NBER TECHNICAL PAPER SERIES

ERROR COMPONENTS IN GROUPED DATA: WHY IT'S NEVER WORTH WEIGHTING

William T. Dickens

Technical Working Paper No. 43

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1985

I would like to thank Chris Martin and Phillip Bokovoy for research assistance, the participants in the Berkeley empirical micro seminar for comments, and the Institute of Industrial Relations for generous research support. I would also like to thank Martin Dooley, James Robinson, and Douglas Wholey for help in obtaining data. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Technical Paper #43
February 1985

Error Components in Grouped Data: Why it's Never Worth Weighting

ABSTRACT

When estimating linear models using grouped data researchers typically weight each observation by the group size. Under the assumption that the regression errors for the underlying micro data have expected values of zero, are independent and are homoscedastic, this procedure produces best linear unbiased estimates.

This note argues that for most applications in economics the assumption that errors are independent within groups is inappropriate. Since grouping is commonly done on the basis of common observed characteristics, it is inappropriate to assume that there are no unobserved characteristics in common. If group members have unobserved characteristics in common, individual errors will be correlated. If errors are correlated within groups and group sizes are large then heteroscedasticity may be relatively unimportant and weighting by group size may exacerbate heteroscedasticity rather than eliminate it. Two examples presented here suggest that this may be the effect of weighting in most non-experimental applications. In many situations unweighted ordinary least squares may be a preferred alternative. For those cases where it is not, a maximum likelihood and an asymptotically efficient two-step generalized least squares estimator are proposed. An extension of the two-step estimator for grouped binary data is also presented.

William T. Dickens
Department of Economics
University of California
Berkeley, CA 94720
(415) 642-5452

Despite the proliferation of micro data sets with extensive information on individuals, researchers often find it necessary to use group average data to estimate models of economic behavior. For example, some recent exercises have used the average population characteristics of states, counties or SMSAs.[1] When such data is used to estimate linear models, researchers typically weight each group observation by the group size. Under the assumption that the regression error for the individual observations used to compute the group means are independent and identically distributed, this weighting eliminates the heteroscedasticity in the regression error caused by grouping. Many econometrics texts use this as an example of the appropriate use of weighted least squares.[2]

This note argues that for most applications in economics, this method is inappropriate. There is good reason to believe that individual observations within groups are not independent when group membership is determined by a common observed characteristic. If observations are not independent, the variance of the regression errors for the grouped data may not vary substantially with group size. When heteroscedasticity introduced by differences in group size is inconsequential, weighting by group size will introduce more heteroscedasticity rather than eliminate it, causing coefficient estimates to be inefficient and estimates of the variance of the coefficients to be biased. Two examples are presented to demonstrate this. In both cases results suggest that weighting by group size produces estimates which are inferior to ordinary least squares (OLS). Further, for many purposes OLS results are inconsequentially different from ideal estimates. For those cases where OLS is not adequate, alternative estimators are proposed.

I. Grouped Data with Error Components

Consider the standard linear regression model with grouped data.

$$Y_{ij} = X_{ij}\beta + \varepsilon_{ij}$$

describes the relation between the dependent variable Y, one or more independent variables X, a vector of coefficients β, and an error ε. The first subscript denotes group, the second the individual within the group. When the micro data are not available the coefficients β can be estimated using group means

$$\bar{Y}_i = \sum_{j=1}^{N_i} Y_{ij}/N_i$$

$$\bar{X}_i = \sum_{j=1}^{N_i} X_{ij}/N_i$$

where N_i is the size of group i. If each ε_{ij} is independently and identically distributed (i.i.d.) with mean zero and variance σ^2 then the errors in the equation

$$\bar{Y}_i = \bar{X}_i\beta + \bar{\varepsilon}_i$$

will have mean zero and variance σ^2/N_i. Estimating equation 2 using OLS will produce unbiased estimates of β but not minimum variance estimates. Minimum variance estimates can be obtained by estimating the equation

(3) $$\sqrt{N_i}\,\bar{Y}_i = \sqrt{N_i}\,\bar{X}_i\beta + v_i.$$

The error $v_i = \sqrt{N_i}\,\bar{\varepsilon}_i$ has a constant variance σ^2.

The problem with this as a practical approach to grouped data is the assumption that the ε_{ij}'s are independent. The errors represent unobserved determinants of Y. To assume these errors are independent is to assume that individuals in the same group share no common unobserved determinants. When grouping is done by common characteristics such as geographic location or industry, this assumption is untenable. We know that for most observed characteristics two people within a state, SMSA, or occupation will most likely differ by less than two people from different groups. Why shouldn't we expect the same of unobserved characteristics?

If people within groups share common unobserved characteristics this can be represented by writing

$$\varepsilon_{ij} = \gamma_i + u_{ij}$$

where u_{ij} is an individual error component--those unobserved determinants of Y unique to the individual--and γ_i is a shared group component. If both the u's and γ's are i.i.d. with expected value zero and variance σ_u^2 and σ_γ^2 respectively, then $\text{var}(\bar{\varepsilon}_i) = \sigma_\gamma^2 + \sigma_u^2/N_i$.

