Industrial Relations Section
Princeton University
Working Paper #159
Revised: November 1982

ON LEAST SQUARES ESTIMATION WHEN THE DEPENDENT VARIABLE IS GROUPED*

Mark B. Stewart
University of Warwick and
Princeton University

* I am grateful to Jim Burridge, Peter Burridge, William Greene, Nick Kiefer, Mervyn King, Mark Salmon, Ken Wallis and an anonymous referee for helpful comments and to the Manpower Services Commission for permission to use the National Training Survey data. The final revision of the paper was completed whilst I was visiting Princeton University whose hospitality is gratefully acknowledged.

ABSTRACT

"On Least Squares Estimation When the Dependent Variable is Grouped"

by

Mark B. Stewart

This paper examines the problem of estimating the parameters of an underlying linear model using data in which the dependent variable is only observed to fall in a certain interval on a continuous scale, its actual value remaining unobserved. A Least Squares algorithm for attaining the Maximum Likelihood estimator is described, the asymptotic bias of the OLS estimator derived for the normal regressors case and a "moment" estimator presented. A "two-step estimator" based on combining the two approaches is proposed and found to perform well in both an economic illustration and simulation experiments.

1. 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 now well 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 interval on a continuous scale, its actual value remaining unobserved. The data are also censored in the usual sense in that both end intervals are assumed to be open-ended.[1] A number of Least Squares estimators that approximate Maximum Likelihood are derived and compared and the results of Greene (1981) on the asymptotic bias of OLS extended to this case.

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 x 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 interval of the real line. The real line is divided into K intervals, the k-th being given by (A_{k-1}, A_k) and these K intervals exhaust the real line. Thus $A_0 = -\infty$ and $A_K = +\infty$, i.e. the first and K-th intervals are "open-ended". The information on the dependent variable is which of these K intervals it falls into, i.e. an indicator variable k_i $(1 \leq k_i \leq K)$ is observed for each i.

This type of problem is important in much applied work, such variables being encountered in 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. Appropriate techniques for the analysis of its earnings variable are thus urgently needed. A number of variables in the General Household Survey (see Office of Population Censuses and Surveys (1978) for details) also give rise to this type of problem. In particular housing expenses, length of time with the present employer and duration of unemployment are all grouped in that survey.[2]

Analysis of these large survey data sets is usually undertaken on one of the commonly available general statistical packages, the range of suitable software being severely restricted by the sample sizes involved. (The National Training Survey, for example, contains approximately 54,000 observations.) Maximum Likelihood routines when available on these packages are generally ruled out by the sample size restrictions and the applied researcher prepared to write special programs may well be discouraged by the execution costs involved. Fast and easy Least Squares techniques that approximate Maximum Likelihood are thus of great value and to be preferred to the ad hoc methods that will be used in their absence. This paper provides and illustrates such techniques for the problem under consideration.

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,

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

where $Z_k = (A_k - \underset{\sim}{x}_i' \underset{\sim}{\beta})/\sigma$ and f and F are the density function and cumulative distribution of the standard normal respectively. Hence the requisite estimation of the conditional expectations requires estimates of $\underset{\sim}{\beta}$ 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 2 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 3 derives a "moment" estimator for the normal regressors case, extending the recent work of Olsen (1980) and Greene (1981). Section 4 then considers a Least Squares "two-step estimator" involving use of the moment estimator in conjunction with early termination of the convergent algorithm. Alternative Least Squares approximations and the full Maximum Likelihood are then illustrated and compared in Section 5 by the estimation of an earnings equation using NTS data. In Section 6 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 7 presents some conclusions.

2. MAXIMUM LIKELIHOOD ESTIMATION

The log likelihood function for the model outlined in the previous section is given by

$$\log L = \sum_{k=1}^{K} \sum_{i\varepsilon k} \log \left\{ F\left[(A_k - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\sigma\right] - F\left[(A_{k-1} - \underset{\sim}{x}_i'\underset{\sim}{\beta})/\sigma\right]\right\}$$

$$= \sum_i \log \left\{ F_k - F_{k-1}\right\}$$

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

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

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

The final first-order condition can be rearranged to give an estimate of σ^2 as follows. The conditional variances of the y_i are given by
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

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

Thus the first J of the first-order conditions can be written 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 the usual 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:

