GROUPING ESTIMATORS ON HETEROSCEDASTIC DATA

TONY LANCASTER
University of Birmingham

This paper gives numerical comparisons of the efficiency of Ordinary Least Squares (OLS) and Grouping Estimators in simple linear regression. The disturbances are assumed to have unequal variances, and an assumption is made about the form of this heteroscedasticity. It is shown that for some types of heteroscedasticity a Grouping Estimator can be more efficient than Ordinary Least Squares.

WE CONSIDER the problem of efficient estimation in simple linear regression with heteroscedastic disturbances. The model is,

$$Y_i = \alpha + \beta X_i + \epsilon_i, \qquad (i = 1, 2, \cdots n). \tag{1}$$

The quantities involved are assumed to satisfy the assumptions of the Gauss-Markov theorem, (in particular, X is taken as non-stochastic), except that the variances of the disturbances are not necessarily equal. The minimum variance, unbiased, linear estimator, (BLUE), for this model is Aitken's Generalised Least Squares Estimator, which amounts to application of OLS after division through by the standard deviations of the disturbances. Other unbiased linear estimators are inefficient, have greater variances, in general.

One well-known exception to this rule, however, arises in the special case in which (a) $\alpha=0$, and (b) var. $\epsilon_i \propto X_i$. In this case the BLUE of β is,

$$b_0 = \overline{Y}/\overline{X}, \tag{2}$$

the ratio of the means of Y and X, which can be regarded as a member of the class of Grouping Estimators.

Following Nair & Shrivastava [6] a Grouping Estimator is formed in the following way. With k determining variables, $(X_1, \cdots X_k)$ and k parameters, $(\beta_1, \cdots \beta_k)$ to be determined, divide the n $(\geq k)$ sets of observations, $(Y_i, X_{1i}, \cdots X_{ki}; i=1, 2, \cdots, n)$ into k exclusive groups. The j'th group contains n_j (>0) sets of observations where the n_j satisfy $\Sigma n_j \leq n$. For each group find the means of Y and the X's. Then estimates of the parameters are found as the coefficients of the plane passing through the k points $(\overline{Y}_j, \overline{X}_{1j}, \cdots, \overline{X}_{kj}; j=1, 2, \cdots, k)$. Note that this only defines, for any n and k, a class of estimators the number of whose members is the number of different ways in which the observations can be divided into k groups.

For the case $k=1$, regression through the origin, this leads to the estimators,

$$b_1 = \overline{Y}_1/\overline{X}_1, \tag{3}$$

where $\overline{Y}_1$ and $\overline{X}_1$ are the means of n_1 of the n observations. When $n_1=n$, all observations are used, this estimator is the ratio estimator, 2, which happens to coincide with the BLUE when the disturbance variances are proportional to X.

For simple linear regressions, $k=2$ with one X a dummy variable, this leads to the slope estimators,

$$b_1 = (\bar{Y}_1 - \bar{Y}_2)/(\bar{X}_1 - \bar{X}_2), \tag{4}$$

where the $\bar{Y}$'s and $\bar{X}$'s are the means of exclusive groups containing n_1 and n_2 observations respectively. Particular members of this class are the estimators of Wald [10], Bartlett [2], and the authors cited above. The estimator of Wald has $n_1 = n_2 = n/2$ while that of Bartlett and Nair & Shrivastava has $n_1 = n_2 = n/3$. For both, one group contains the smallest X values and the other the largest.

For general k the estimators are defined by the matrix equation[1]

$$b_1 = (GX)^{-1}Gy. \tag{5}$$

X and y are the matrix $(n \times k)$ and vector $(n \times 1)$ of the n ungrouped observations on k determining variables and the dependent variable. G is a grouping matrix $(k \times n)$ of the type defined by Prais & Aitchison [7], and has the effect of replacing sets of elements of a column vector by their means.[2]

2

The observation that the minimum variance estimator and a Grouping Estimator coincide, although for one special sort of heteroscedasticity and a very simple regression model, suggests that it may be worth examining the relative efficiency of Grouping Estimators under heteroscedasticity in more detail. We shall make this examination for the simple linear regression model under the following pair of assumptions.