Now if σ_γ^2 equals zero (no shared error component), and if the smallest groups have N_i's in the low hundreds while the largest groups have nearly ten-thousand members, heteroscedasticity will be substantial. The value of the $\bar{\varepsilon}_i$'s will differ by over 1000%. On the other hand, if σ_γ^2 equals σ_u^2, and group sizes are in the hundreds or higher, the variance of observations will differ by less than 1% no matter how large the differences in group sizes. With group sizes in the thousands or millions, even if σ_u^2 is substantially larger than σ_γ^2, heteroscedasticity

will be minimal and unweighted OLS will differ insubstantially from best-linear-unbiased (BLU) estimates. But, if there are large differences in group sizes, weighting each group observation by group size will introduce considerable heteroscedasticity, produce inefficient parameter estimates and biased estimates of the standard errors.

Of course, if the variance of the group error component is sufficiently small relative to the variance of the individual component, weighting by group size will be appropriate. It has been argued above that since nearly all applications of this technique in economics use data which are grouped by some common characteristic, σ_γ^2 is likely to be large. To demonstrate this for one particular case, consider the following.

II. An Example

Dooley (1982) estimates reduced form equations for labor force participation and fertility using a number of data sources. One of these sources is 1970 census data on average characteristics of women and households for each of 79 SMSAs. Separate equations are estimated for black and white women in a number of different age groups. Before estimating, Dooley weights each observation by the number of ever married women of the particular age group and race living in each SMSA. A replication of his results for the age-race category with the smallest group sizes is presented in the first column of table 1.[3] This category is chosen since for a given ratio of group to individual error component variances, the smaller the group sizes, the worse the heteroscedasticity.

What are the sizes of the error component's variances? There are a number of approaches that could be taken to determine this. Two approaches are used here. For the first it is assumed that the $\bar{\varepsilon}_i$'s are normally

distributed. The error variances and the coefficients are estimated simultaneously using maximum likelihood.[4] Column 4 of table 1 presents these estimates for the error components model. An alternative approach is to obtain initial consistent estimates of β, to compute $\hat{\varepsilon}_i = Y_i - X_i\hat{\beta}$ and then to regress the $\hat{\varepsilon}_i^2$'s on the $1/N_i$'s and a constant. The appendix demonstrates that the coefficient of the $1/N_i$'s will be a consistent estimate of σ_u^2 and the constant will be a consistent estimate of σ_γ^2. The appendix also demonstrates that these estimates of the error components variances can be used to construct an asymptotically efficient estimate of β by multiplying each observation by $\sqrt{\hat{\sigma}_\gamma^2 + \hat{\sigma}_u^2/N_i}$. The estimates presented in column 3 of table 1 were obtained by repeating this procedure using the estimates of $\hat{\beta}$ obtained one time as the initial consistent estimates for the next repetition. The process was continued until two consecutive estimates of both error component variances were the same to three decimal places. Column 1 presents group size weighted estimates, and column 2 presents unweighted OLS estimates for comparison. Both weighted and unweighted data were used to obtain initial estimates of $\hat{\beta}$. Both initial estimates converged to the same final values.

Using either the maximum likelihood or the two-step approach, the estimated variances for the group error components in both equations are substantial. Still the estimated variances of the individual error components are between nine hundred and sixteen thousand times larger than the variances of the common error components. The smallest group size is 568 and the largest is 52,690. By these estimates, the variances of the errors of smallest and largest observation could differ by over 2200%.[5] How much of a difference will this make for estimation?

The first concern is that using group sizes as weights will lead to a loss of efficiency. As a standard for comparison the efficiency of group size weighted estimates may be compared to ideal GLS estimates. If the ratio of σ_u^2 to σ_γ^2 was known to equal R, we could obtain minimum variance estimates by weighting each observation i by

$$\frac{1}{\sqrt{1+R/N_i}}$$

Column 1 in table 2 presents the difference between the variance of the group size weighted estimates and the ideal estimates as a percent of the variance of the ideal estimates.[6] The computation is made assuming the true variances of the error components are equal to the maximum likelihood estimates in table 1. Column 2 presents the same figures for unweighted OLS estimates.

The results are striking. Despite large difference in the variance of the largest and smallest groups' error terms, group size weighted estimates are substantially less efficient than properly weighted estimates. Coefficient variances are in some cases twice as large as those of the ideal estimates. Also surprising is the comparison with unweighted OLS which is nearly as efficient as the ideal estimator. Using the estimated variances from the iterated two step method produces results which are more favorable to unweighted OLS.

Using improper weights may also produce biased estimates of the variance-covariance matrix of the coefficients. Columns 3 and 4 of table 2 present the percent bias of the estimated standard errors of the coefficients for the group size weighted and unweighted estimators.[7] The

bias is substantial for the group size weighted estimator and relatively small for the unweighted estimator. Once again, using the two-step variance estimates would produce results more favorable to unweighted OLS.

It should also be remembered that the group for which these results were computed was the smallest of those analyzed by Dooley. Other groups had similar variance components but efficiency losses and biases were bigger for group size weighted estimates and smaller for unweighted OLS because of the larger group sizes.

Despite the relatively good performance of unweighted OLS, it is possible to do better, at least if the number of groups is large. Maximum likelihood is one alternative. The estimator is asymptotically efficient and the corresponding estimator of the coefficient's covariances is also consistent. However, maximum likelihood estimates are biased in small samples. Since group average data sets typically have few observations this may be a concern. One must also know the distribution of the errors a priori to use maximum likelihood.

Like maximum likelihood, the two-step GLS coefficient estimates are consistent and efficient. However, the expected value of the estimates is not defined and the small sample properties of the estimator cannot be analyzed without imposing a priori bounds on the regression errors.

None of these three estimators (OLS, ML, or TSGLS) is clearly superior for practical application. All that is clear is that, at least for these data and these estimates of the ratio of the variances of the error components, weighting by group size is inappropriate.