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

The final first-order condition can be rearranged to give an estimate of σ^2 as follows. The conditional variances of the y_i are given by

$$\text{Var}\,(y_i|k_i,\underset{\sim}{x}_i) = \sigma^2 \left\{\left[\frac{Z_{k-1}f(Z_{k-1}) - Z_kf(Z_k)}{F(Z_k) - F(Z_{k-1})}\right] - \left[\frac{f(Z_{k-1}) - f(Z_k)}{F(Z_k) - F(Z_{k-1})}\right]^2 + 1\right\}$$

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

Hence the final first-order condition can be written as

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

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

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

where $d = \sum_i(1 - v_i)$. Hence the likelihood conditions can be solved by iterating between $\hat{\underset{\sim}{m}}$ and $(\hat{\underset{\sim}{\beta}}, \hat{\sigma})$.[3] This method of solving the likelihood conditions can be seen to be an application of the EM Algorithm discussed by Dempster et al. (1977).[4] Hence convergence is guaranteed, and the likelihood is increased at each iteration.[5]

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. These are given by

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

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

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

where $M_r = \dfrac{Z_{k-1}^r f_{k-1} - Z_k^r f_k}{F_k - F_{k-1}}$.[6]

3. A MOMENT ESTIMATOR FOR THE NORMAL REGRESSORS CASE

This section examines the inconsistency of OLS in the grouped dependent variable model, and derives a moment estimator which is consistent in the case when the regressors are normally distributed. 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

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

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

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

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

$$\underset{\sim}{c} = \underset{\sim}{S}_{xx}^{-1} \underset{\sim}{S}_{xg} \; ; \qquad a = \bar{g} - \bar{\underset{\sim}{x}}'\underset{\sim}{c} \; ; \qquad s^2 = S_{gg} - \underset{\sim}{S}_{xg}' \underset{\sim}{c} \; ;$$

where $\underset{\sim}{S}_{xx}$, $\underset{\sim}{S}_{xg}$ and S_{gg} are the appropriate sample moments and tend in probability to their population equivalents.

To examine the inconsistency in the OLS estimates, the following moments of the observed random variables are required.

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

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

where $\quad B_k = (A_k - \mu_y)/\sigma_y$.

$$E(g^2) = \sum_{k=1}^{K} q_k^2 \{F(B_k) - F(B_{k-1})\} = \psi \text{ , say.}$$

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

The conditional expectation here is given by

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

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

Thus

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

$$= \underset{\sim}{\mu}_x \lambda + \underset{\sim}{\sigma}_{xy} \theta$$

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

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

$$\text{plim } \underset{\sim}{S}_{xg} = E(\underset{\sim}{x}g) - \underset{\sim}{\mu}_x E(g) = \underset{\sim}{\sigma}_{xy} \theta$$

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

Thus

$$\text{plim } \underset{\sim}{c} = \underset{\sim}{\Sigma}_{xx}^{-1} \underset{\sim}{\sigma}_{xy} \theta = \underset{\sim}{\gamma} \theta \neq \underset{\sim}{\gamma} \text{ , in general.}$$

Thus all the OLS slope coefficient estimates are inconsistent by the same proportion. Turning to a,

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

$$= \lambda - \underset{\sim}{\mu}_x' \underset{\sim}{\gamma} \theta$$

$$= \lambda + (\alpha - \mu_y)\theta \ .$$

Finally turning to s^2,

$$\text{plim } S_{gg} = \text{Var}(g) = \psi - \lambda^2$$

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

Thus

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

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

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

Clearly the OLS estimates are in general inconsistent. However, consistent estimates can easily be derived from them using the following simple adjustments.

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

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

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

where $\hat{\lambda}$, $\hat{\psi}$, $\hat{\theta}$ are evaluated at consistent estimates $\hat{\mu}_y$ and $\hat{\sigma}_y$. 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.

