Econometrica, Vol. 51, No. 3 (May, 1983)

MORE EFFICIENT ESTIMATION IN THE PRESENCE OF HETEROSCEDASTICITY OF UNKNOWN FORM

BY J. G. CRAGG[1]

Recent work on inference in regressions with heteroscedastic disturbances leads to the possibility of more efficient estimators. The conditions needed to exploit the possibility are discussed. A sampling experiment indicates that the gains can be achieved even in small samples.

1. INTRODUCTION

CONCERN ABOUT HETEROSCEDASTICITY in linear regressions usually stems from its making least-squares estimates inefficient and rendering their estimated standard errors inconsistent. The effects need not be minor, as Geary **[3]** and Goldfeld and Quandt **[5]** noted. However, the problem of inconsistent standard errors can be overcome easily. Several authors, particularly Eicker **[1, 2]**, Hartley, Rao, and Kiefer **[7]**, Rao **[10]**, Hinkley **[9]** and White **[12]**, have pointed out that the covariance matrix of the least-squares estimator itself can be estimated consistently under very wide conditions. As a result, asymptotically appropriate inferences can be drawn using the least-squares estimates even when the form of the heteroscedasticity is unknown.

This paper shows that the same approach may be used to produce more efficient estimates when heteroscedasticity matters. Unlike the standard methods of adjusting for heteroscedasticity, such as those of Glejser **[4]**, Goldfeld and Quandt **[5]**, or Harvey **[8]**, the proposed estimator does not require specification of the exact form of the heteroscedasticity. The estimator is developed in Section 2 of the paper, while Section 3 examines the nature of the gain in efficiency and likely ways of exploiting the possibility.

The basic justification for the suggested estimator is asymptotic. Since the gain in efficiency smacks of getting something for nothing, one may be skeptical that efficiency gains really exist for practical-sized samples. A small sampling experiment, described in Section 4, revealed that efficiency gains may be realized. However, the method does have costs if used in small samples when it is not appropriate. This is a property shared with other methods. For example, Zellner **[14]** demonstrated it for his estimator of the seemingly unrelated regression (SUR) model. In addition, the proposed consistent covariance matrix, which is the White **[12]** matrix in the case of least-squares, shows significant biases that lead to serious problems of inference.

[1] Part of the work reported here was done while visiting the Institute of Economics and Statistics, Oxford, financed in part by the Social Sciences and Humanities Research Council of Canada. I have benefited from discussions with E. R. Berndt, C. L. Gilbert, and D. Stapleton. Two anonymous referees of an earlier version have been extremely helpful.

2. MORE EFFICIENT ESTIMATION

Consider the usual linear regression model

(1) $$y = X\beta + \epsilon,$$

where y is a $T \times 1$ vector of observed variables, X is a $T \times K$ matrix of exogenous variables with typical row x_t', and β is a $K \times 1$ vector of coefficients. The elements of ϵ, ϵ_t, are independent random variables for which

(2) $$E(\epsilon_t) = 0$$

and

(3) $$E(\epsilon_t^2) = \sigma_t^2,$$

so that

(4) $$E(\epsilon\epsilon') = \Sigma = \operatorname{diag}(\sigma_1^2, \dots, \sigma_T^2).$$

In addition, the existence of certain absolute moments of ϵ is assumed to ensure the applicability of the central limit theorem and the law of large numbers. To this end, it will be sufficient to assume that for some $\delta > 0$ there exists a $\Delta < \infty$ such that for all t

(5) $$E(|\epsilon_t|^{(2+\delta)}) < \Delta.$$

Suppose that, in addition to the x_t, there are available G auxiliary variables, p_t. The observations on these variables form a $T \times G$ matrix P. We let $Q = [X \;\; P]$ and $q_t' = (x_t', p_t')$. The exact nature of these variables will be considered in Section 3, but we may note that the elements of $x_t x_t'$—other than the ones in which the constant, if any, is involved—are likely candidates for inclusion in p_t. We shall treat the elements of q_t as being fixed variables. If they are considered random, the assumptions and analyses of White **[12]** may be extended readily to the present model.[2] Though stronger than generally needed (*cf.* Eicker **[2]**), we shall assume that the absolute values of σ_t^2 and of the absolute values of the elements of the q_t vectors are uniformly bounded, that the rank of Q is $(K + G)$, and that $\lim_{T\to\infty} Q'Q/T$ and $\lim_{T\to\infty} Q'\Sigma Q/T$ exist as well defined, finite, positive-definite matrices.