A. X is continuously distributed in two parameter lognormal form.

B. $E(\epsilon_i^2) = \lambda X_i^p$, $\quad (\lambda > 0)$.

Assumption A has two parts. The first is that the discrete distribution of the n values of X is in fact continuous. The point of this fiction is that it enables us to replace sums of the form ΣX_i^p by integrals of X^p. It does not seem too unreasonable except that for very small n it leads to nonsense results. It is a device also used by Theil [9] in a similar context. Note that we still assume the n values of X fixed in repeated samples.

The second part is the lognormality assumption. This has empirical justification for much cross-sectional economic data. It has the implication that all X values have the same sign, which we shall take to be positive.

Assumption B appears the more restrictive of the two. It states that the variance of the disturbance (or, equivalently, of y) is proportional to a power of X. This power, p, will be referred to as the degree of heteroscedasticity. The case $p=0$ gives constant variance, homoscedasticity.

There appears to have been little empirical work on the subject of heteroscedasticity in regression analysis. Goldberger [3] states that the form B with $p=2$, occurs in the savings-income relationship, and Johnston [4] suggests that "something of this kind may be expected in budget studies." Morgan et al [5] investigated the residual variance around savings-income regressions and found that "a good approximation states that the standard deviation of savings

---

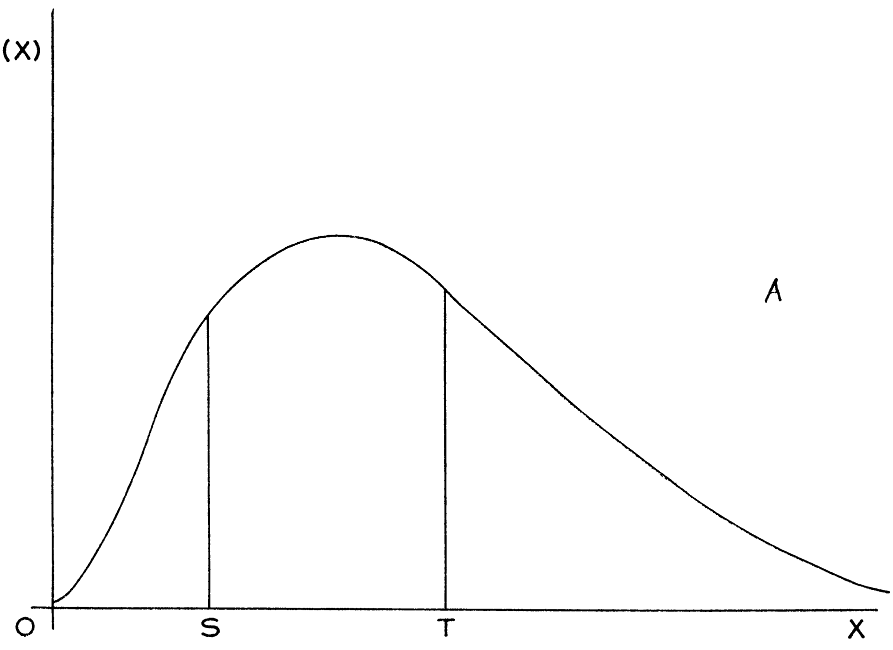
[1] Note the analogy with the equations of Instrumental Variable Estimation, $(Z^TX)^{-1}Z^Ty$, where Z is the matrix of observations on k instrumental variables.

[2] Actually G differs slightly from the matrix described by Prais and Aitchison in that it will have columns of zeros corresponding to those $(n - \Sigma_j n_j)$ observations not included in any of the k groups.

within each of several income classes is proportional to the average income of each class". (p. 203). Appendix 2 describes a small study of the dependence of the variance of company dividend payments on company profits using pooled cross-section time series data. The evidence favours a relationship of the form B with a value of p of about 1.5.