III. Is It Ever Worth Weighting?

The calculations presented in table 2 were done assuming that the variances of the error components were those estimated using maximum likelihood. However, the estimates are not exact so the sensitivity of the results in table 2 to changes in the assumption about the ratio of the variance of the error components should be considered.

The values for table 2 were recomputed assuming that the standard deviation of the group error component was over-estimated by two standard errors and that the standard deviation of the individual component was underestimated by two standard errors.[8] Even at these values, the loss of efficiency using group size weighted estimates is over 10% for 5 of the 13 coefficients and standard errors are understated by as much as 20%. The unweighted estimator is also very inefficient and seriously biased. Still, it produces less biased estimates of the coefficient variances for several coefficients. Other variations of this sensitivity analysis produce similar results, but perhaps these findings are unique to these data.

From the discussion in section I, we know that heteroscedasticity introduced by different group sizes is most likely to be large when groups are small on average but vary a great deal in size. It will also be large when the individual error variance is large relative to the shared error variance. It is difficult to know a priori when individual error variance will be large relative to group error variance. However, it is possible to find an example with small but highly variable group size.

IV. Grouped Binary Data

The example presented here is a model of voting in union representation elections similar to that estimated by Dickens, Wholey and Robinson

(1984).[9] The groups are prospective collective bargaining units voting on whether they should be represented by a union. The independent variables are characteristics of the bargaining unit thought to influence how workers will vote or proxies for such characteristics. The average number of workers voting in these elections is 56 and the actual numbers range from 2 to 16,953.[10]

One standard approach to the estimation of models of this type is minimum χ^2 logit (Berkson, 1953). Taking this approach it is assumed that

$$L_i = \log\left[\left(p_i + \frac{1}{2N_i}\right)/\left(1 - p_i + \frac{1}{2N_i}\right)\right] = X_i\beta$$

where p_i is the probability of a person voting union in election i, and X_i is a vector of characteristics of the i^{th} bargaining unit.[11] Adding the transformation of the observed probabilities to each side of the equation and subtracting the transformation of the observed proportion of workers voting union $(\hat{p}_i)$ yields

$$\hat{L}_i = \log\left[\left(\hat{p}_i + \frac{1}{2N_i}\right)/\left(1 - \hat{p}_i + \frac{1}{2N_i}\right)\right] = X_i\beta + \hat{L}_i - L_i \quad .$$

The difference between the two logit variables on the R.H.S. is an unobserved random variable. Gart and Zweifel (1967) suggest several good approximations to the variance of this random variable, so the method of minimum χ^2 logit involves dividing the transformed true probabilities for the i^{th} group and the X_i's by the square root of the approximate variance for the i^{th} group and running OLS on the weighted data to obtain estimates of $\hat{\beta}$. Since the variance of $\hat{L}_i - L_i$ is declining in N_i this procedure is analogous to WLS using group size weights from the case of a continuous dependent variable.

The problem with weighted minimum χ^2 estimation as a practical approach to this problem is that one must assume that the observed X's are all the bargaining unit characteristics which determine the probability of workers in an election voting union. But, it is likely that there are many unobserved characteristics that affect the probability of voting union, so a more appropriate specification is .

$$\hat{L}_i = X_i\beta + \gamma_i + \hat{L}_i - L_i$$

where γ_i is an i.i.d. group error component representing the unobserved determinants of L_i . With this specification the β's can be efficiently estimated using a two-step procedure analogous to the one used above for continuous data. Initial consistent estimates of the β's can be obtained by regressing the unweighted $\hat{L}_i$'s on the X_i's . Then a consistent estimate of σ_γ^2 can be constructed as

$$\hat{\sigma}_\gamma^2 = \frac{\sum_{i=1}^{M} (\hat{L}_i - X_i\hat{\beta})^2 - \hat{V}_i}{M - K}$$

where $\hat{V}_i$ is a consistent approximation[12] to the variance of $\hat{L}_i - L_i$.

Each $\hat{L}_i$ and X_i can then be multiplied by $\sqrt{1/(\hat{\sigma}_\gamma^2 + \hat{V}_i)}$ and the weighted data used to obtain asymptotically efficient estimates of $\hat{\beta}$. Table 3 presents unweighted OLS, weighted minimum χ^2 , and asymptotically efficient two-step GLS estimates of the voting model. The coefficient estimates for all three specifications are qualitatively similar. With the exception of the constant and the coefficient on the number of voters the signs are all the same. However, there is a big difference between the estimated standard errors for the weighted minimum χ^2 and the other two estimators. Since the variance is assumed to be known for the minimum χ^2 estimator, and much smaller than it truly is,

the coefficient standard errors are seriously underestimated. Table 4 shows the bias in the estimated standard errors for this sample, assuming the GLS estimate of the variance of the group error components is correct. The table also presents the efficiency of OLS and minimum χ^2 relative to the ideal GLS estimator. The bias is small for unweighted OLS and very large for minimum χ^2 logit. The efficiency comparisons are also stark -- OLS is nearly as efficient as GLS, while the minimum χ^2 estimates have true variances which are up to five times greater than those of GLS.