The assumptions allow inference to proceed using the ordinary least-squares estimator, at least asymptotically. Denote this estimator by

(6) $$\hat{\beta}^L = (X'X)^{-1}X'y$$

for which the variance-covariance matrix is

(7) $$V(\hat{\beta}^L) = E(\hat{\beta}^L - \beta)(\hat{\beta}^L - \beta)' = (X'X)^{-1}X'\Sigma X(X'X)^{-1}.$$

[2] This includes the assumption $E(\epsilon_t \mid Q) = 0$, which with fixed Q is subsumed in (2).

As noted by Eicker [2] and more recently by White [12] in a more general context, the assumptions made imply that[3] $\sqrt{T}(\hat{\beta}^L - \beta) \overset{A}{\sim} N(0, A)$ with

$$A = \lim_{T\to\infty} (X'X/T)^{-1}(X'\Sigma X/T)(X'X/T)^{-1} \tag{8}$$
$$= \lim_{T\to\infty} \left[TV(\hat{\beta}^L) \right].$$

Furthermore, the residuals, $e_t = y_t - x_t'\hat{\beta}^L$, provide a means for estimating A consistently since, defining $S = \text{diag}(e_1^2, e_2^2, \ldots, e_T^2)$,

$$\text{plim}(X'SX/T) = \lim_{T\to\infty} X'\Sigma X/T. \tag{9}$$

These considerations then provide asymptotic justification for treating the matrix $\hat{V}(\hat{\beta}^L)$, defined as

$$\hat{V}(\hat{\beta}^L) = (X'X)^{-1}X'SX(X'X)^{-1}, \tag{10}$$

as being the variance-covariance matrix of a normally distributed $\hat{\beta}^L$.

This procedure for dealing with heteroscedasticity only offers advantage over the usual least-squares covariance-matrix estimator if the variation in σ^2 is related to that of x_t. Explicitly let $\bar{\sigma}^2 = \sum_{t=1}^T \sigma_t^2/T$, $\delta_t = \sigma_t^2 - \bar{\sigma}^2$, $\delta' = (\delta_1, \ldots, \delta_T)$, and $\Delta = \text{diag}(\delta_1, \ldots, \delta_T)$. Then (7) can be written as

$$E(\hat{\beta}^L - \beta)(\hat{\beta}^L - \beta)' = \bar{\sigma}^2(X'X)^{-1} + (X'X)^{-1}X'\Delta X(X'X)^{-1}. \tag{11}$$

Hence if the δ_t tend to be orthogonal to the elements of $x_t x_t'$, in the sense that

$$\lim_{T\to\infty} X'\Delta X/T = \lim_{T\to\infty} \sum_{t=1}^{T} x_t x_t' \delta_t / T = 0, \tag{12}$$

there would be no need to use the suggested procedure instead of the usual OLS one.[4] When condition (12) does not hold, it is reasonable, even within the assumptions made, to expect to be able to find estimators which are more efficient than least-squares (at least asymptotically). It may even be possible to find such an estimator when condition (12) holds but the ϵ_t are heteroscedastic.

Consider the estimator

$$\tilde{\beta}^A = \left[X'Q(Q'\Sigma Q)^{-1}Q'X \right]^{-1} X'Q\left[Q'\Sigma Q \right]^{-1} Q'y \tag{13}$$
$$= \beta + \left[X'Q(Q'\Sigma Q)^{-1}Q'X \right]^{-1} X'Q\left[Q'\Sigma Q \right]^{-1} Q'\epsilon$$

[3] $\overset{A}{\sim}$ indicates convergence in distribution.

[4] Since $\sum_{t=1}^T s_t^2/T$ is of form (9), it has probability limit $\bar{\sigma}^2$ while, when (12) holds, (8) reduces to $\bar{\sigma}^2(X'X/T)^{-1}$. When σ_t^2 varies, $\hat{\beta}$ is not efficient even when (12) holds in that the Aitken estimator, if it were feasible, would be more efficient. While (12) does not imply that $\lim_{T\to\infty}(X'\Sigma^{-1}X/T) = \lim_{T\to\infty}\sum_{t=1}^T(1/\sigma_t^2)/T \lim_{T\to\infty}(X'X/T)$, when this is true the usual relationship of the arithmetic to the harmonic mean shows that $(\lim_{T\to\infty}\sum_{t=1}^T(1/\sigma_t^2)/T))^{-1} < \bar{\sigma}^2$ when σ_t^2 varies over time.

