pseudo-random numbers from GO5DDF, and the lognormal variates were generated as

$$L_i = m.\exp(s.N_i)$$

where m is the median, s the shape parameter and N_i a standard normal speudo-random number from GO5DDF. Each sample was initialised from the real-time clock.

Results for the eight experiments are given for five estimators in Tables 4 to 6. The five estimators are:

Table 3: Characteristics of Experiments

Experiment	Distribution of x_2	σ_2^2	Distribution of u	σ^2	$\mu_y = \alpha$	ρ^2	mean proportion $< A_1$	mean proportion $> A_{K-1}$
1	Normal	1	Normal	2	5	.5	.023	.023
2	Normal	0.2	Normal	2.8	5	.3	.023	.023
3	Normal	3	Normal	4	5	.5	.079	.079
4	Normal	1	Normal	2	3	.5	.157	.001
5	$\chi^2(2)$	1	Normal	2	5	.5	.018	.029
6	Lognormal (m=.5,s=1.0)	1	Normal	2	5	.5	.015	.026
7	Normal	1	$\chi^2(2)$	2	5	.5	.010	.036
8	Normal	1	Lognormal m=.5,s=1.0)	2	5	.5	.006	.030

Notes:
1. x_1 generated by N(0,1) distribution in all experiments.
2. Sample size = 1000 in all experiments.
3. 50 replications performed for each experiment.
4. In experiments 5 and 6 distribution of x_2 is standardised to have mean zero and variance 1.
5. In experiments 7 and 8 distribution of u is standardised to have mean zero and variance 2.
6. x_2 and u have mean zero in all experiments.

Table 4: Mean Biases

	Initial Iteration Only	Initial Iteration + Moment Adjustments	Two Iterations	The "Two-step Estimator"	Fully Iterated Maximum Likelihood	OLS on Ungrouped Data
Experiment 1:						
β_1	-.0315	-.0068	-.0080	-.0067	-.0068	-.0030
β_2	-.0213	.0037	.0019	.0032	.0032	.0038
σ	.0233	.0087	.0069	.0067	.0066	.0082
Experiment 2:						
β_1	-.0305	-.0058	-.0066	-.0056	-.0057	-.0048
β_2	-.0029	.0225	.0214	.0224	.0223	.0172
σ	.0042	-.0020	-.0043	-.0043	-.0045	.0011
Experiment 3:						
β_1	-.0308	.0027	-.0009	.0018	.0015	.0044
β_2	-.0271	.0065	.0028	.0055	.0051	.0051
σ	.0222	-.0079	-.0102	-.0102	-.0110	-.0033
Experiment 4:						
β_1	-.0477	.0014	-.0037	.0014	.0012	.0016
β_2	-.0474	.0017	-.0032	.0019	.0017	.0020
σ	.0351	.0033	.0013	.0008	.0003	-.0002
Experiment 5:						
β_1	-.0289	-.0040	-.0027	-.0012	-.0002	-.0003
β_2	-.0538	-.0296	-.0067	-.0037	.0016	.0029
σ	.0147	-.0002	-.0018	-.0020	-.0012	.0022
Experiment 6:						
β_1	-.0430	-.0180	-.0166	-.0151	-.0127	-.0100
β_2	-.1758	-.1542	-.0674	-.0608	.0070	.0097
σ	.0208	.0080	-.0048	-.0052	-.0059	-.0027
Experiment 7:						
β_1	-.0354	-.0101	-.0106	-.0091	-.0093	.0012
β_2	-.0414	-.0162	-.0161	-.0146	-.0148	-.0049
σ	-.0733	-.0902	-.1002	-.1005	-.1015	-.0066
Experiment 8:						
β_1	-.0383	-.0109	-.0126	-.0106	-.0108	.0072
β_2	-.0399	-.0125	-.0142	-.0122	-.0124	-.0019
σ	-.2912	-.3144	-.3256	-.3262	-.3271	.0016

