



The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search

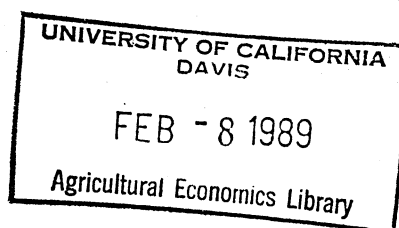
<http://ageconsearch.umn.edu>

aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

DRAFT



Estimation of Variances in the Grouped Heteroskedasticity Model

by

James K. Binkley

August 1988

AAEA paper presented at its annual meetings,

Purdue University, West Lafayette, Indiana.

Dept. of agricultural economics

Introduction

A common form of heteroskedasticity occurs when data used to estimate a relation are drawn from two or more distinct populations. Formally, suppose we have the model

$$Y_i = X_i \beta + e_i, \quad i = 1 \dots M \quad (1)$$

in which Y_i and e_i are n_i -element vectors and X_i is a $k \times n_i$ matrix, all pertaining to group i . β is a k -element vector of coefficients, invariant over groups. σ_i^2 is the variance of e_i , which in general differs from σ_j^2 , $j \neq i$.

This model is usually termed the grouped heteroskedasticity model. If the error variances were known, estimation of β would be accomplished by straightforward generalized least squares (GLS), i.e., OLS on data weighted by σ_i^{-1} . In practice, (1) is estimated either by OLS (knowingly or unknowingly ignoring the heteroskedasticity) or by an operational version of GLS ("feasible GLS") in which estimates of the error variances replace the unknown values.

Some aspects of this model have received detailed examination in the literature. Taylor (1977, 1978) analyzed the gain of feasible GLS over OLS and its loss compared to GLS. He found the former typically large and the latter typically small. Swamy and Mehta (1979) examined conditions under which feasible GLS is not optimal and developed an alternative estimator.

However, there is one aspect of the grouped heteroskedasticity model that has not been considered. There are two standard approaches to estimating the error variances, differing in whether the information about the common β coefficients is imposed. Although different authors make different recommendations, there is little if any information available on the comparative performance of alternative procedures. The purpose of this paper

is to provide such information. Since any weighting scheme results in unbiased estimates, the concern is estimator efficiency and reliability in making inferences. The analysis primarily will be based on Monte Carlo simulation. Although with this method results can depend on the specific nature of the experiment, it is a useful tool for detecting broad differences among alternative procedures. We find that, under some conditions, such differences exist for the grouped heteroskedasticity model. Since in most cases these conditions can be identified, the results of the study are of potential practical use.

Alternative Approaches

The standard method to estimate (1) is, as suggested above, weighted least squares. Given $\hat{\sigma}_i^2$, an estimate of σ_i^2 , one computes $Y_i^* = Y_i \hat{\sigma}_i^{-1}$ and $X_i^* = X_i \hat{\sigma}_i^{-1}$ and then obtains $\hat{\beta} = (X^{*'} X^*)^{-1} X^{*'} Y^*$, where X^* and Y^* contain the data from individual groups appropriately stacked. The issue here is the estimation of σ_i^2 . A common and simple method is to use the unbiased estimate from a separate OLS regression on group i , that is,

$$\hat{\sigma}_i^2 = \frac{(Y_i - X_i b_i)' (Y_i - X_i b_i)}{n_i - k}, \quad (2)$$

where $b_i = (X_i' X_i)^{-1} X_i' Y_i$. This is the estimator employed by Taylor and by Swamy and Mehta. It is probably the estimator most often used in practice, since it is readily available from standard computer output. However, most authors are doubtful that this is the best procedure. As stated by Judge et. al., it makes

no allowance for the fact that each b_i is an estimate of the same β . If we incorporate the additional information given by this restriction it is likely that the resulting $\hat{\sigma}_i^2$'s will be more efficient, and, although there will be no difference asymptotically, this may lead to a $\hat{\beta}$ that is more efficient in finite samples. (1985, p. 429)

Their suggestion is to obtain a pooled OLS estimate of β , i.e.,

$b = (X'X)^{-1}X'Y$, and then take

$$\hat{\sigma}_i^2 = \frac{(Y_i - X_i b)'(Y_i - X_i b)}{n_i - k} \quad (3)$$

This is somewhat more troublesome to employ, since it requires the partitioning of a residual vector.