for which

$$(14) \qquad V(\tilde{\beta}^A) = \left[X'Q(Q'\Sigma Q)^{-1}Q'X\right]^{-1}.$$

The assumed bounded nature of q_t and assumption (5) ensure (just as they do for $X'\epsilon/\sqrt{T}$ in deriving the results for $\hat{\beta}^L$) that the Liapounov Central Limit Theorem applies to $Q'\epsilon/\sqrt{T}$. Thus,

$$(15) \qquad \sqrt{T}(\tilde{\beta}^A - \beta) \overset{A}{\sim} N\left[0, \lim_{T\to\infty} TV(\tilde{\beta}^A)\right].$$

This estimator arises as the Aitken estimator of the (artificial) system

$$(16) \qquad Q'y = Q'X\beta + Q'\epsilon$$

and so is the best linear unbiased estimator of form $\tilde{\beta} = CQ'y$, where C can be considered a fixed matrix. Since $\hat{\beta}^L$ is itself of this form, $\tilde{\beta}^A$ is more efficient than $\hat{\beta}^L$, unless $\hat{\beta}^L = \tilde{\beta}^A$. The form corresponds to White's **[13]** instrumental-variable estimator, but Q includes X.

Formula (13) does not provide a feasible estimator since it involves the unknown matrix Σ. The feasible equivalent may be taken to be

$$(17) \qquad \hat{\beta}^A = \left[X'Q(Q'SQ)^{-1}Q'X\right]^{-1}X'Q(Q'SQ)^{-1}Q'y.$$

Since Q shares all the relevant properties of X, condition (9) carries over so that $\text{plim}(Q'SQ/T) = \lim_{T\to\infty} Q'\Sigma Q/T$. Hence

$$(18) \qquad \text{plim}\left[\sqrt{T}(\hat{\beta}^A - \beta) - \sqrt{T}(\tilde{\beta}^A - \beta)\right] = 0$$

so that (15) provides the asymptotic distribution of $\hat{\beta}^A$. The covariance matrix of the estimator can be estimated consistently since, letting $\hat{V}(\hat{\beta}^A) = (X'Q(Q'SQ)^{-1}Q'X)^{-1}$:

$$(19) \qquad \text{plim}\left[T\hat{V}(\hat{\beta}^A)\right] = \text{plim}\left(X'Q/T(Q'SQ/T)^{-1}Q'X/T\right)^{-1}$$
$$= \lim_{T\to\infty}\left[TV(\tilde{\beta}^A)\right].$$

Analysis of $\tilde{\beta}^A$ will therefore reveal the gains in asymptotic efficiency obtained by using $\hat{\beta}^A$ rather than $\hat{\beta}^L$.

3. FINDING AUXILIARY VARIABLES

Exploitation of the possible increase in efficiency offered by $\tilde{\beta}^A$ requires finding variables to include in p_t to make $\tilde{\beta}^A$ more efficient than $\hat{\beta}^L$. A necessary and sufficient condition for p_t to increase efficiency is that[5]

$$(X'\Sigma X)^{-1}X'\Sigma P \neq (X'X)^{-1}X'P. \tag{20}$$

Equation (20) states that the regressions of P on X must not have the same coefficients as those of $\Sigma^{1/2}P$ on $\Sigma^{1/2}X$. Condition (20) may also be written as

$$X'\Sigma\big(P - X(X'X)^{-1}X'P\big) = X'\Delta\big(P - X(X'X)^{-1}X'P\big) \neq 0 \tag{21}$$

which states that the residuals of the (multivariate) regression of P on X must not be orthogonal to ΔX.

Insight into the nature of the gain in efficiency and desirable characteristics of auxiliary variables is provided by observing that $\tilde{\beta}^A$ is an instrumental-variable estimator of a rather arcane sort. If Σ were known, the efficient, Aitken estimator would be obtained by regressing $\Sigma^{-1/2}y$ on $\Sigma^{-1/2}X$ to produce

$$\tilde{\beta}^E = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \tag{22}$$

for which, of course,

$$V(\tilde{\beta}^E) = (X'\Sigma^{-1}X)^{-1}. \tag{23}$$

Instead of this regression, we use the instrumental-variable estimator in which the instruments for $\Sigma^{-1/2}X$ are its values predicted by its regression on $\Sigma^{1/2}Q$. That is, the instruments are the values of $\Sigma^{1/2}Q(Q'\Sigma Q)^{-1}Q'X$ arising from the regressions