Table 5: Mean Bias to Standard Deviation Ratios

	Initial Iteration Only	Initial Iteration + Moment Adjustments	Two Iterations	The "Two-step Estimator"	Fully Iterated Maximum Likelihood	OLS on Ungrouped Data
Experiment 1:						
β_1	-0.68	-0.14	-0.17	-0.14	-0.14	-0.06
β_2	-0.46	0.08	0.04	0.07	0.07	0.08
σ	0.73	0.27	0.21	0.21	0.20	0.29
Experiment 2:						
β_1	-0.66	-0.12	-0.14	-0.12	-0.12	-0.10
β_2	-0.02	0.17	0.16	0.17	0.17	0.13
σ	0.14	-0.07	-0.14	-0.14	-0.15	0.03
Experiment 3:						
β_1	-0.47	0.04	-0.01	0.02	0.02	0.07
β_2	-0.89	0.20	0.09	0.17	0.16	0.16
σ	0.58	-0.21	-0.28	-0.28	-0.30	-0.08
Experiment 4:						
β_1	-1.05	0.03	-0.08	0.03	0.03	0.04
β_2	-1.07	0.04	-0.07	0.04	0.04	0.05
σ	1.09	0.10	0.04	0.03	0.01	-0.01
Experiment 5:						
β_1	-0.69	-0.09	-0.06	-0.03	-0.01	-0.01
β_2	-1.40	-0.75	-0.16	-0.09	0.04	0.07
σ	0.46	-0.01	-0.06	-0.06	-0.04	0.07
Experiment 6:						
β_1	-0.97	-0.40	-0.38	-0.34	-0.29	-0.23
β_2	-1.93	-1.65	-0.93	-0.85	0.14	0.26
σ	0.56	0.21	-0.13	-0.14	-0.16	-0.08
Experiment 7:						
β_1	-0.82	-0.23	-0.24	-0.20	-0.21	0.03
β_2	-0.96	-0.37	-0.36	-0.33	-0.33	-0.11
σ	-1.58	-1.90	-2.15	-2.16	-2.18	-0.10
Experiment 8:						
β_1	-0.99	-0.28	-0.33	-0.27	-0.28	0.14
β_2	-0.89	-0.27	-0.31	-0.27	-0.27	-0.04
σ	-5.57	-5.79	-6.26	-6.27	-6.31	0.06

Table 6: Root Mean-Square Errors

	Initial Iteration Only	Initial Iteration + Moment Adjustments	Two Iterations	The "Two-step Estimator"	Fully Iterated Maximum Likelihood	OLS on Ungrouped Data
Experiment 1:						
β_1	.0562	.0480	.0485	.0484	.0484	.0469
β_2	.0514	.0478	.0474	.0476	.0476	.0488
σ	.0397	.0338	.0334	.0334	.0333	.0301
Experiment 2:						
β_1	.0550	.0472	.0476	.0475	.0475	.0484
β_2	.1312	.1363	.1357	.1360	.1359	.1378
σ	.0307	.0308	.0311	.0312	.0312	.0342
Experiment 3:						
β_1	.0720	.0681	.0674	.0677	.0676	.0622
β_2	.0409	.0331	.0319	.0324	.0324	.0316
σ	.0440	.0385	.0378	.0379	.0379	.0397
Experiment 4:						
β_1	.0658	.0481	.0465	.0467	.0465	.0431
β_2	.0649	.0466	.0448	.0449	.0447	.0433
σ	.0476	.0321	.0326	.0327	.0328	.0289
Experiment 5:						
β_1	.0510	.0432	.0427	.0427	.0427	.0423
β_2	.0662	.0493	.0427	.0427	.0442	.0401
σ	.0351	.0322	.0313	.0313	.0312	.0317
Experiment 6:						
β_1	.0618	.0487	.0470	.0465	.0456	.0455
β_2	.1980	.1804	.0988	.0940	.0506	.0389
σ	.0427	.0388	.0370	.0371	.0376	.0355
Experiment 7:						
β_1	.0559	.0454	.0456	.0454	.0455	.0480
β_2	.0599	.0471	.0473	.0469	.0470	.0429
σ	.0866	.1019	.1105	.1108	.1116	.0676
Experiment 8:						
β_1	.0544	.0408	.0407	.0401	.0401	.0513
β_2	.0599	.0473	.0476	.0472	.0472	.0535
σ	.2958	.3190	.3298	.3303	.3312	.1396

(i) Initial Iteration Only: a single iteration of the Maximum Likelihood algorithm of Section II with the dependent variable constructed on the basis of consistent estimates of the parameters of the marginal distribution.