The assumptions, taken together, have the advantage of making the analysis and calculations relatively simple.[3]

The efficiency of a linear unbiased estimator, denoted by E, is equal to its variance divided into that of the BLUE. The OLS estimator of β in 1 is well defined, as is the minimum variance estimator, given assumption B. Out of the class of Grouping Estimators of β we shall consider two members, that associated with Wald and that associated with Bartlett. These Grouping Estimators, which correspond to two possible ways of arranging the observations into two groups, are illustrated in figure A



$\gamma(X)$ is the lognormal frequency function and a typical function has been drawn in. The points S and T partition the X distribution into three groups. For both the Wald and Bartlett estimators the interval $0<x<S$ represents the observations allocated to Group 1 and the interval $T<x<\infty$ the observations allocated to Group 2. Denoting by $\Gamma(X)$ the lognormal distribution function we have;

$$b_1\,(1)\text{: Wald's Estimator, } \Gamma(S) = \Gamma(T) = 1/2,$$

$$b_1\,(2)\text{: Bartlett's Estimator, } \Gamma(S) = 1 - \Gamma(T) = 1/3.$$

[3] Theil [8] found an expression for the variance of an OLS Estimator in the presence of heteroscedasticity of the type in which the standard deviation of y is proportional to its mean. His result is very complex though he made fewer assumptions than we shall. It may also be noted that a good deal is now known about the efficiency of Grouping Estimators of the Wald and Bartlett type on homoscedastic models, see, for example, Theil & Van Ijzeren [9].

Wald's estimator uses all observations and splits them at the median of the X distribution, while Bartlett's estimator discards the central third of the X observations, and uses only the tails of the distribution.

Now using assumptions A and B we require to find expressions for the variances of the BLUE, OLS, and the two Grouping Estimator, and hence to derive expressions for the efficiencies of the latter three. To illustrate the calculations involved we shall derive an expression for the variance of a Grouping Estimator of the Wald or Bartlett type, in which the X observations used comprise tails of equal area of the distribution.

$$b_1 = (\bar{Y}_1 - \bar{Y}_2)/(\bar{X}_1 - \bar{X}_2) = \beta + (\bar{\epsilon}_1 - \bar{\epsilon}_2)/(\bar{X}_1 - \bar{X}_2).$$

Hence,

$$\begin{aligned} \text{var.}\, b_1 &= \mathrm{E}(\bar{\epsilon}_1 - \bar{\epsilon}_2)^2/(\bar{X}_1 - \bar{X}_2)^2, \\ &= \lambda(\Sigma X_i^p/n_1^2 + \Sigma X_i^p/n_2^2)/(\bar{X}_1 - \bar{X}_2)^2, \end{aligned} \tag{6}$$

using assumption B, the uncorrelatedness of the ϵ_i, and the non-stochastic nature of X. The first sum is taken over the n_1 observations in group 1 and the second over the n_2 in group 2. We now make use of assumption B by setting,

$$n_1/n = \Gamma(S); \qquad n_2/n = 1 - \Gamma(T) = \Gamma(S);$$

for these are the proportions of the observations falling into the two groups. Further, $\Sigma X_i^p/n_1$ is the mean value of X^p in the interval $0 < X < S$, and similarly for $\Sigma X_i^p/n_2$, so we set,

$$\sum X_i^p/n_1 = \int_0^S X^p d\Gamma(X)/(\Gamma(S));$$

$$\sum X_i^p/n_2 = \int_T^\infty X^p d\Gamma(X)/(\Gamma(S)).$$

This then gives,[4]

$$\text{var.}\, b_1 = \frac{\lambda\left[\int_0^S X^p d\Gamma(X) + \int_T^\infty X^p d\Gamma(X)\right]}{n\left[\int_0^S X d\Gamma(X) \;-\int_T^\infty X d\Gamma(X)\;\right]^2}$$

The problem then reduces to evaluating integrals of the lognormal distribution which is easily done using Aitchison & Brown's (1) theorem on moment distributions.[5]

Expressions for the variances of the BLUE and OLS Estimators are derived in a similar way and expressions for the efficiencies may then be derived. The efficiencies of the OLS and Grouping Estimators are found to depend upon p, the degree of heteroscedasticity and upon η the coefficient of variation of the distribution of X. They do not depend upon λ, or upon n, the sample size, or upon the location parameter of the X distribution, though all these enter into the variances of the estimators.