4. A LEAST SQUARES TWO-STEP ESTIMATOR

A Least Squares two-step estimator that approximates Maximum Likelihood will be valuable in applied work for the reasons outlined in Section 1. In the case of non-normally distributed regressors 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 2 means that Least Squares approximations to the full Maximum Likelihood 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 2 and 3 will be beneficial. An iteration of the monotonically convergent algorithm will improve the moment estimator (in the sense of increasing the value of the likelihood). It is likely to be particularly useful when the regressors are non-normal, but will also result in a more efficient estimator in the normal regressors case. 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 OLS estimates based on any appropriate $q_k (k = 1,\ldots,K)$ satisfying $A_{k-1} < q_k < A_k$ however ad hoc the choice. However, given the consistent estimates of μ_y and σ_y required for the adjustment factors, consistent estimates of the conditional expectations of the marginal distribution for y can be obtained and used for the q_k,

$$q_k = \hat{\mu}_y + \hat{\sigma}_y \frac{f(\hat{B}_{k-1}) - f(\hat{B}_k)}{F(\hat{B}_k) - F(\hat{B}_{k-1})} \qquad (k = 1,\ldots,K)$$

where $\hat{\beta}_k = (A_k - \hat{\mu}_y)/\hat{\sigma}_y$. OLS estimation of $\underset{\sim}{\beta}$ using these q_k is then equivalent to one iteration of the algorithm described in Section 2 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.[7] A weakness with these initial OLS estimates is that, inevitably, the information contained in the explanatory variables for any given observation is not utilised in the construction of the estimated conditional expectations. Hence one iteration of the Maximum Likelihood algorithm of Section 2 is likely to produce considerable improvements in the approximation to the Maximum Likelihood estimates.

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 are evaluated on the basis of the parameters of the conditional distribution as described in Section 2. 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 μy and σy are 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.

5. AN ILLUSTRATION - THE ESTIMATION OF EARNINGS EQUATIONS

This section illustrates the methods presented above in the context of the estimation of earnings equations. The "two-step estimator" and some others suggested by the previous sections 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) and the sample used here consists of 5352 full-time manual male employees in manufacturing. 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 interval in the 1975 General Household Survey. 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.[8] The single iteration on the basis of the estimated marginal distribution (Table 1(a)) 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 estimate of σ differs from the Maximum Likelihood estimate by less than 1% compared with 34% and 15%, and the likelihood value attained differs from the maximum by 3 as compared with 418 and 112.

Comparing the moment estimates 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 sensitive to 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.

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 ML	1st ad hoc method (see text)	% difference from ML	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

Approximate Standard errors are given in parentheses.

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

Method	Maximum Likelihood	initial iteration + moment adjustment	% difference from ML	1st ad hoc method + moment adjustment	% difference from ML	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
λ	.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 (.0086)	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

Finally, interspersing the two iterations with the moment estimator adjustments to give the "two-step estimator" proposed in Section 4 (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 and is superior to the various alternatives considered. Whilst the convergence of the algorithm of Section 2 is monotonic, the improvements in estimates are much smaller in all cases for the remaining iterations. (An additional four iterations are required for convergence when the largest proportional change permitted in a parameter estimate is 10^{-5}.) The computational savings are considerable: the "two-step estimator" takes less than 40% of the time of the Maximum Likelihood estimator (and less than one quarter of the time needed to attain the M.L. estimator by Newton-Raphson).[9]

When data from the General Household Survey were utilised to compare the Maximum Likelihood estimates on artificially grouped data (using the NTS grouping) with the estimates from using the original (ungrouped) data, there was fairly close agreement between the Maximum Likelihood estimates and the OLS estimates using the original ungrouped data and the correlation between the complete earnings data and the final Maximum Likelihood estimates of the conditional expectations was .97. The consequences of grouping do not appear to be too severe in this particular case.[10]

6. SENSITIVITY TO SAMPLE PROPERTIES - A SIMULATION EXERCISE

In order to ascertain how dependent are the favourable results of the previous section on the particular sample involved a number of Monte Carlo experiments were conducted examining the sensitivity of the estimators to non-normality in the underlying distribution, the proportion of observations in the open-ended groups, the multiple correlation and the extent of asymmetry in the grouping (relative to the distribution of y). The underlying model used in all the experiments is given by

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

with $\beta_1 = \beta_2 = 1.0$. The grouping was performed with ten groups 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 2. 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 the 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 standardized in each experiment to have zero mean and prescribed variances (denoted σ_2^2 and σ^2 respectively). Fifty 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 $\sigma = 2$ (giving a multiple correlation of .5). The resultant marginal distribution of y is normal and the grouping symmetric about its mean with on average approximately 2½% of the observations in each of the open-ended groups. The estimators are then examined in a number of different situations. Experiments 2 and 3, by varying σ^2 and σ_2^2, vary ρ^2 and the proportion in the open-ended groups

Table 2: 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 $> A_{I-1}$
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 3: 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 4: Simulation 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	0.0[illegible]

Table 5: 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	.1596

respectively. In experiment 4 the grouping is asymmetric relative to μ_y. Experiments 5 and 6 use two convenient skewed distributions to generate x_2, while experiments 7 and 8 use the same two distributions to generate non-normal disturbances.[11] In both cases the distributions are standardised to give the same mean and variance as experiment 1.