$$\Sigma^{-1/2}X = \Sigma^{1/2}Q\Pi + V. \tag{24}$$

The difference in the precision matrices of $\tilde{\beta}^E$ and $\tilde{\beta}^A$,

$$V(\tilde{\beta}^E)^{-1} - V(\tilde{\beta}^A)^{-1} = X'\Sigma^{-1}X - X'Q(Q'\Sigma Q)^{-1}Q'X, \tag{25}$$

is simply the unexplained generalized sum of squares of the regression (24). In other words, the precision of $\tilde{\beta}^A$ falls short of that of $\tilde{\beta}^E$ exactly by the extent to which $\Sigma^{1/2}Q$ fails to "explain" $\Sigma^{-1/2}X$ in (24).[6]

[5]Estimator $\tilde{\beta}^A$ will be more efficient than $\hat{\beta}^L$ if $V(\hat{\beta}^L)^{-1} \neq V(\tilde{\beta}^A)^{-1}$. Define $F = [P'\Sigma P - P'\Sigma X(X'\Sigma X)^{-1}X'\Sigma P]$, which is positive-definite by the assumed positive-definiteness of $Q'\Sigma Q$. The standard rules for inversion of partitioned matrices then yield

$$\begin{aligned} V(\tilde{\beta}^A)^{-1} - B(\hat{\beta}^L)^{-1} = {} & [X'X(X'\Sigma X)^{-1}X'\Sigma P - X'P]F^{-1} \\ & \times [P'\Sigma X(X'\Sigma X)^{-1}X'X - P'X]. \end{aligned}$$

Condition (20) follows immediately.

[6]Similarly, the improvement in precision of $\tilde{\beta}^A$ over $\hat{\beta}^L$ in footnote 5 is simply the increase in the generalized explained sum of squares obtained by including $\Sigma^{1/2}P$ along with $\Sigma^{1/2}X$ in the instrumental-variable regression (24).

This way of looking at matters immediately suggests what role the auxiliary variables play, and so may suggest ways of finding them. Rewriting (24) as

$$\Sigma^{1/2}(\Sigma^{-1}X) = \Sigma^{1/2}Q\Pi + V, \tag{26}$$

we can see that the procedure implicitly uses the weighted least-squares fit of $\Sigma^{-1}X$ on Q. We are therefore looking for variables that "help" this regression.

We saw in condition (12) that heteroscedasticity is apt to be particularly a problem to usual least-squares inference when it is related to the x_t. When heteroscedasticity is recognized as a problem, we might expect that elements of $\sigma_t^{-2}x_t$ can be treated as some functions of x_t, say $\sigma_t^{-2}x_t = g(x_t)$. This in turn might suggest using a polynomial approximation for the unknown functions:

$$g(x_t)' = g_0' + x_t'G_1 + z_t'G_2 + \dots . \tag{27}$$

Here z_t is the vector of the $J = K(K-1)/2$ distinct elements of $x_t x_t'$ that are not elements of x_t (assuming that x_t contains a constant) and the G_1 and G_2 are matrices of coefficients. There is, however, no need to stop with a quadratic approximation. The elements of $x_t z_t'$ and of $z_t z_t'$ (distinct from those already included), can also be expected to improve the fit of (26) and so the efficiency of $\tilde{\beta}^A$.

The approximation in (27) need not be limited to the powers of x_t. If it is known that σ_t^2 depends on other, observed variables then it would seem sensible to include in p_t these variables, their own cross-products and their cross-products with x_t. Such variables may even be found when δ_t is unrelated to the x_t so that no direct benefit is available from the use of z_t.

These suggestions indicate that the number of candidates for possible inclusion in P may be large. The size of this number, provided that it is fixed, does not matter for the asymptotic results and inclusion of any variable having any additional explanatory power in the instrumental-variable regression (24) will increase the asymptotic efficiency of $\tilde{\beta}^A$. However, in practical applications T is given and the number of auxiliary variables to be used may be at the discretion of the investigator. Certainly, the number of elements of q_t, which is $K+G$, cannot exceed T for then the needed rank condition is not met. If $K+G=T$, $\tilde{\beta}^A$ reduces identically to the Aitken estimator $\tilde{\beta}^E$ because, of course, equation (24) then fits perfectly. $\tilde{\beta}^A$, however, becomes $(X'S^{-1}X)^{-1}X'S^{-1}y$. This is a very different sort of estimator from $\tilde{\beta}^A$, involving weighted sums of inverses of individual random variables rather than inverses of weighted sums as does $\hat{\beta}^A$. The presumption justifying use of the estimator in practical situations must be that a reasonably small number of auxiliary variables goes a long way to accounting for the dependent variables in (24). Even so, one may well suspect that the potential gain may not be substantial in practical situations while the switch from the unknown Σ to the calculated S may not be innocuous.