V. Conclusion

When estimating linear models using grouped data, weighting each observation by the size of the group is only appropriate if individual error terms are not correlated within groups. If the correlation is large or if group size is large, group size weighted estimates will be inefficient and coefficient variance estimates may be badly biased. This turns out to be true for both examples presented here. This suggests that the use of group sizes as weights is unlikely to be appropriate for applications where data are grouped by some common characteristic. Only if group assignment is entirely random, such as in experimental situations, does weighting by group size make sense. In many cases OLS may be a preferred alternative. For both examples examined here OLS coefficient estimates were nearly as efficient as GLS and estimated standard errors suffered little bias. Consistent estimates which are asymptotically efficient can be obtained with little additional computational difficulty, using maximum likelihood or two-step GLS.

Footnotes

1. For example, Cogan (1982) uses state average data on black teenagers to examine the causes of changes in their labor force participation rates. Higgs (1982) uses county average data to examine property accumulation by black farmers before World War I, and Dooley (1982) uses SMSA average data to investigate the determinants of women's labor force participation and fertility.
2. For example see Maddala (1977, p. 268-279), Johnston (1984, p. 293-296) and Rao and Miller (1971, p. 116-121).
3. See Dooley (1982, p. 503-505) for a complete description of the data. Dooley's results presented in his article are estimated using seemingly unrelated regressions. To simplify the presentation only single equation techniques are discussed here. The extension to multiple equation estimation is straight forward.
4. The Berndt, Hall, Hall and Hausman (1974) iterative method was used. The program was written in APL and implemented on a SP9000 micro-computer. The likelihood function is described in the Appendix.
5. The reader may note that the standard errors of the individual error components are implausibly large -- a four to eight children per woman difference in fertility and a 117 to 136 percent difference in labor force participation. This probably indicates that the group error components are not homoscedastic but have variances which decrease with group size. Although this changes the interpretation of the variances, it does not affect the analysis being performed here.
6. The formula for computing these variances can be found in the Appendix.

7. These formula are also in the Appendix.

8. These are extreme values. A Wald test rejects the hypothesis that the true standard deviations of the error components are equal to or more extreme than these values at the .001 level.

9. The labor force participation equation in the last section also involves a binary dependent variable. In that case the relatively small differences between SMSA's in the participation rate justifies the assumption of a linear probability model, and the very large size of the groups justifies the normal approximation to the binomial distribution. Neither of those conditions is met here.

10. Dickens, Wholey and Robinson (1984) contains a complete description of the data.

11. Since in some of the elections all or none of the workers voted union, the standard logit transformation of the observed proportions voting union $(\log[\hat{p}/(1-\hat{p})])$ is undefined. This is the modified logit transformation suggested by Haldane (1955).

12. Gart and Zweifel (1967) analyze several approximations of varying accuracy and complexity. All are consistent as group size goes to infinity. The better approximations have biases of less than 10% when $N_i p_i$ is greater than 1.5. $\hat{\sigma}_\gamma^2$ is consistent in the sense that by choosing M and a lower bound for the N_i's sufficiently large, the probability that $\hat{\sigma}_\gamma^2$ lies within δ of σ_γ^2 can be made arbitrarily close to one for any value of δ.

References

Berkson, J. "A Statistically Precise and Relatively Simple Method of Estimating the Bio-Assay with Quantal Response, Based on the Logistic Function," Journal of the American Statistical Association, 48, 565-99.

Berndt, E.K., B.H. Hall, R.E Hall, and J.A. Hausman, "Estimation and Inference in Non-Linear Structural Models," Annals of Economic and Social Measurement, Vol. 3, #4 (Oct. 1974), p. 653-56.

Chow, Gregory C. Econometrics. New York: McGraw-Hill, 1983.

Cogan, John F., "The Decline in Black Teenage Employment: 1950-70," American Economic Review, Vol. 72, no. 4 (September 1982), pp. 621-638.

Dickens, William T., Douglas Wholey, and James Robinson, "Bargaining Unit, Union, Industry, and Locational Correlate of Union Support in Certification and Decertification Elections," University of California at Berkeley, Department of Economics Working Paper No. 183 (March 1984).

Dooley, Martin, "Labor Supply and Fertility of Married Women: An Analysis with Grouped and Individual Data from the 1970 Census," Journal of Human Resources, Vol. 17, #4 (Fall, 1982), p. 499-532.

Gart, John J. and Zweifel, James R. "On the Bias of Various Estimators of the Logit and its Variance with Application to Quantal Bio-assay," Biometrika, Vol. 54, no. 1 (, 1967), pp. 181-187.

Haldane, J.B.S. "The Estimation and Significance of the Logarithm of a Ratio of Frequencies," Annals of Human Genetics, Vol. 20, no. (, 1955), pp. 309-11.

Higgs, Robert, "Accumulation of Property by Southern Blacks Before World War I," American Economic Review, Vol. 72, no. 4 (September, 1982), pp. 725-737.

Johnston, J. Econometric Methods. 3rd ed. New York: McGraw-Hill, 1984.

Maddala, G.S. Econometrics. New York: McGraw-Hill (1977).

Rao, Potluri and Roger LeRoy Miller. Applied Econometrics. Belmont, CA: Wadsworth, 1971.