[4] For Wald's Estimator this expression simplifies further since S = T.

[5] (1), p. 12, Theorem 2.6.

In order to study numerically the way in which the efficiency of the different estimators depends upon p and η it was decided to evaluate efficiency for a number of values of these parameters. The range of η, the coefficient of variation, was from 0.1 to 1.0 by intervals of 0.1. This covers the values which might normally be expected on empirical data. The range of p chosen was from -2 to 2 by intervals of 0.5. The interval from 0 to 2 covers types of heteroscedasticity ranging from constant variances to variances varying as the square of X. Negative values of p were considered because of the frequency with which economists use regression on deflated data. If, for the model 1, ϵ is heteroscedastic of degree p then after division through by X, the deflated disturbance will be heteroscedastic of degree $p-2$, which may be negative.[6]

The Table given in Appendix 1 gives the computed expressions for the efficiency of the OLS and the two Grouping Estimators. For each of the 9 values of p, efficiency is tabulated against η. The diagrams below present some of the results in graphical form. In diagrams 1 (a) to 1 (d) the efficiency of the OLS and Bartlett estimators is plotted against η for four values of p. In diagrams 2 (a) to 2 (c) the efficiency of these estimators is plotted against p for three values of η.

The principal features of the results are as follows.

1. In almost all cases the efficiency of the estimators diminishes as the variance of the distribution of the X observations increases. This is an obvious consequence of our assumption, B, about the form of the heteroscedasticity. The more widely spread is X, the greater the variation in the disturbance variances, and the greater the penality for failing to take this into account in the estimation procedure.

2. While the OLS Estimator has maximum efficiency, of unity of course, when the data is homoscedastic, $p=0$, the Grouping Estimator has maximum efficiency when the disturbance variance is proportional to X, $p=1$. For Bartlett's Estimator this efficiency is about 80% and almost independent of the dispersion of X. That the maximum efficiency is at $p=1$ is presumably related to the fact noted earlier that when $\alpha=0$ a Grouping Estimator coincides with the minimum variance estimator when the disturbance variance is proportional to X.

3. For heteroscedasticity of degree 1 or more the Grouping Estimator is more efficient than OLS, except when the X values show very little dispersion. The superiority of the Grouping Estimator is greater the more widely spread are the X observations.

4. Where the coefficient of variation of X exceeds about 0.5 the efficiency of the OLS Estimator falls off very rapidly with departures from homoscedasticity. The same applies to the efficiency of the Grouping Estimator with departures from heteroscedasticity of degree 1.

5. Bartlett's estimator is almost always more efficient than Wald's. This extends the well-known fact of the superiority of the 'three group' procedure for homoscedastic models.

---

[6] $E(\epsilon_i/X_i)^2=(1/X_i)^2\cdot E(\epsilon_i^2)=(1/X_i)^2\cdot\lambda X_i^p=\lambda X_i^{p-2}$.