4. A SAMPLING EXPERIMENT

A small sampling experiment was conducted to investigate whether the asymptotic gain in efficiency really can be obtained in practice. The experiment used a model specified in the spirit of the one used by Goldfeld and Quandt [**5**]. The equation investigated had only a constant and one explanatory variable:

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t . \tag{28}$$

The values of the scalars x_t were obtained by using independent log-normally distributed pseudo-random variables.[7] They were then held fixed for the different replications of the experiment. The ϵ_t were independent, normally distributed pseudo-random variables with means zero and variances:

$$\sigma_t^2 = \gamma_1 + \gamma_2 x_t + \gamma_3 x_t^2 . \tag{29}$$

In the initial experiment the sample size, T, was set at 25 while the parameters were $\beta' = \{1.0, 1.0\}$ and $\gamma' = \{.1, .2, .3\}$. This assured substantial heteroscedasticity.

The auxiliary variables were chosen as $p_{jt} = x_t^{(j+1)}$, $j = 1, \ldots, G$, and the estimator was investigated for $G = 1, \ldots, 4$. An experiment consisted of 1000 replications using independent samples from the specified model, holding the x_t fixed. Because a variety of different types of questions might be asked of the data, no attempt was made to use variance-reducing techniques, such as those discussed in Hammersley and Handscomb [**6**].

Two features of the estimators are worth noting. First, the definition of S and expressions (1) and (17) show that $(\hat{\beta}^A - \beta)$ is an odd function of ϵ. Hence, if the ϵ are symmetrically distributed about zero, so is $(\hat{\beta}^A - \beta)$. In consequence, there is no point in being concerned about central tendency when a normal distribution is specified for ϵ_t. Second, multiplication of ϵ by a scalar simply multiplies $(\hat{\beta}^A - \beta)$ by the same scalar and changes all population and calculated variances by the square of the scalar. Therefore, we need not be concerned about the explanatory power of the model chosen. In other words, the absolute values of the γ parameters, but not their values relative to each other, are a matter of indifference and the level selected is essentially arbitrary. In consequence, we can legitimately present all results in relative terms. Similarly, neither S nor $(\hat{\beta}^A - \beta)$ actually depend on β, so these parameter values can be selected arbitrarily. By contrast, the specification of the x_t variables and of the form of heteroscedasticity are matters that do affect the results.

The findings of the initial experiment about efficiency of the estimators are summarized in Table I. The specification resulted in there being a substantial

[7] That is, they were $\exp(\eta_t)$ where η_t was an $N(0, 1)$ variable. All pseudo-random variables were generated using the UBC Computing Center program RANDOM [**11**], which employs Marsaglia's Rectangle-Wedge-Tail Method, applied to the output of a linear congruent method, to produce REAL*4 normal pseudo-random numbers. All subsequent computations were performed in double precision using the UBC Amdahl 470-V/8 computer.

TABLE I

EFFICIENCY GAINS—INITIAL EXPERIMENT

Variances of Estimates and Estimates of Variances as Proportions of Least-Squares Variances.
(Standard Deviations in Parentheses)

	β_1 Asymptotic	β_1 Actual	β_1 Estimated	β_2 Asymptotic	β_2 Actual	β_2 Estimated
Least Squares	1.000	1.011	.764	1.000	.980	.701
		(.047)	(.016)		(.047)	(.020)
$G=1$	.409	.478	.400	.590	.742	.442
		(.021)	(.005)		(.033)	(.009)
$=2$	.278	.337	.286	.471	.629	.309
		(.016)	(.004)		(.029)	(.006)
$=3$	.254	.331	.266	.445	.626	.270
		(.016)	(.004)		(.029)	(.005)
$=4$	.247	.346	.247	.437	.661	.230
		(.018)	(.004)		(.031)	(.004)
Aitken	.244	—	—	.432	—	—

difference between least-squares and the (true) Aitken estimator. The model is also such that large gains in efficiency would arise from using $\tilde{\beta}^A$, if it were feasible. The variances of these estimators are shown in the columns of Table I giving the asymptotic variances, in the sense of the diagonal elements of $(X'Q(Q'\Sigma Q)^{-1}Q'X)^{-1}$, relative to those of $(X'X)^{-1}X'\Sigma X(X'X)^{-1}$.