Table 1

ESTIMATED COEFFICIENTS AND STANDARD ERRORS FOR A MODEL OF LABOR FORCE PARTICIPATION AND FERTILITY FOR BLACK WOMEN AGED 45-49 YEARS

Dependent Variable: Labor Force Participation Rate (percent)	Group Size Weighted Estimates	Unweighted Estimates	Iterated Two-Step Estimates	Maximum Likelihood Estimates
Constant	52.819	60.124	59.546	59.247
	(8.278)	(7.538)	(7.345)	(7.296)
Wife's wage ($100's/year)	.495	.620	.691	.706
	(.130)	(.146)	(.139)	(.119)
Husband's income ($100's/year)	-.496	-.521	-.554	-.562
	(.120)	(.104)	(.104)	(.091)
Nonlabor income ($100's/year)	-1.179	-1.472	-1.582	-1.665
	(1.519)	(1.193)	(1.212)	(1.361)
Education (Years completed)	1.709	.924	.911	.942
	(1.036)	(.858)	(.869)	(.839)
Percent rural SMSA residents	.050	-.076	-.066	-.063
	(.115)	(.096)	(.093)	(.099)
Unemployment rate in SMSA	-.609	-.573	-.635	-.655
	(.388)	(.354)	(.357)	(.389)
S.E. Individual Error Component	297.771	--	117.286	135.525
				(43.574)
S.E. Group Error Component	--	4.921	3.833	3.451
				(.746)

(continued)

Dependent Variable: Fertility (children/married women)	Group Size Weighted Estimates	Unweighted Estimates	Iterated Two-Step Estimates	Maximum Likelihood Estimates
Constant	5.201 (.442)	4.811 (.436)	4.840 (.433)	4.920 (.432)
Wife's wage ($100's/year)	-.023 (.007)	-.006 (.008)	-.008 (.008)	-.011 (.010)
Husband's income ($100's/year)	.000 (.006)	.011 (.006)	-.010 (.006)	-.008 (.006)
Nonlabor income ($100's/year)	.259 (.080)	.121 (.069)	.130 (.070)	.152 (.080)
Education (Years completed)	-.174 (.055)	-.098 (.050)	-.105 (.050)	-.120 (.047)
Percent rural SMSA residents	.025 (.006)	.019 (.006)	.020 (.005)	.021 (.006)
S.E. Individual Error Component	15.916	--	4.643	8.355 (3.028)
S.E. Group Error Component	--	.258	.251	.197 (.052)

Sample: 79 SMSAs (standard errors in parentheses)

Table 2

EFFICIENCY AND BIAS ANALYSIS FOR GROUP SIZE WEIGHTED AND UNWEIGHTED ESTIMATES OF LABOR FORCE PARTICIPATION AND FERTILITY MODELS

Dependent Variable: Labor Force Participation Rate	Relative Efficiency[1]		Percent Bias in Estimated Coefficient Standard Error[2]	
	Group Size Weighted Estimates	Unweighted Estimates	Group Size Weighted Estimates	Unweighted Estimates
Constant	63.87	7.38	-17.99	- .91
Wife's wage	90.08	11.26	-36.49	.14
Husband's income	95.02	7.97	-23.85	-4.93
Nonlabor income	79.49	6.22	-14.15	- 5.85
Education	83.30	9.22	-18.91	- 6.59
Percent Rural SMSA residents	49.69	3.99	- 5.72	1.12
Unemployment rate	70.44	12.16	-22.75	- 6.74
Dependent Variable: Labor Force Participation Rate				
Constant	60.38	8.82	-17.56	-0.90
Wife's wage	105.08	13.80	-41.66	0.28
Husband's income	103.11	9.94	-28.92	-4.73
Nonlabor income	79.00	7.46	-15.40	- 6.15
Education	77.94	10.94	-18.58	- 7.05
Percent rural SMSA residents	45.09	5.04	-4.86	1.07

1. Relative efficiency is defined as the percent difference between the variance of the estimator and the variance of the minimum variance estimator. See the Appendix for the formulas used to compute these numbers.

2. The bias in the Estimated Coefficients' standard errors are defined as the percent difference between the expected value of the standard estimate of the coefficient standard errors and the true standard errors given the structure of the error components. See the Appendix for the formula used to compute these numbers.

Table 3

ESTIMATED COEFFICIENTS AND STANDARD ERRORS FOR
A MODEL OF VOTING IN UNION CERTIFICATION ELECTIONS

Dependent Variable: Haldane-Logit Transformation of the Percent Voting Union	Weighted Minimum χ^2 Estimates	Unweighted Estimates	Iterated Two-step Estimates
Constant	-.482	.022	-.058
	(.018)	(.083)	(.082)
Number of eligible voters in thousands	.152	-.051	-.012
	(.002)	(.053)	(.048)
Dummy Variable: Did management consent to the election?	.304	.324	.314
	(.016)	(.050)	(.052)
Log of number of months between petition for election and date election was held	-.070	-.168	-.147
	(.004)	(.021)	(.020)
Difference between average union and nonunion wages in Industry ($/hour)	.080	.079	.073
	(.004)	(.015)	(.015)
Difference between the standard deviation of union and nonunion wages in industry ($/hour)	-.049	-.060	-.054
	(.002)	(.009)	(.009)
Percent of workers in industry who are black	1.091	1.785	1.582
	(.076)	(.423)	(.411)
Percent of workers in industry who are union members	.209	.019	.080
	(.015)	(.074)	(.073)
Percent of workforce in industry which is unemployed	-.453	-2.338	-1.965
	(.213)	(.878)	(.878)
Dummy Variable: Election held in a state with a right to work law	.003	.007	.002
	(.007)	(.039)	(.038)
Dummy Variable: Election held in a Southern state	-.009	-.066	-.058
	(.007)	(.038)	(.037)
Standard error of group error component			1.523

Data: 13,545 union certification elections held between 1977 and 1979
(standard erros in parentheses)

Table 4
EFFICIENCY AND BIAS ANALYSIS FOR GROUP SIZE WEIGHTED AND UNWEIGHTED ESTIMATES OF A MODEL OF VOTING IN UNION CERTIFICATION ELECTIONS