Results for the eight experiments are given in Tables 3 to 5 for five estimators using grouped data and, for purposes of comparison, for OLS using ungrouped data. Table 3 gives the mean biases of the estimates of β_1, β_2 and σ, and Table 5 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 simulation standard deviation will be distributed approximately as t with 49 degrees of freedom. These ratios are presented in Table 4.[12]

Compare first the Maximum Likelihood estimates with the results from applying OLS to the ungrouped data. When the disturbances are normally distributed (both estimators are then consistent), the mean biases are all small and none are significantly different from zero (see Table 4). 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 relatively small and confirming findings with non-experimental data. In the case of non-normal disturbances (only OLS on ungrouped data consistent) the mean biases in the slope coefficients 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 increased. 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 well in these experiments. The root mean-square errors are very similar to those for the Maximum Likelihood estimator in all experiments (including those where u is non-normal) and the mean biases are only significantly different from zero in the cases where those for Maximum Likelihood are. The relative performance of the "two-step estimator" does not appear to be impaired by a reduction in ρ^2, an increase in the proportion of observations in the open-ended groups or asymmetric grouping. When x_2 is generated by non-normal distributions, 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. When 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.

Comparing 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, where the "two-step estimator" does, as expected, provide a considerable improvement in the estimation of β_2. The mean bias and root mean-square error in experiments 5 and 6 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. Whilst use of the moment adjustments reduces the mean biases in some cases, the improvements are not major. Finally, the root mean-square errors and mean biases for the initial iteration only estimator tend to be larger than those for the other estimators. In general either the moment adjustments or a second iteration or both appear to give substantial improvements.

7. 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 two-step estimator which approximates Maximum Likelihood is proposed. It involves a particular application of the "moment" estimator in conjunction with early termination of the monotonically convergent algorithm and is constructed as follows:

(i) calculate the sample cumulative frequency distribution of k, C_k, and estimate μ_y and σ_y by regressing $F^{-1}(C_k)$ on A_k $(k = 1,\ldots,K-1)$;

(ii) construct q_k $(k = 1,\ldots,k)$ as the conditional expectations of the marginal distribution (see Section 4), construct $\underset{\sim}{g}$ by allocating q_k to all observations in group k, and regress $\underset{\sim}{g}$ on $\underset{\sim}{X}$ (saving $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$) to give the one-step estimates of $\underset{\sim}{\beta}$ and σ;

(iii) calculate $\hat{\lambda}$, $\hat{\psi}$, $\hat{\theta}$ as defined in Section 3 and use them to adjust the one-step estimates to give the moment estimates.

(iv) construct $\hat{m}_i$ as the conditional expectations $\underset{\sim}{x}_i'\hat{\underset{\sim}{\beta}} + \hat{\sigma}M_0$ (see Section 2) and regress $\underset{\sim}{m}$ on $\underset{\sim}{X}$ (using the already extant $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$) to give the two-step estimate of $\underset{\sim}{\beta}$. Adjust the moment estimate of σ according to footnote 3 to give the two-step estimate of σ.

This "two-step estimator" should be of considerable value in applied work. In an application to the estimation of earnings functions from NTS data it was found to be superior to all the alternatives considered and to provide a very good approximation to the full Maximum Likelihood estimator. These findings were confirmed by a number of simulation experiments.

FOOTNOTES

1. An important difference from the Tobit and related models is the absence of any completely observed data. This prevents direct extension of the methods of, for example, Heckman (1979) or Fair (1977) to this problem from being useful.

2. In addition there are many other surveys where it is possible to obtain only grouped data and instances of this may increase in the future with the Privacy Act. When government departments allow access to individual data they may be reluctant to provide information on the exact values of certain variables and prepared to produce them only in ranges. Much income data, for example, comes in this form.