The feasible estimators, $\hat{\beta}^A$, were found to be substantially more efficient than least-squares. The columns of Table I labelled "actual" record the average squared errors, relative to the population variances of least-squares, that is, $(\hat{\beta}_i^A - \beta_i)^2/V(\hat{\beta}^L)_{ii}$. The indicated differences from least-squares are all highly significant. Not surprisingly, however, the gains in efficiency over least-squares are significantly less than those of $\tilde{\beta}^A$. Furthermore, increasing G after a small number appears not to produce further gains in efficiency. The main import of the findings is there are feasible efficiency gains even in very small samples when heteroscedasticity is pronounced.

Success in estimating the covariance matrices using $\hat{V}(\hat{\beta})$ was not very great. The "estimated" column of Table I shows the average of these estimates. They are all, including those for least-squares, biased downwards. The least-squares bias arises from $X'SX$ being a biased estimator of $X'\Sigma X$. It was, however, less biased than the diagonal elements[8] of $\{\sum_{t=1}^T s_t^2/(T-2)\}(X'X)^{-1}$. For $\hat{\beta}^A$, $\hat{V}(\hat{\beta}^A)$ was not always significantly smaller than $V(\tilde{\beta}^A)$, and relative to $V(\tilde{\beta}^A)$ the downard biases were not always as large as those of $\hat{V}(\hat{\beta}^L)$.

The finding that the gain in efficiency using $\hat{\beta}^A$ in small samples is less than the increase that would be obtained with $\tilde{\beta}^A$ makes one suspect that achieving any gain at all is dependent on there being substantial heteroscedasticity. This

[8] $\sum_{t=1}^T s_t^2/(T-2)$ is, with heteroscedasticity, a biased estimator of $\sum_{t=1}^T \sigma_t^2/T$, in this case a downward biased estimator.

TABLE II

EFFECTS OF DIFFERENT LEVELS OF HETEROSCEDASTICITY

Differences of Variances from Those of Least Squares as Per Cent of Least-Squares Population Variances.
(Standard Deviations in Parentheses)

Heteroscedasticity Level	β_1 0	1	2	3	4
$G = 1$	7.3	− 2.5	− 27.2	− 53.3	− 60.7
	(1.8)	(2.0)	(3.0)	(3.9)	(4.1)
= 2	8.3	− 3.1	− 31.9	− 67.4	− 78.2
	(2.0)	(2.2)	(3.1)	(3.9)	(4.1)
= 3	11.5	− 0.3	− 30.1	− 68.0	− 80.2
	(2.3)	(2.4)	(3.1)	(3.9)	(4.1)
= 4	13.9	2.0	− 27.6	− 66.5	− 79.6
	(2.4)	(2.6)	(3.2)	(3.8)	(4.0)
Aitken	0.0	− 9.6	− 38.3	− 75.6	− 90.2

Heteroscedasticity Level	β_2 0	1	2	3	4
$G = 1$	9.5	3.7	− 13.2	− 23.8	− 26.1
	(2.2)	(2.3)	(3.0)	(3.3)	(3.4)
= 2	13.1	4.9	− 16.7	− 35.1	− 40.0
	(2.5)	(2.7)	(3.1)	(3.3)	(3.4)
= 3	14.7	6.1	− 15.5	− 35.4	− 41.2
	(2.6)	(2.7)	(3.0)	(3.3)	(3.3)
= 4	15.6	7.7	− 12.5	− 31.8	− 38.0
	(2.7)	(2.8)	(2.9)	(3.1)	(3.1)
Aitken	0.0	− 9.9	− 34.6	− 56.8	− 66.1

suspicion is confirmed in experiments using different levels of heteroscedasticity. Five levels of increasing heteroscedasticity were defined by the γ' vectors $\{2.0, 0., 0.\}$, $\{.6, .3, 0.\}$, $\{.3, .2, .1\}$, $\{.1, .2, .3\}$, and $\{0., .2, .4\}$. They are referred to as levels 0 through 4 respectively. Level 3 is the one used in the initial experiment. As noted, only the relative values of the γ parameters matter. The experiment used the same values of x_t as the initial experiment and the same underlying random variables were used to generate the ϵ_t.