Dependent Variable: Haldane-Logit Transformation of the Percent Voting Union	Relative Efficiency[1]		Percent Bias in Estimated Coefficient Standard Errors[2]	
	Weighted Minimum χ^2 Estimates	Unweighted Estimates	Weighted Minimum χ^2 Estimates	Unweighted Estimates
Constant	361.50	3.52	-89.88	- .42
Number of eligible voters in thousands	43.92	.36	-95.99	10.42
Dummy Variable: Did management consent to the election?	282.34	4.15	-84.66	- 4.60
Log of number of months between petition for election and date election was held	317.52	3.19	-91.54	1.58
Difference between average union and nonunion wages in Industry ($/hour)	397.53	3.75	-89.16	-1.70
Difference between the standard deviation of union and nonunion wages in industry ($/hour)	389.13	4.05	-89.15	-2.04
Percent of workers in industry who are black	382.60	3.27	-91.56	1.30
Percent of workers in industry who are union members	470.38	3.53	-91.62	-.20

(continued)

1. Relative efficiency is defined as the percent difference between the variance of the estimator and the variance of the minimum variance estimator. See the Appendix for the formulas used to compute these numbers.

2. The bias in the Estimated Coefficients' standard errors are defined as the percent difference between the expected value of the standard estimate of the coefficient stamdard errprs and the true standard given the structure of the error components. See the Appendix for the formula used to compute these numbers.

Dependent Variable: Haldane-Logit Transformation of the Percent Voting Union	Relative Efficiency[1]		Percent Bias in Estimated Coefficient Standard Errors[2]	
	Weighted Minimum χ^2 Estimates	Unweighted Estimates	Weighted Minimum χ^2 Estimates	Unweighted Estimates
Percent of workforce in industry which is unemployed	419.08	3.72	-89.36	-1.75
Dummy Variable: Election held in a state with a right to work law	507.03	3.65	-92.27	.09
Dummy Variable: Election held in a Southern state	484.58	3.57	-92.14	.30

Appendix

Log likelihood function for variance components model:

$$L(Y|X,\beta,\sigma_\mu^2,\sigma_\gamma^2) = -\frac{M}{2}\log(2\pi) - \frac{1}{2}\left[\sum_{i=1}^{M}\log\left(\sigma_\gamma^2 + \frac{\sigma_\mu^2}{N_i}\right) + \frac{\left(y_i - X_i\beta\right)^2}{\left(\sigma_\gamma^2 + \frac{\sigma_\mu^2}{N_i}\right)}\right]$$

where Y, X, β, σ_μ^2, σ_γ^2 and N_i are defined as in section 1. M is the number of group observations.

Formula used to compute values in Tables 2 and 4:

Define $\bar{X}$ as the matrix of unweighted grouped data, $\hat{X}$ as grouped data where each row is multiplied by $\sqrt{N_i}$, and $\tilde{X}$ as the grouped data where each row is multiplied by $(1/\sqrt{\sigma_\gamma^2 + \sigma_\mu^2/N_i})$. Define $\bar{Y}$, $\hat{Y}$ and $\tilde{Y}$ similarly. Then

$$\hat{\beta}_u = (\bar{X}'\bar{X})^{-1}\bar{X}'\bar{Y}$$

$$\hat{\beta}_S = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{Y}$$

$$\hat{\beta}_I = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y},$$

and

$$\text{Var}\,(\hat{\beta}_u) = (\bar{X}'\bar{X})^{-1}\bar{X}'\begin{bmatrix}\bar{\sigma}_1^{-2} & & & 0\\ & \bar{\sigma}_2^{-2} & & \\ & & \ddots & \\ 0 & & & \bar{\sigma}_M^{-2}\end{bmatrix}\bar{X}(\bar{X}'\bar{X})^{-1}$$

$$\text{Var}\ (\hat{\beta}_S) = (\hat{X}'\hat{X})^{-1}\hat{X}' \begin{bmatrix} \hat{\sigma}_1^2 & & & 0 \\ & \hat{\sigma}_2^2 & & \\ & & \ddots & \\ 0 & & & \hat{\sigma}_M^2 \end{bmatrix} \hat{X}(\hat{X}'\hat{X})^{-1}$$

$$\text{Var}\ (\hat{\beta}_I) = (\tilde{X}'\tilde{X})^{-1} \quad .$$

where

$$\bar{\sigma}_i^2 = \sigma_\gamma^2 + \sigma_\mu^2/N_i$$

$$\hat{\sigma}_i^2 = \sigma_\gamma^2 N_i + \sigma_\mu^2 \quad .$$

The relative efficiency of the i^{th} coefficient estimate for the unweighted and group size weighted estimators are defined as:

$$[\text{Var}\ (\hat{\beta}_u)_{ii} - \text{Var}\ (\hat{\beta}_I)_{ii}/\text{Var}(\hat{\beta}_I)_{ii}$$

$$[\text{Var}\ (\hat{\beta}_S)_{ii} - \text{Var}\ (\hat{\beta}_I)_{ii}/\text{Var}(\hat{\beta}_I)_{ii}$$