In his 1977 paper, Taylor notes the likely inefficiency of (2) but uses it because of its popularity and tractability. In his text, Johnston (1984) recommends (3) without discussion, although in a related case, estimating a demand system with cross-equation restrictions, he states it is a "moot point" whether the first stage residuals for estimating the error structure should be based on OLS with or without the constraints. Kmenta (1986) implicitly endorses (3), since in his discussion of the cross-sectionally heteroskedastic and timewise autoregressive model he states that first stage residuals should be obtained from OLS on all groups simultaneously. Pindyck and Rubinfeld (1985) also say that pooled residuals should be used. The one exception is the textbook by Fomby, et al., in which only (2) is mentioned.

Since (3) incorporates additional information, it is plausible that it is the better estimator. However, this should not be accepted without question. Although it is obviously consistent, it is biased in small samples. This follows from the fact that (2) is unbiased. Since b_i is the estimate of β that minimizes (2), (3) is always larger. It therefore has an upward bias. It is reasonable to expect the bias to increase with the difference in error variances. This suggests that whether and to what extent (3) is better than (2) depends upon the degree of heteroskedasticity.

So far the discussion has been concerned with initial estimates of the σ_i^2 's, which lead to estimates of β . But these estimates can themselves be

used to obtain new estimates of the variances, which in turn generate revised estimates of β , and so on in an iterative fashion. Based on (3), such a method is iterative maximum likelihood under normality (Oberhofer and Kmenta, 1974), and this provides another approach to estimating β . Since it is a maximum likelihood method, it has optimal large sample properties. Whether it is an improvement in small samples depends on the performance of the revised variance estimates.

The Monte Carlo Experiment

The Monte Carlo experiment was conducted using the following model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + e_i, \quad i = 1 \dots M$$

where i refers to group. The core of the analysis was conducted with $M=2$, the case considered by Taylor and by Swamy and Mehta. All β 's were set at 1. e_i was $N(0, \sigma_i^2)$, with differences in variances as described below. Each experiment was replicated 1000 times. The general procedure was to estimate σ_i^2 in different ways and to use the estimates in weighted least squares.

Three methods of estimating σ_i^2 were employed:

WLS1: $\hat{\sigma}_i^2$ from formula (2)

WLS2: $\tilde{\sigma}_i^2$ from formula (3)

WLS3: σ_i^2 estimated with (3), using the residuals from WLS2. This is two iterations of MLE estimation.

The model was also estimated with OLS and GLS, i.e., using the true variances as weights. Experiments were conducted with three sample sizes for each group: 10, 20, and 40.

The X 's, which were fixed in each individual experiment but differed across experiments, were generated as uniform random variables. Two approaches were employed. In the first, the variables for each group were $\mu(10,30)$ in all experiments. In the second, the variation in the data was increased along with that of the error by generating the elements of X_i as

$\mu(10\sigma_i, 30\sigma_i)$. We will refer to these as X1 and X2, respectively. With X1, heteroskedasticity can be thought of as arising simply due to differences in the amount of unexplained behavior, whereas with X2 it can be thought of as due to size differences, as might occur with data from different countries.¹

Results

The efficiency of the different weighted least squares estimates of β was measured as follows. In each experiment of 1000 replications, the sample variance of each of the four coefficients in the model as estimated by the different methods was calculated. GLS was used as a benchmark by taking ratios of the sample variances of the other estimators to that of GLS. These were averaged over the four coefficients. Thus, letting r_i be such an average, we would expect $r_i \geq 1$, although sampling error can reverse this. If $r_i < r_j$, this is evidence that method i is more efficient than method j . These ratios are presented in table 1 for the three sample sizes. In each section, the first line is for the homoskedastic case and movement down the columns involves a greater degree of heteroskedasticity, which we define as the ratio of the larger variance to the smaller and which will be denoted by λ .