(ii) Initial Iteration + Moment Adjustments : the moment estimator based on the OLS estimates in (i).

(iii) Two Iterations : a second iteration applied to (i).

(iv) The "Two-Step Estimator" : a second iteration applied to (ii).

(v) Fully Iterated Maximum Likelihood : the algorithm of Section II iterated to convergence.

For purposes of comparison the results of applying OLS to the ungrouped data are also given. Table 4 gives the mean biases of the estimates of β_1, β_2 and σ, and Table 6 gives the equivalent root mean-square errors. If the estimates obtained from each experimental replication are assumed to be asymptotically normal, the ratio of the mean bias to its estimated standard deviation will be distributed approximately as t with 49 degrees of freedom. These ratios are presented in Table 5. Whilst they can be generated easily enough from the entries in Tables 4 and 6, they provide useful summary statistics.

Comparing first the Maximum Likelihood estimates with the results

from applying OLS to the ungrouped data, it is important to distinguish the last two experiments from the rest. When the disturbances are normally distributed (experiments 1 to 6) both estimators give consistent estimates, whilst in experiments 7 and 8 only OLS on the ungrouped data does. Thus in experiments 1 to 6 the mean biases for both are all small and none are significantly different from zero (see Table 5). In addition the root mean-square errors for the two estimators are very similar, suggesting that the loss of precision due to the grouping is small when the disturbances are normally distributed and confirming the findings of the previous section (Table 2). In the case of non-normal disturbances (experiments 7 and 8) the mean biases in the slope coefficients (β_1 and β_2) for both estimators are again insignificantly different from zero and the root mean-square errors are very similar both to one another and to those in the earlier experiments. However the Maximum Likelihood estimate of σ has a mean bias that is much larger and significantly different from zero in both experiments and the root mean-square error is much increases. Hence, not unexpectedly, the accuracy of the estimation of σ is much reduced when the disturbances have a skewed distribution, i.e. when the wrong conditional distribution has been assumed.

The "two-step estimator' performs very well in these experiments. The root mean-square errors are very similar to those for the Maximum Likelihood estimator in all experiments (including the experiments where u is non-normal) and the mean biases are never significantly different from zero except in the cases when those for the Maximum Likelihood estimator are.

Experiment 2 (reduced ρ^2) exhibits a slight increase in the mean bias and root mean-square error of the estimate of β_2. This is due to the relative reduction in σ_2^2 and is only in line with that exhibited by OLS on ungrouped data. The relative performance of the "two-step estimator" does not appear to be impaired by a reduction in ρ^2.

Experiment 3 (enlarged open-ended groups) gives a slight increase in the root mean-square error of the estimate of β_1, but again this is only in line with that exhibited by OLS on ungrouped data. In this case the relative variance of x_1 has been reduced by the increase in σ^2 and σ_2^2. The relative performance of the "two-step estimator" does not appear to be impaired by an increase in the proportion of observations in the open-ended groups (the degree of censoring) either.

Experiment 4 (assymmetric grouping) produces no increases in any of the mean biases or root mean-square errors. Again the results parallel OLS on ungrouped data and no impairment in the relative performance of the "two-step estimator" is evident.

In experiments 5 and 6 x_2 is generated by non-normal distributions. Experiment 5 exhibits little change in the mean biases or root mean-square errors. In experiment 6 (the more skewed) the mean bias and root mean-square error of the estimate of β_2 are somewhat increased, but the mean bias is still not significantly different from zero.

In experiments 7 and 8 where u is generated by non-normal

distributions the mean biases and root mean-square errors move in parallel with those for the Maximum Likelihood estimator. The comments made earlier on the performance relative to OLS on ungrouped data apply equally here, but the performance of the "two-step estimator" relative to the Maximum Likelihood estimator is as good as before.

Overall the evidence suggests that the "two-step estimator" provides satisfactory estimates in all cases where the Maximum Likelihood estimator does and only in the case of the most skewed x_2 distribution is its relative performance impaired in any way.

Turning to the other estimators considered, the moment estimator, as expected, performs equally well in the experiments where x_2 is generated by a normal distribution, but less well in the remaining two. In experiments 1 to 4 the mean biases and root mean-square errors are similar to those for the "two-step estimator" and the Maximum Likelihood estimator. The moment estimator appears to give just as good an approximation in these cases as the "two-step estimator". The position is similar in experiments 7 and 8.