The standard estimate for the variance of the unweighted and weighted estimators are

$$\hat{V}(\hat{\beta}_u) = \frac{\bar{Y}'(I-\bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}')\bar{Y}}{(M - K)} (\bar{X}'\bar{X})^{-1}$$

$$\hat{V}(\hat{\beta}_S) = \frac{\hat{Y}'(I-\hat{X}(\hat{X}'\hat{X})^{-1}\hat{X}')\hat{Y}}{(M - K)} (\hat{X}'\hat{X})^{-1}$$

where K is the number of X variables. The expected value of $\hat{V}(\hat{\beta}_\mu)$ and $\hat{V}(\hat{\beta}_S)$ are

$$E(\hat{V}(\hat{\beta}_u)) = (\bar{X}'\bar{X})^{-1} \frac{\sum_{i=1}^{M} (\sigma_\gamma^2 + \sigma_\mu^2 / N_i)(1-\bar{E}_{ii})}{(M-K)}$$

$$E(\hat{V}(\hat{\beta}_u)) = (\hat{X}'\hat{X})^{-1} \frac{\sum_{i=1}^{M} (\sigma^2 N_i + \sigma_\mu^2)(1-\bar{E}_{ii})}{(M-K)}$$

where

$$\bar{E}_{ii} = (\bar{X}(\bar{X}'\bar{X})^{-1}\bar{X}')_{ii}$$

$$\hat{E}_{ii} = (\hat{X}(\hat{X}'\hat{X})^{-1}\hat{X}')_{ii}.$$

Finally, the percent bias of the i^{th} coefficient standard error is defined as

$$[E(\hat{V}(\hat{\beta}_u))_{ii} / \text{Var } (\hat{\beta}_u)_{ii}]^{\frac{1}{2}} - 1$$

and

$$[E(\hat{V}(\hat{\beta}_S))_{ii} / \text{Var } (\hat{\beta}_u)_{ii}]^{\frac{1}{2}} - 1 \quad .$$

For the grouped binomial dependent variable case, σ_μ^2/N_i is replaced by the variance approximation presented later in this appendix. To get the "true variance" $\hat{p}_i$ is replaced by p_i where p_i is assumed to equal $X_i\hat{\beta}_I$.

Proof of the consistency of the two-step estimates of σ_γ^2 and σ_μ^2 and the asymptotic efficiency of the two-step estimator:

Theil (1971, p. 399) shows that if

i) $\underset{M\to\infty}{\text{plim}}\ M^{-1}\ \bar{X}'(\hat{V}^{-1} - V^{-1})\ \bar{X} = 0$

and

i) $\underset{M\to\infty}{\text{plim}}\ M^{-\frac{1}{2}}\ \bar{X}'(\hat{V}^{-1} - V^{-1})\ \bar{\varepsilon} = 0$

where V is the true covariance matrix of the $\bar{\varepsilon}$s and $\hat{V}$ is an estimate of that matrix then the estimators

$$(\bar{X}'V^{-1}\bar{X})^{-1}\bar{X}'V^{-1}\bar{Y}$$

and

$$(\bar{X}'\hat{V}^{-1}\bar{X})^{-1}\bar{X}'\hat{V}^{-1}\bar{Y}$$

have the same limiting distribution. Since the first estimator is the true GLS estimator, and is therefore asymptotically efficient, if conditions i) and ii) hold the second estimator, the two-step estimator, will also be asymptotically efficient.

To show that conditions i) and ii) hold,

first define

$$\hat{\varepsilon}_i = y_i - X_i(\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}$$

where the $\tilde{}$'s denote an arbitrary weighting of the X's and y's . Consistent estimates of σ_γ^2 and σ_μ^2 can be obtained as

$$\begin{bmatrix} \hat{\sigma}^2_\gamma \\ \hat{\sigma}^2_\mu \end{bmatrix} = Q^{-1} \begin{bmatrix} \sum_{i=1}^{M} \hat{\varepsilon}^2_i \\ \sum_{i=1}^{M} \hat{\varepsilon}^2_i / N_i \end{bmatrix} = \begin{bmatrix} \sigma^2_\gamma \\ \sigma^2_\mu \end{bmatrix} + Q^{-1} \begin{bmatrix} \sum_{i=1}^{M} Z_i \\ \sum_{i=1}^{M} Z_i / N_i \end{bmatrix}$$

where

$$Q = \left[\begin{array}{c|c} M & \sum_{i=1}^{M} 1/N_i \\ \hline \sum_{i=1}^{M} 1/N_i & \sum_{i=1}^{M} 1/N_i^2 \end{array}\right]$$

and

$$Z_i = (X_i(\beta-\hat{\beta}))^2 + \bar{\varepsilon}_i X_i(\beta-\hat{\beta}) + \bar{\varepsilon}^2 - \sigma^2_\gamma - \sigma^2_\mu/N_i \quad .$$

Assuming $\lim_{M\to\infty} Q/M = Q'$ and $|Q'| \neq 0$ and that the N_i's are bounded, $\hat{\sigma}^2_\gamma$ and $\hat{\sigma}^2_\mu$ will be consistent iff

$$\operatorname*{plim}_{M\to\infty} \sum_{i=1}^{M} Z_i/M = 0 \quad .$$

Expanding the first term of Z_i we get K^2 terms of the form

$$\operatorname*{plim}_{M\to\infty} (\beta-\hat{\beta})_j \, (\beta-\hat{\beta})_k \sum_{i=1}^{M} \frac{X_{ij} X_{ik}}{M} \quad .$$