The results of varying the degree of heteroscedasticity on the efficiency of the estimator are shown in Table II, which records in percentage terms averages of $[(\hat{\beta}_i^A - \beta_i)^2 - (\hat{\beta}_i^L - \beta_i)^2]/V(\hat{\beta}_L^L)_{ii}$, that is the average actual proportional difference in square errors of the auxiliary-variable estimator and least-squares. With no heteroscedasticity, the auxiliary-variable estimator is significantly less efficient than least-squares.[9] With heteroscedasticity of level 1, significant differences were not found for β_1 while the auxiliary-variable estimator of β_2 was significantly weaker only for the larger values of G. With more heteroscedasticity, the auxiliary-variable estimators became steadily more efficient.

Varying the level of heteroscedasticity had little effect on the qualitative

[9] The significance arises from the greater power obtained by comparing each auxiliary-variable estimate with the corresponding least-squares estimate. For neither heteroscedasticity levels 0 or 1 are the actual relative variances significantly greater than unity.

performance of the estimated variances. They remained highly significantly biased downward relative to the actually observed values. While there was substantial variation among the various experiments, no pattern was evident.

Heteroscedasticity level 0 actually represents homoscedasticity, and so offers a good opportunity to investigate the White [**12**] test for heteroscedasticity. This test conducted at the .05 level rejected the null hypothesis of homoscedasticity 4.9 per cent of the time. As heteroscedasticity increased, so did the frequency of rejection of the null hypothesis, the frequencies being 17.9 per cent, 56.2 per cent, 89.7 per cent and 94.8 per cent as the level of heteroscedasticity went from 1 to 4.

The problems of $\hat{\beta}^A$ being significantly less efficient than $\tilde{\beta}^A$ and of all estimated variances being biased downward are features of the finite-sample rather than the asymptotic distributions. It is of some interest to see how rapidly they may disappear as the sample size increases. Experiments were therefore conducted with 100 and 1000 observations. The sample size was expanded by drawing further values from the distribution originally used. This altered the values of expressions such as $Q'\Sigma Q/T$ and the results may reflect these differences as well as changes in sample size.

The results of altering the sample size are shown in Table III. The actual variances did tend to get closer to the asymptotic values. With $T = 1000$, the

TABLE III

DIFFERENT SAMPLE SIZES:
RATIOS OF AVERAGE ACTUAL AND ESTIMATED VARIANCES TO ASYMPTOTIC
(Standard Deviations in Parentheses)

1. Actual

	β_1			β_2		
$T=$	25	100	1000	25	100	1000
$G = 1$	1.17	1.12	1.00	1.26	1.17	1.00
	(.05)	(.05)	(.05)	(.06)	(.05)	(.04)
$= 2$	1.22	1.15	.97	1.34	1.21	1.03
	(.06)	(.05)	(.04)	(.06)	(.06)	(.05)
$= 3$	1.30	1.24	1.06	1.41	1.32	1.12
	(.06)	(.06)	(.05)	(.07)	(.06)	(.05)
$= 4$	1.40	1.28	1.20	1.51	1.37	1.26
	(.07)	(.06)	(.05)	(.07)	(.06)	(.06)

2. Estimated

	β_1			β_2		
$T=$	25	100	1000	25	100	1000
Least Squares	.764	.793	.855	.701	.803	.866
	(.018)	(.019)	(.017)	(.021)	(.017)	(.015)
$G = 1$	.977	.951	.945	.749	.890	.953
	(.013)	(.008)	(.008)	(.015)	(.010)	(.007)
$= 2$	1.032	.868	.881	.656	.736	.889
	(.016)	(.008)	(.005)	(.012)	(.010)	(.006)
$= 3$	1.047	.897	.942	.606	.705	.918
	(.017)	(.008)	(.004)	(.011)	(.010)	(.005)
$= 4$	.999	.909	.937	.527	.689	.886
	(.017)	(.008)	(.004)	(.010)	(.010)	(.006)

differences were not significant for lower values of G. Improvements, but not as complete ones, were found for the estimated variances. Significant downward biases of $\hat{V}(\hat{\beta})_{ii}$ as estimates of actual variances continued to be found as T increased. The most dramatic changes occurred for the auxiliary-variable estimators of β_1: significant downward bias relative to the asymptotic value appeared when T changed from 25 to 100. Even by $T = 1000$, the biases had not disappeared for either coefficient.[10]

The downward bias of the estimated variances can have serious effects on the *de facto* sizes of standard tests for the values of the coefficients. This was indeed the case. When $(\hat{\beta}_i - \beta_i)/\hat{V}(\hat{\beta})_{ii}^{1/2}$ was treated as $N(0,1)$ to test the (true) $H_0: \beta_i = 1.0$ at the .05 level, it was found[11] that the frequency of rejection was significantly greater than five per cent except when $G = 1$ with $T = 1000$. The weakness of the tests affects least-squares as well as the auxiliary-variable estimators. Indeed, with considerable heteroscedasticity, the problem of incorrect size was less important for the auxiliary-variable estimator with low values of G than for least-squares when coefficient β_1 was under consideration. Thus, even for testing and only considering size, the auxiliary-variable estimator may be preferable to least-squares. Since the standard errors of the auxiliary-variable estimator are smaller than those of least-squares, considerations of power then indicate clear advantage to using the auxiliary-variable estimator.