Let us first consider the results for OLS. Comparing the columns for OLS to the others, we see that at a low level of heteroskedasticity, OLS was superior to all the weighted least squares estimators. However, this superiority did not last long. When the variance in one equation was twice that in the other, OLS lost its advantage, even at the smallest sample size. This agrees with Taylor's (1977) analytic results. It does suggest that unless λ exceeds two - which it often may not in models of this sort - ignoring heteroskedasticity may at worst generate only a negligible loss in efficiency. But clearly, the loss can be very large, as occurred with X1 at the higher levels of heteroskedasticity examined. The much poorer performance

Table 1. Ratio of Sample Variance of Estimators to that of GLS, Two Group Case.

		X1				X2			
λ		OLS	WLS1	WLS2	WLS3	OLS	WLS1	WLS2	WLS3
n = 10	1.0	1.000	1.147	1.065	1.134	-----	-----	-----	-----
	1.1	1.004	1.124	1.054	1.117	1.006	1.147	1.068	1.140
	1.5	1.029	1.135	1.055	1.118	1.034	1.127	1.052	1.092
	2.0	1.095	1.145	1.077	1.144	1.010	1.116	1.075	1.109
	3.0	1.249	1.107	1.056	1.075	1.203	1.094	1.077	1.086
	5.0	1.610	1.123	1.140	1.114	1.306	1.139	1.098	1.117
	10.0	2.739	1.069	1.193	1.045	1.612	1.154	1.159	1.145
	15.0	3.172	1.097	1.332	1.106	2.121	1.107	1.218	1.143
	20.0	4.725	1.094	1.587	1.135	1.953	1.141	1.168	1.127
n = 20	50.0	11.203	1.049	2.721	1.338	2.347	1.139	1.303	1.203
	1.0	1.000	1.052	1.038	1.060	-----	-----	-----	-----
	1.1	.999	1.070	1.053	1.080	1.001	1.065	1.049	1.073
	1.5	1.040	1.060	1.041	1.065	1.049	1.076	1.058	1.080
	2.0	1.120	1.060	1.049	1.063	1.062	1.044	1.033	1.045
	3.0	1.297	1.061	1.056	1.062	1.186	1.047	1.042	1.049
	5.0	1.758	1.030	1.045	1.028	1.449	1.049	1.065	1.058
	10.0	2.566	1.037	1.080	1.034	1.484	1.020	1.029	1.019
	15.0	3.884	1.013	1.116	1.011	2.015	1.060	1.107	1.071
n = 40	20.0	5.103	1.015	1.161	1.011	1.980	1.029	1.055	1.031
	50.0	11.062	1.012	1.419	1.014	1.831	1.055	1.073	1.057
	1.0	1.000	1.029	1.025	1.032	-----	-----	-----	-----
	1.1	1.003	1.028	1.025	1.031	1.001	1.034	1.030	1.038
	1.5	1.042	1.026	1.023	1.027	1.036	1.020	1.018	1.021
	2.0	1.128	1.024	1.021	1.026	1.102	1.025	1.025	1.026
	3.0	1.306	1.010	1.009	1.011	1.204	1.019	1.020	1.020
	5.0	1.772	1.016	1.022	1.015	1.348	1.018	1.016	1.018
	10.0	2.997	1.011	1.033	1.008	1.487	1.026	1.025	1.027
	15.0	3.934	1.007	1.042	1.006	1.673	1.018	1.024	1.018
	20.0	5.458	1.008	1.054	1.006	1.544	1.020	1.024	1.020
	50.0	11.791	1.004	1.091	1.004	1.962	1.014	1.029	1.015

of OLS here than in the X2 case follows since, at any given λ , there is much less variation in the pooled X1 data. As a result, the OLS estimates are more variable.