In experiments 5 and 6 the "two-step estimator" does, as expected, provide a considerable improvement in the estimation of β_2 compared with the moment estimator. The mean bias and root mean-square error are much larger for the moment estimator and the mean bias is bordering on significance in the case of the more skewed of the two distributions.

The two iterations (without moment adjustments) estimator results in root mean-square errors very similar to those for the "two-step estimator" in all experiments. The mean biases are only

significant in the experiments where those for the "two-step estimator" and Maximum Likelihood estimator are; and they also are fairly similar. In the case of experiments 1 to 3 the moment adjustments (i.e. comparing the "two-step estimator") do not appear to improve the estimator in the sense of reducing the mean biases. In experiments 4 to 8 there are slight reductions in the mean biases of the estimates of β_1 and β_2, but the improvement does not appear to be a major one.

Finally turning to the initial iteration only estimator, the root mean-square errors and mean biases tend to be larger than those for the other estimators. The mean biases of the estimate of β_2 are significant, or close to, in experiments 5 and 6 and the significance is considerably greater than that for both the moment and two iterations estimators. This latter comment is also the case for all parameters in experiment 4 (assymmetric grouping). In all cases either the moment adjustments or a second iteration or both seem beneficial.

In conclusion, these experiments suggest:

(i) that the loss of precision due to such grouping is only slight when the disturbances are normally distributed;

(ii) that the estimation of σ suffers when the wrong conditional distribution is chosen, but that the slope parameter estimates are much less affected and may not be unduly impaired;

(iii) that the "two-step estimator" performs very well in all cases where the Maximum Likelihood estimator does and

provides most satisfactory approximations to the Maximum Likelihood estimator;

(iv) that the moment estimator performs equally well when the regressors are normally distributed, but that the "two-step estimator" provides considerable improvements in the case of non-normal regressors;

(v) that the two iterations estimator also performs well in all situations and that while the moment adjustments improve the performance in some cases they may not be necessary;

(vi) that the initial iteration estimator is substantially improved by either the moment adjustments or a second iteration.

VII: Conclusions

This paper has examined the problem of estimating the parameters of an underlying linear model on the basis of data in which the dependent variable is grouped. An algorithm for attaining the Maximum Likelihood solutions has been described. This algorithm has been shown to be a special case of the EM algorithm and hence to have the property of monotonic convergence. The results of Greene (1981) on the asymptotic bias of OLS have been extended to the grouped dependent variable model and a "moment" estimator derived for the normal regressors case. A Least Squares approximation to the Maximum Likelihood estimator involving use of a particular application of the "moment" estimator in conjunction with early termination of the monotonically convergent algorithm is proposed and found in an illustration to provide a useful and satisfactory estimator. The application to the estimation of earnings functions from NTS data found the proposed "two-step estimator" to be superior to the ad hoc methods examined, the various straight moment estimators, some estimators based on the Maximum Likelihood algorithm alone, and various combinations thereof and to provide a very good approximation to the full Maximum Likelihood estimator. This was confirmed by a number of simulation experiments. Estimation of earnings functions from GHS data to compare the Maximum Likelihood estimates with those based on the original (ungrouped) data demonstrated considerable agreement between the coefficients and also between the two sets of predictions. In the case of the particular grouping examined (that employed in the NTS) the consequences of grouping do not appear to be too severe. This finding was also broadly confirmed by the simulation experiments.

References

Aitchison, J. and J.A.C.Brown (1966) - The Lognormal Distribution : with Special Reference to its Uses in Economics (Cambridge University Press).

Dempster, A.P., N.M.Laird and D.B.Rubin (1977) - "Maximum Likelihood from Incomplete Data via the EM Algorithm", Journal of the Royal Statistical Society, Series B, 39, 1-38.

Greene, W.H. (1981) - "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model", Econometrica, 505-514.

Manpower Services Commission (1978) - People and their Work.

Numerical Algorithms Group (1981) - NAG Library Manual (Oxford).

Office of Population Censuses and Surveys (1978) - The General Household Survey 1975 (HMSO, London).

Olsen, R.J. (1980) - "Approximating a Truncated Normal Regression with the Method of Moments", Econometrica, 48, 1099-1105.