6

The reason why these Grouping Estimators turn out to be relatively efficient when the disturbance variance varies with X is that grouping amounts to weighting the observations. Taking the arithmetic means of the largest and of the smallest observations and treating the resulting pair as single observations is to weight down the largest X observations, (with large disturbance variances), and to weight up the smaller X values. Since the optimal estimation procedure is to weight the observations in inverse proportion to their variances the grouping method comes closer to the optimum than does ordinary (unweighted) Least Squares. Of course when the disturbance variance varies *inversely* with X the grouping method weights in the wrong direction and is in

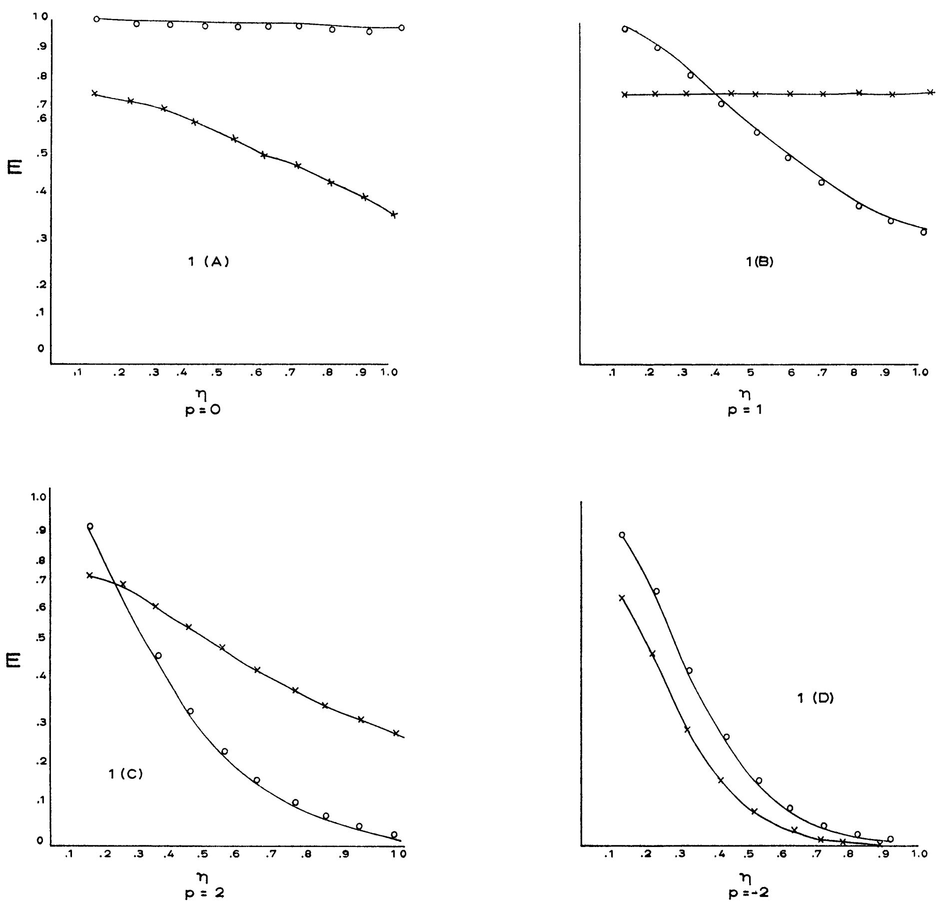


DIAGRAM 1

EFFICIENCY OF ESTIMATION (E) AS A FUNCTION OF THE DISPERSION OF X, (η), FOR FOUR VALUES OF p

o OLS (b_2); x Grouping (b_1 (2))