3. In terms of the M_r defined at the end of this section the updating rule for $\hat{\sigma}^2$ is given by

$$\hat{\sigma}^{2(n+1)} = \hat{\sigma}^{2(n)} \sum_i M_0^2 / \sum_i (M_0^2 - M_1)$$

the M_0 and M_1 being evaluated at $\hat{\underset{\sim}{\beta}}^{(n)}$ and $\hat{\sigma}^{2(n)}$.

4. Note that $\hat{\sigma}^2$ satisfies

$$\sum_i (\underset{\sim}{x}_i' \hat{\underset{\sim}{\beta}})^2 + N\hat{\sigma}^2 = \sum_i \hat{E}(y_i^2 | k_i, \underset{\sim}{x}_i)$$

as required for the M-step of the EM algorithm. Note also that the computer algorithm presented by Wolynetz (1979) based on the EM algorithm is not directly applicable here since no completely observed data are available.

5. Convergence to a local maximum is guaranteed by the EM algorithm. The log likelihood for this problem is concave and hence this local maximum is the global maximum. See Burridge (1981).

6. Since $\sum_i x_{ij} M_0 = 0$ (for all 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 evaluated at the Maximum Likelihood solution. Note that the asymptotic variance-covariance matrix of $\hat{\underset{\sim}{\beta}}$ alone can be written as $\hat{\sigma}^2(X'\hat{G}X)^{-1}$ where the elements of $\hat{G}$ are functions of the $\hat{M}_r$ alone.

7. The adjustment factors for the moment estimator described in Section 3 are applied direct to the OLS estimates. Hence the OLS estimate of σ should be used in this step rather than the iterative estimate derived in Section 2. If however the initial iteration estimates are to be used on their own, or with additional iterations without the moment adjustments, then the estimate of Section 2 should be used.

8. The standard errors presented are calculated by inversion of the information matrix and hence are only approximate. In the case of the

"two-step estimator" they should be fairly good approximations given the proximity of the estimates to the Maximum Likelihood estimates. The fact that the updating rule for $\hat{\beta}$ can be written as

$$\hat{\beta}^{(n+1)} = \hat{\beta}^{(n)} + \hat{\sigma}^{2(n)} \; (X'X)^{-1} \; \{\delta \; \log \; L/\delta\beta\}^{(n)}$$

suggests the use of $\hat{\sigma}^2(X'X)^{-1}$ as an alternative approximation. (This is equivalent to approximating G by I). This produces standard errors which are smaller than those given by 2.4% to 3.5%. The derivation of the asymptotic variance-covariance matrix of the "two-step estimator" and the comparison of alternative approximations to it remains a task for future research.

9. The "two-step estimator" took 4.7 seconds to execute on an IBM 3081 as compared with 12.0 seconds for the fully iterated Maximum Likelihood estimator (by the algorithm of Section 2). The use of this algorithm itself exhibits a saving over Newtown-Raphson due to only requiring one matrix inversion: Maximum Likelihood estimation by Newton-Raphson with the moment estimates as initial estimates took 19.5 seconds despite requiring two less iterations, whilst the moment estimator plus one Newton-Raphson step took 8.4 seconds. All times are of course notional and differ considerably from machine to machine. The ratios given in the text should however be fairly similar across machines.

10. Details, reported in an earlier version of the paper, are available from the author on request.

11. The NAG function G05DDF 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 G05DDF, and the lognormal variates were generated as L = m.exp(s.N) where m is the median, s the shape parameter and N a standard normal pseudo-random number from G05DDF. Each sample was initialised from the real-time clock.

12. These ratios can of course be generated easily enough from the entries in Tables 3 and 5. However, they provide convenient summary statistics.

REFERENCES

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

Burridge, J. (1981) - "A Note on Maximum Likelihood Estimation for Regression Models Using Grouped Data," Journal of the Royal Statistical Society, Series B, 43, 41-45.

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.

Fair, R. C. (1977) - "A Note on the Computation of the Tobit Estimator," Econometrica, 45, 1723-1727.

Greene, W. H. (1981) - "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model," Econometrica, 49, 505-514.

Heckman, J. J. (1979) - "Sample Selection Bias as a Specification Error," Econometrica, 47, 153-162.

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

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.

Wolynetz, M. S. (1979) - "Maximum Likelihood Estimation in a Linear Model from Confined and Censored Normal Data," Applied Statistics, 28, 196-206.