The most interesting aspect of the Monte Carlo results involves comparing WLS1 and WLS2. It is evident that, as conjectured above, which was the better estimator in the two group case depended upon λ . If this was low, then the use of residuals from a pooled OLS regression on all $2n$ observations was the better procedure. However, with stronger heteroskedasticity (generally, λ exceeding 3 or 5), the method using residuals from separate regressions became superior. This occurred in both cases and for all three sample sizes. For $n=40$, and with X2 regardless of sample size, this superiority was typically not of large consequence. However, for $n=10$ or 20 and severe heteroskedasticity the advantage of WLS1 was substantial under data condition X1.²

The reason for these results was made evident by examining some additional information from the experiments. As expected, the estimates of error variance using formula (3) were always biased upward. However, the bias for the smaller variance was always proportionately larger. As a result, λ was underestimated. This is illustrated by data in table 2, which are estimates of λ for WLS1 and WLS2 for the experiments with $n=20$. It is quite clear from these that with strong heteroskedasticity formula (3) can lead to a weighting scheme that only partly eliminates the problem, in which case its efficiency suffers. This is more serious with the X1 data because of the weaker OLS estimates used to generate the pooled residuals.

Generally, the results for WLS1 and WLS2 display two counteracting forces. If λ is low, then the loss associated with implicitly assuming the error variances are the same is also small and is outweighed by the benefits of a larger sample. With large differences in error variances, the relative

Table 2. Estimated Degrees of Heteroskedasticity, $n = 20$.

Actual	X1		X2	
	WLS1	WLS2	WLS1	WLS2
1.0	.99	.98	-----	-----
1.1	1.09	1.08	1.05	1.03
1.5	1.44	1.38	1.50	1.39
2.0	1.99	1.85	1.98	1.78
3.0	2.96	2.63	2.90	2.59
5.0	4.98	4.00	5.04	4.18
10.0	10.13	6.73	10.06	8.35
15.0	15.29	8.73	14.57	11.04
20.0	20.29	9.83	19.85	15.38
50.0	48.58	13.06	47.53	38.87

importance of these two effects reverses. As a result, in the first case WLS2 is more efficient while WLS1 is in the second.³

WLS3 is a second iteration of WLS2 and as such represents the first two iterations of maximum likelihood estimation. The results in table 1 indicate that with severe heteroskedasticity this method substantially improved on WLS2. But even at the largest sample size considered, unless λ exceeded three there was no gain - indeed a slight loss - associated with a second iteration of WLS2. Furthermore, the experimental results suggest that if λ is large the iterative method is at best only marginally better than the conventional procedure using separate OLS residuals. Although additional iterations might bring additional gains, a brief analysis described below suggests this is likely to be important in few cases. Furthermore, for the two-group case there does not seem to be much room for improvement if sample size is reasonably adequate.

From these results it is fairly evident that no one estimation method dominates the others and no method is dominated by the others. Each was best in several individual experiments. If we categorize the degree of heteroskedasticity as "low," "medium," or "high," with the specific definitions depending upon sample size, the corresponding optimal technique would seem to be OLS, WLS2, and either WLS1 or WLS3.

How, then, should one proceed? In a world of total ignorance of λ , WLS1 appears to be the minimax approach. In the experiments, it often outperformed the others and was measurably less efficient only with small λ and a small sample (when OLS was best). It is also easy to implement. However, one need not proceed in complete ignorance. Information about λ is readily available from estimates of regression variance for the individual groups. If this suggests λ is quite large, then one would not be inclined to employ WLS2 except possibly as the first step in an iterative process. This is

particularly true if the explanatory variables have the characteristics of our X1 data, which is also easily ascertained. If λ is small, then the logical choice would seem to be between OLS and WLS2.

Additional Experiments

The two-group results show that use of pooled residuals in the first stage of estimation leads to a gain due to the larger sample size and a loss due to ignoring the heteroskedasticity. Thus, the relative performance of WLS2 should improve with more groups, since the number of observations in the first stage of WLS1 remains constant. To investigate this, a set of experiments with more groups was conducted. The error variances were chosen in an attempt to get a reasonable coverage of the possibilities. The results are presented in table 3, and they have the same interpretation as before. Here we have confined attention to the smaller sample sizes. The heteroskedasticity is characterized by spelling out the variances, with σ_1^2 always being one.