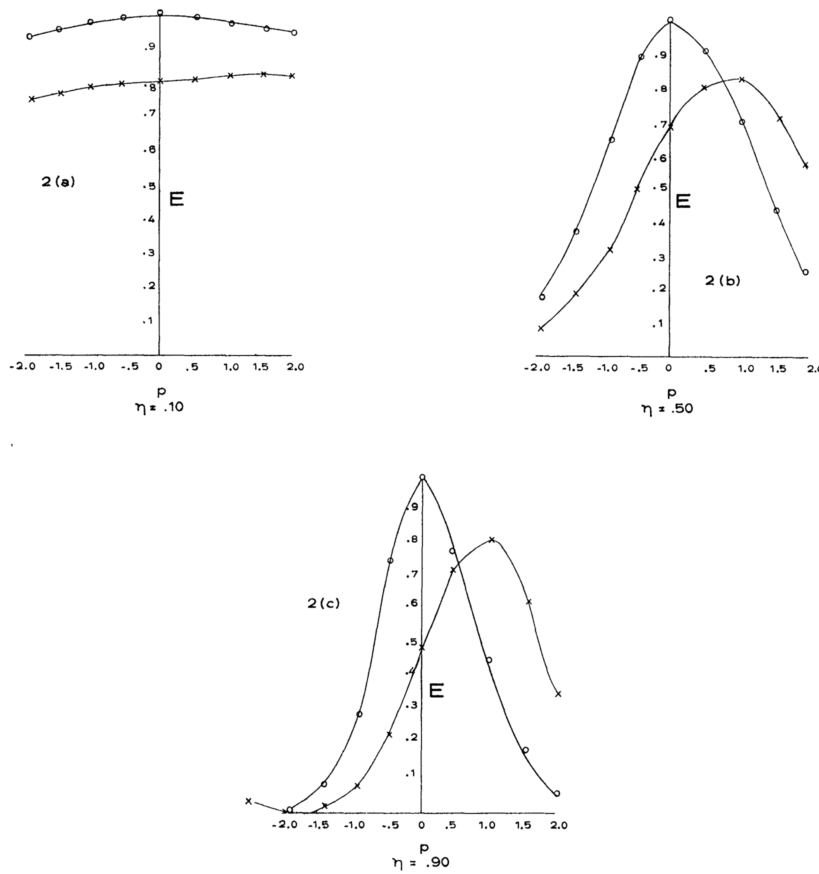


DIAGRAM 2

EFFICIENCY OF ESTIMATION AS A FUNCTION OF THE DEGREE OF HETEROSCEDASTICITY (p) FOR THREE VALUE OF η

o OLS (b_2) ; x Grouping (b_1 (2))

consequence further from the optimum than OLS, as emerges clearly in diagram 2.

Our results suggest that, in addition to its well-known computational simplicity, the Grouping Estimator performs better than OLS in the presence of heteroscedasticity of a type which the scanty evidence suggests is common with economic data. But there seems a great need for investigations into the frequency and form of heteroscedasticity in economic variables.

REFERENCES

[1] Aitchison, J., Brown, J. A. C., *The Lognormal Distribution.* Cambridge, 1957.
[2] Bartlett, M. S., "Fitting a Straight Line when Both Variables are Subject to Error." *Biometrics*, V, 1949.
[3] Goldberger, A. S., *Econometric Theory.* Wiley, 1964.
[4] Johnston, J., "*Econometric Methods.*" McGraw-Hill, 1963.

[5] Katona, G., Klein, L. R., Lansing, J. B., Morgan, J. N. *Contributions of Survey Methods to Econometrics*. Columbia, 1954.

[6] Nair, K. R., Shrivastava, M. P., "On a Simple Method of Curve Fitting." *Sankhya*, 6, 1942.

[7] Prais, S. J., Aitchison, J., "The Grouping of Observations in Regression Analysis." *Revue de l'Institut International de Statistique* 22, 1954.

[8] Theil, H., "Estimates and their Sampling Variances of the Parameters of certain Heteroscedastic Distributions." *Revue de l'Institut International de Statistique*, 19, 1951.