Since $\hat{\beta}$ is a consistent estimator of β, these terms will equal 0. The second term of Z_i can be expanded to obtain K terms of the form

$$\operatorname*{plim}_{M\to\infty} (\beta-\hat{\beta})_j \sum_{i=1}^{M} \frac{X_{ij} \varepsilon_i}{M} \quad ,$$

all of which have the probability limits equal to zero since $\hat{\beta}$ is consistent. Finally, the last three terms can be rewritten

$$\operatorname*{plim}_{M\to\infty} \sum_{i=1}^{M} \gamma_i^2/M - \sigma_\gamma^2 + \sum_{i=1}^{M} \frac{\bar{\mu}_i^2}{N_i M} - \frac{\sigma_\mu^2}{N_i M} + \sum_{i=1}^{M} \frac{\gamma_i \bar{\mu}_i}{N_i M} \quad .$$

The first term has a probability limit of σ_γ^2 so the first two terms cancel. Since the N_i's are all positive integers, the absolute value of the third term is less than or equal to

$$\operatorname*{plim}_{M\to\infty} \left(\sum_{i=1}^{M} \bar{\mu}_i^2/M \right) - \sigma_\mu^2 = 0 \quad .$$

By the assumptions that the γ_i's and the μ_{ij}'s are independent and that the N_i's are positive integers, the probability limit of the last term is zero. Thus $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_\mu^2$ are consistent estimates of σ_γ^2 and σ_μ^2 .

Now, to show that condition (i) holds, note that the ij^{th} element of the matrix may be written

$$\operatorname*{plim}_{M\to\infty} \sum_{k=1}^{M} \frac{(\bar{X})_{ki}(\bar{X})_{kj}}{M} \left[\frac{1}{\hat{\sigma}_\gamma^2 + \hat{\sigma}_\mu^2/N_k} - \frac{1}{\sigma_\gamma^2 + \sigma_\mu^2/N_k} \right] =$$

$$\operatorname*{plim}_{M\to\infty} \sum_{i=1}^{M} \frac{(\bar{X})_{ki}\,(\bar{X})_{kj}}{M} \; V(N_k) \quad .$$

If the N_k's are positive integers with an upper bound of U and values of $\hat{\sigma}_\gamma^2$ or $\hat{\sigma}_\mu^2 < 0$ are treated as $\hat{\sigma}_\gamma^2$ or $\hat{\sigma}_\mu^2 = 0$, then

$$V(N_k) \leq \max(V(1), V(U))$$

and

$$\operatorname*{plim}_{M\to\infty} \sum_{k=1}^{M} \frac{(\bar{X})_{ki}\,(\bar{X})_{kj}}{M}\, V(N_k) \leq \operatorname*{plim}_{M\to\infty} \max(V(1),V(U)) \sum_{k=1}^{M} \frac{(\bar{X})_{ki}\,(\bar{X})_{kj}}{M} \quad .$$

Since $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_\mu^2$ are consistent estimates of σ_γ^2 and σ_μ^2, both $V(1)$ and $V(u)$ go to zero in probability as M goes to infinity. If $\operatorname*{plim}_{M\to\infty} \bar{X}'\bar{X}/M = Q$, a matrix of constants, then the first condition holds.

Similarly the ij^{th} element of the second condition can be written:

$$\operatorname*{plim}_{M\to\infty} \sum_{k=1}^{M} \frac{(\bar{X})_{ki}\bar{\varepsilon}_k}{\sqrt{M}}\left[\frac{1}{\hat{\sigma}_\gamma^2 + \hat{\sigma}_\mu^2/N_k} - \frac{1}{\sigma_\gamma^2 + \sigma_\gamma^2/N_k}\right] \quad .$$

By the same argument made above, this can be rewritten as

$$\operatorname*{plim}_{M\to\infty} \max(V(1),V(U)) \sum_{k=1}^{M} \frac{(\bar{X})_{ki}\,\bar{\varepsilon}_k}{\sqrt{M}} \quad .$$

Once again the consistency of $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_\mu^2$ assures that $V(1)$ and $V(U)$ have a zero probability limit. The distributional assumptions made in section I are sufficient for

$$\operatorname*{plim}_{M\to\infty} \sum_{k=1}^{M} \frac{(\bar{X})_{ki}\,\bar{\varepsilon}_k}{\sqrt{M}}$$

to have a limiting normal distribution. Thus the second condition holds and two-step estimation is fully efficient.

The proof of the consistency and efficiency of the two-step GLS minimum χ^2 logit follows the same steps as above, except that $\hat{V}_i$ replaces $\hat{\sigma}_\mu^2/N_i$, the consistency of $\hat{V}_i$ is proved by Gart and Zweifel (1967) and the consistency of $\hat{\sigma}_\gamma^2$ follows easily from the consistency of $\hat{V}_i$.

Approximation to the Variance of $\hat{L}_i - L_i$

The approximation used here is one developed by Goodman and reported by Gart and Zweifel (1967). It is an unbiased estimate of the true variance of $\hat{L}_i - L_i$, except for terms of $O(N_i^{-4})$ or higher, and converges in probability to the true variance as $N_i \to \infty$.

$$\hat{V}_i = V_3 + [(N_i+1)^2/(2N_i^2)] \left[\frac{1}{N_i\hat{p}_i+1} - \frac{1}{N_i-N_i\hat{p}_i+1}\right]^2 - \frac{V_3}{2}\left[V_3^2 - \frac{4V_3}{N_i} + \frac{4}{N_i^2}\right]$$

where

$$V_3 = \frac{N_i + 1}{N_i}\left[\frac{1}{N_i\hat{p}_i+1} + \frac{1}{N_i-N_i\hat{p}_i+1}\right],$$

N_i is the number of workers voting, and $\hat{p}_i$ is the proportion voting union.