The relative performance of WLS2 indeed improved. This is especially apparent with $n = 10$. At low levels of heteroskedasticity, it clearly outperformed WLS1 regardless of the nature of the explanatory variables, and with X2 data it did so with virtually all levels of heteroskedasticity. It thus appears that with several groups and a quite small sample (with $n = 10$ there were six degrees of freedom for each group), WLS2 will often be preferred to WLS1, more so than in the two group case. As before, however WLS1 appears much more efficient with X1 data and large differences in error variance.⁴ Indeed, the largest advantage of WLS1 over WLS2 in the study occurred in this group of experiments, suggesting that the correct choice requires particular attention to the data under these conditions.

Table 3. Ratio of Sample Variances of Estimators to That of GLS, Various Three and Four Group Cases, $\sigma_1^2 = 1$.

σ_2^2	σ_3^2	σ_4^2	X1				X2			
			OLS	WLS1	WLS2	WLS3	OLS	WLS1	WLS2	WLS3
n = 10										
1	1	--	1.000	1.195	1.088	1.163	-----	-----	-----	-----
1	3	--	1.258	1.220	1.102	1.155	1.263	1.235	1.125	1.163
2	3	--	1.230	1.232	1.120	1.177	1.149	1.181	1.081	1.125
3	3	--	1.240	1.198	1.139	1.181	1.135	1.188	1.098	1.139
1	5	--	1.670	1.204	1.121	1.144	1.451	1.168	1.135	1.132
5	5	--	1.645	1.192	1.193	1.172	1.184	1.176	1.060	1.103
1	10	--	2.732	1.203	1.158	1.101	2.066	1.163	1.122	1.111
10	10	--	2.192	1.196	1.252	1.145	1.305	1.182	1.102	1.139
5	10	--	1.952	1.222	1.188	1.146	1.471	1.200	1.129	1.138
1	20	--	4.576	1.199	1.295	1.157	2.718	1.187	1.282	1.171
10	20	--	3.570	1.129	1.351	1.140	1.622	1.190	1.137	1.157
1	50	--	10.712	1.132	1.404	1.126	2.730	1.200	1.318	1.207
50	50	--	10.532	1.048	2.840	1.407	1.637	1.190	1.108	1.108
1	1	1	1.000	1.277	1.111	1.185	1.000	1.306	1.110	1.194
3	3	3	1.233	1.290	1.116	1.158	1.138	1.222	1.082	1.126
2	3	4	1.255	1.268	1.102	1.148	1.209	1.231	1.103	1.148
5	10	20	2.623	1.224	1.244	1.127	1.731	1.256	1.159	1.169
50	50	50	8.640	1.141	2.491	1.324	1.822	1.208	1.105	1.116
n = 20										
1	1	--	-----	-----	-----	-----	1.000	1.096	1.072	1.101
1	3	--	1.256	1.079	1.055	1.071	1.246	1.060	1.050	1.058
2	3	--	1.212	1.067	1.058	1.073	1.125	1.063	1.043	1.061
3	3	--	1.305	1.098	1.077	1.086	1.155	1.065	1.046	1.060
1	5	--	1.693	1.067	1.061	1.065	1.514	1.073	1.063	1.070
5	5	--	1.653	1.088	1.094	1.075	1.236	1.086	1.062	1.083
1	10	--	2.753	1.083	1.086	1.083	1.881	1.073	1.080	1.072
10	10	--	2.815	1.035	1.094	1.028	1.365	1.065	1.051	1.062
5	10	--	2.201	1.059	1.100	1.057	1.555	1.073	1.058	1.061
1	20	--	4.625	1.064	1.068	1.065	2.406	1.071	1.075	1.064
10	20	--	3.576	1.036	1.130	1.041	1.743	1.063	1.058	1.058
1	50	--	12.354	1.066	1.166	1.062	2.119	1.088	1.093	1.082
50	50	--	10.426	1.027	1.595	1.045	1.766	1.074	1.067	1.077
1	1	1	1.000	1.085	1.049	1.071	-----	-----	-----	-----
3	3	3	1.248	1.108	1.082	1.092	1.166	1.082	1.056	1.073
2	3	4	1.271	1.092	1.064	1.085	1.209	1.095	1.064	1.079
5	10	20	3.118	1.049	1.111	1.042	1.952	1.077	1.076	1.064
50	50	50	8.399	1.026	1.466	1.038	1.970	1.078	1.067	1.061