APPENDIX I. TABLES OF EFFICIENCIES

		b_2	$b_1(1)$	$b_1(2)$	b_2	$b_1(1)$	$b_1(2)$	b_2	$b_1(1)$	$b_1(2)$
	.10	.924	.583	.720	.956	.599	.743	.980	.613	.762
	.20	.733	.450	.543	.839	.501	.614	.925	.548	.678
	.30	.509	.297	.346	.680	.377	.452	.842	.457	.562
	.40	.316	.172	.192	.516	.258	.302	.743	.360	.438
	.50	.181	.089	.096	.371	.164	.187	.640	.270	.325
η	.60	.097	.042	.044	.256	.098	.109	.541	.196	.232
	.70	.050	.019	.019	.172	.056	.061	.450	.138	.162
	.80	.025	.008	.008	.114	.031	.033	.372	.095	.110
	.90	.013	.003	.003	.074	.017	.018	.305	.065	.075
	1.00	.006	.001	.001	.048	.009	.010	.250	.044	.050
		$p=-2$			$p=-1.0$			$p=0.0$		
	.10	.995	.624	.777	1.000	.631	.787	.995	.636	.792
	.20	.981	.587	.731	1.000	.616	.769	.981	.635	.791
	.30	.958	.532	.663	1.000	.592	.742	.960	.632	.787
	.40	.929	.467	.582	1.000	.562	.707	.934	.629	.783
	.50	.896	.400	.498	1.000	.528	.667	.905	.624	.778
η	.60	.860	.335	.417	1.000	.492	.624	.877	.619	.772
	.70	.823	.277	.345	1.000	.455	.580	.849	.614	.765
	.80	.787	.226	.282	1.000	.420	.538	.822	.608	.758
	.90	.751	.184	.229	1.000	.386	.497	.797	.602	.750
	1.00	.718	.149	.185	1.000	.354	.458	.774	.595	.742
		$p=-1.5$			$p=-0.5$			$p=0.5$		
	.10	.980	.638	.793	.957	.636	.789	.925	.631	.780
	.20	.927	.641	.793	.846	.635	.777	.748	.616	.743
	.30	.853	.646	.793	.708	.632	.758	.554	.592	.690
	.40	.770	.652	.793	.573	.629	.737	.391	.562	.630
	.50	.690	.660	.794	.457	.624	.713	.270	.528	.569
η	.60	.615	.669	.795	.363	.619	.690	.184	.492	.512
	.70	.550	.678	.797	.288	.614	.668	.126	.455	.459
	.80	.493	.688	.799	.230	.608	.648	.086	.420	.413
	.90	.443	.698	.802	.184	.602	.629	.058	.386	.372
	1.00	.400	.708	.805	.148	.595	.612	.040	.354	.336
		$p=1.0$			$p=1.5$			$p=2.0$		

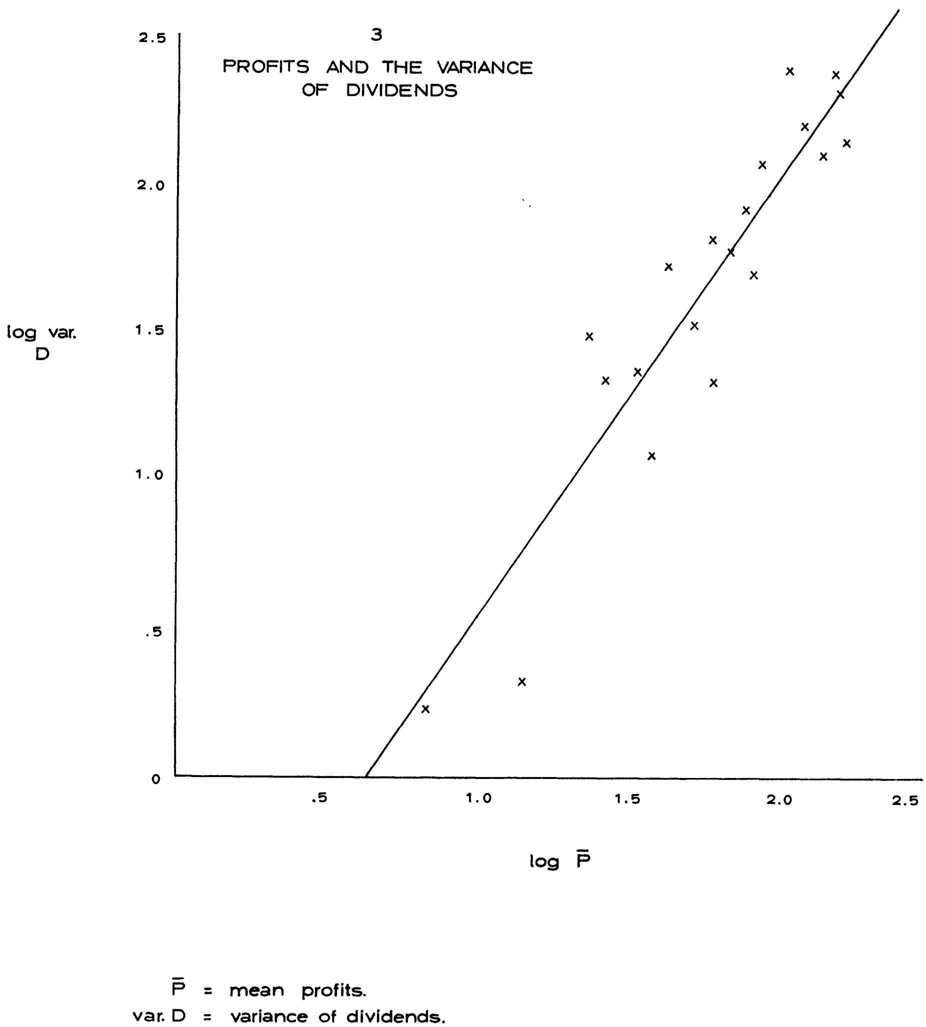
Notes: b_2 : OLS
$b_1(1)$: "Wald's" Two-Group
$b_1(2)$: "Bartlett's" Three Group