With $n = 20$, the pattern of results was similar, but in corresponding experiments either the magnitude of the WLS2 advantage over WLS1 declined or its disadvantage increased. Thus, for example, while in the X2 case WLS2 was superior in about the same set of experiments (actually one less), this superiority was seldom of any significance and certainly less than before. The difference between $n=10$ and 20 illustrates that increasing sample size is a more important consideration when it is small in the first place. In fact, with $n=20$ the results do not greatly differ from the corresponding ones for the two-group case. The only change of any importance is the indication that with several groups and X2-type data, users of the pooled approach are nearly assured of a gain regardless of the extent of heteroskedasticity (assuming it exists at all). However, the magnitude of the gain is not likely to warrant any additional effort that may be required.

Now consider WLS3, the iterative approach. With $n = 20$, its performance was similar to that in the two-group case, tending to parallel that of WLS1. But with ten observations, in several instances it substantially improved on both WLS1 and WLS2. These occurred with X1 data, in the middle to upper ranges of heteroskedasticity. Still, the fact that in the majority of experiments its variance exceeded that of WLS1 and/or WLS2 suggests that iterative estimation to improve efficiency is at best problematic.

However, our results involve an estimator with just two iterations. In practice, more would typically be used, generally until some convergence criterion is met. Although it is beyond the scope of this paper to extensively analyze this procedure, a representative group of experiments was repeated with more iterations added to WLS3. For all cases in which WLS3 was inferior to WLS2 (and several others as well), further iterations made matters worse, even with 40 observations per group. Only with strong heteroskedasticity was there an improvement. Even then, the efficiency

achieved after five iterations was at best only marginally better than that of WLS1, the much simpler procedure. This is similar to the result obtained by Kmenta and Gilbert (1968) in the case of seemingly unrelated regression. They also found that iterations often reduced estimator efficiency.

Firm conclusions about iterative estimation in the grouped heteroskedasticity model will require additional research. However, based on our results we can conclude that its routine application is unwise. Furthermore, the evidence here is that, unless sample size is quite small, it is of limited value. Although a second iteration often significantly improved on WLS2, additional iterations seldom did. More typically, efficiency was reduced, and in most cases, the simpler approach provided by WLS1 was nearly as good and in numerous instances better.

Summary and Concluding Remarks

In this paper we have examined the estimation of error variances in the grouped heteroskedasticity model and the implications for feasible generalized least squares estimation of the common coefficient vector. These variances are always estimated using residuals from a first stage OLS. This can be applied to each group separately or on all groups simultaneously. Since the latter incorporates the common coefficient vector into the first stage, it is widely believed to be a more efficient procedure and thus is most often recommended by textbooks.

The results of this study would not support this recommendation. Although using pooled residuals generates a gain due to the larger sample, this comes at the cost of biased variance estimates due to ignoring the heteroskedasticity in the first stage. The desirability of this approach depends upon the relative strength of these two effects. In simulations based on two groups, it tended to be demonstrably better only when heteroskedasticity was not too serious and the sample size was small. With

substantial differences in group variances, residuals from separate regressions led to better estimates. In some cases, the differences were pronounced. It was found that problems with pooled residuals could largely be eliminated by iterating the method once, with further iterations usually reducing efficiency. Generally, only limited support for iterative estimation was found, since in few cases was it more than negligibly better than using separate residuals without iteration.