[9] Theil, H., van Ijzeren J., "On the Efficiency of Wald's Method of Fitting Straight Lines." *Revue de l'Institut International de Statistique*, 25, 1956.

[10] Wald, A., "The Fitting of Straight Lines if Both Variables are Subject to Error." *Annals of Mathematical Statistics*, XL, 1940.

APPENDIX 2. TESTING THE FORM OF HETEROSCEDASTICITY

Company dividends and profits can be assumed, as a reasonable first approximation, to satisfy a simple regression equation of the form described by 1. In addition, profits of firms in the sample used for this test were almost uniformly non-negative so that assumption B could be appropriate to this data. To test this assumption we strictly need repeated observations on dividends, D, for each level of profits, P. Replication is rarely possible with economic data so we must approximate by grouping the data such that within groups the variation in profits is small.



We have 200 pairs of observations which are the dividends and profits of 20 manufacturing companies over 10 consecutive years. These 200 pairs were ordered by the size of profits and divided into 20 groups of 10 pairs of observations. We then computed the variance of dividends and the mean profit within each group and plotted the logarithms of the observations, figure 3. If assumption B is appropriate the observations ought to be scattered around a straight line of slope p and intercept log λ. The Least Squares estimate of p is 1.47 and the r^2 is .88, which suggests that assumption B is not inappropriate for this data. Since, in addition, the profits of these firms are quite well described by a lognormal curve, and have a coefficient of variation of about .65, the tables of appendix 1 suggest that the OLS estimator of β would have an efficiency as low as .3, while the efficiency of Bartlett's estimator would be in the region of .68. Grouping this data would be twice as efficient as applying Ordinary Least Squares.

This doesn't, of course, imply that one cannot do any better than apply Bartlett's estimator to this data. Presumably an approximation to the Generalised Least Squares estimator by dividing through by $x_i^{\hat{p}/2}$ ($\approx x_i^{.75}$) would be still more efficient. What it does mean is that the Grouping estimator stands up to the heteroscedasticity of this data markedly better than Ordinary Least Squares.

We also tested the alternative hypothesis that the variance of dividends is proportional to some power, q, of its mean, i.e. $E(\epsilon_i^2) = (\alpha + \beta X_i)^q$. The case $q = 2$ gives constancy of the *coefficient of variation* of dividends. We regressed the logarithm of the variance of D within groups linearly against the log of its mean and found an r^2 of .87 and an estimate of q of 1.53. The two hypotheses perform equally well because of the smallness of the intercept in the regression equation, for clearly when this is zero the hypotheses are equivalent.