With more than two groups, the relative performance of the pooled procedure improved. When the groups each had few observations, it was often substantially more efficient than the method using separate residuals. However, even then it performed rather badly in some of our experiments, and users need to be aware of this possibility. For moderate to large samples, results of the study suggest that the advantage of pooled OLS in the first stage is not likely to be large even with many groups, and there remain cases in which it can have a substantial disadvantage. Since in general it is more troublesome to employ, we see little recommendation for this method when sample size is adequate.

A conclusion from this study is that none of the methods considered is always best, for this depends upon circumstance. Although this may be unfortunate, it is perhaps comforting that a second conclusion is that the simplest and hence most often used procedure is in many cases probably the best procedure. That a method which ignores information about the common vector of coefficients can often outperform one which uses it might come as a surprise. However, the use of pooled residuals ignores the fact that the error variances are not the same, and results here suggest this can be an important consideration. Perhaps this should not be surprising. It is, after all, the estimation of the error structure that is at issue.

References

- Fomby, Thomas B., R. Carter Hill, and Stanley R. Johnson. Advanced Econometric Methods. (New York: Springer-Verlag, 1984.)
- Johnston, J. Econometric Methods. (New York: McGraw-Hill, 1984.)
- Judge, George G., W.E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee. The Theory and Practice of Econometrics. (New York: Wiley, 1985.)
- Kmenta, Jan. Elements of Econometrics. (New York: Macmillan, 1986.)
- Kmenta, Jan and Roy R. Gilbert. "Small Sample Properties of Alternative Estimators of Seemingly Unrelated Regressions." Journal of the American Statistical Association, 63(1968), 1180-1200.
- Oberhofer, W. and J. Kmenta. "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models." Econometrica, 42(1974), 579-590.
- Pindyck, Robert S. and Daniel L. Rubinfeld. Econometric Models and Economic Forecasts. (New York: McGraw-Hill, 1981.)
- Swamy, P.A.V.B. and J.S. Mehta. "Estimation of Common Coefficients in Two Regression Equations." Journal of Econometrics, 10(1979), 1-14.
- Taylor, W.E. "Small Sample Properties of a Class of Two Stage Aitken Estimators." Econometrica, 45(1977), 497-508.
- Taylor, W.E. "The Heteroskedastic Linear Model: Exact Finite Sample Results." Econometrica, 46(1978), 663-675.

Footnotes

1. However, if size differences primarily affect means, X1 is appropriate. It is important to note that some experiments were tried with normal and log normal data. Results for these were substantially the same as those with uniform data.
2. Such heteroskedasticity can certainly occur. As Taylor notes, "In cross-section studies involving aggregates of vastly different sizes (states or countries), it is not unreasonable to expect [the degree of heteroskedasticity] to be on the order of 50." (1977, p. 504)
3. A contributing factor to the bias of WLS2 is the denominator in (3). Since $\sum n_i$ observations were used to estimate k parameters, $n_i - k$ understates the degrees of freedom. This is clear in the two group case with $\sigma_1^2 = \sigma_2^2$ and $n_1 = n_2$. Then an appropriate divisor for an unbiased estimate would be $1/2(n_1 + n_2 - k) = n_1 - \frac{k}{2}$. However, inappropriate degrees of freedom is not the only source of bias. Consider, for example, the unrealistic case of $\sigma_1^2 = 0$. The estimate of σ_1^2 from (3) would necessarily be non-zero regardless of the denominator.

Although the entire residual vector used in WLS2 has a sample mean of zero, the subvectors used to estimate the group variances do not. This prompted us to investigate a version of WLS2 in which the partitioned error sum of squares in (3) was corrected for its sample mean. This reduces the estimated error variances and hence their bias. However, the performance of this estimator was generally inferior to the version reported in the text.

4. With X1 data, the overestimation of small error variances was more pronounced when they were outnumbered by large error variances. This was less true with X2 data, which explains some of the specific patterns in the results. Notice that several of the experiments in the table can proxy for the case of two groups each with different sample sizes.