The problems arising in least-squares inference from using $\hat{V}(\hat{\beta}^L)$ are only partly due to its being a biased estimator. This bias can be corrected using the MINQUE estimator developed by Rao [**10**]. Using the MINQUE estimates did produce unbiased estimates of the variances of $\hat{\beta}^L$, but not[12] of $\hat{\beta}^A$. Even with unbiased estimates, the random nature of the denominator suggests that the normal distribution would not be appropriate and, on an *ad hoc* basis, one might expect $t(T-2)$ to serve better. Making these two changes did improve the performance of the tests. They did not cure the problems entirely. In many cases the frequency with which the null hypothesis was rejected still significantly exceeded the five per cent which was the nominal significance level used. Clearly a better approximation to the distributions is needed.

5. SUMMARY AND CONCLUSION

This paper has argued that more efficient estimates than least-squares are likely to be available when heteroscedasticity of unknown form is a problem.

[10] The failure of the standard deviation to decline in inverse proportion to the square roots of the increases in sample size probably reflects the changing values of x_t that accompany alterations in T. For β_1, the ratios of the Aitken variance to the least-squares were .24, .20, and .04 for $T = 25$, 100, and 1000 respectively while for β_2 the figures were .43, .35, and .10.

[11] More details about these results are available in the UBC Discussion Paper (81-43) version of this paper, under the same title.

[12] The covariance matrices of $\hat{\beta}^A$ were changed to $(X'Q(Q'\tilde{S}Q)^{-1}_{\sim}Q'X)^{-1}$ for this investigation, where $\tilde{S}$ uses the MINQUE estimates. $\hat{\beta}^A$ was not redefined to use $\tilde{S}$ in its calculation. Trials of this latter change suggested that it did not improve the estimator.

Auxiliary variables can be used to increase efficiency. The cross-products and higher products of the independent variables can be expected to provide the needed auxiliary variables.

A small sampling experiment indicated that the efficiency gains can be realized even in very small samples, but only if there is more than a small amount of heteroscedasticity. The addition of further auxiliary variables encountered diminishing returns at a faster rate than the asymptotic theory suggested.

The major problem encountered was downward bias of the covariance matrices. This appeared to be as serious for White's heteroscedasticity-consistent estimates for least-squares as for the auxiliary-variable estimator. The difficulties produced by this bias disappeared very slowly as sample size increased and were still evident when 1000 observations were used. The results do not indicate that least-squares has an edge. Thus when adjustment for heteroscedasticity of the least-squares covariance matrix seems desirable, use of the auxiliary-variable estimator may well be indicated.

The approach taken to obtain more efficient estimates extends easily to other models. When instrumental-variable estimation is appropriate, the auxiliary-variable formula (17) may still be used, redefining Q to contain the instruments and auxiliary variables rather than the independent variables. When systems of equations are involved, the auxiliary variables may include exogenous variables left out of each equation. Furthermore, noting that the Eicker–White approach extends immediately to the covariance matrices of systems of equations, the auxiliary-variable estimator (17) also extends in straightforward fashion to provide the equivalent of Zellner's SUR estimator and three-stage least-squares. It remains an open question whether with heteroscedasticity these estimators, like the simpler ones considered in this paper, are more satisfactory than the usual estimators with appropriately estimated variance-covariance matrices.

University of British Columbia

Manuscript received May, 1981; revision received June, 1982.

REFERENCES

[1] Eicker, F.: "Asymptotic Normality and Consistency of the Least Squares Estimator for Families of Linear Regressions," *Annals of Mathematical Statistics*, 34(1963), 447–456.

[2] ———: "Limit Theorems for Regressions with Unequal and Dependent Errors," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. by L. LeCam and J. Neyman. Berkeley: University of California Press, 1967, pp. 59–82.

[3] Geary, R. C.: "A Note on Residual Heterovariance and Estimation Efficiency in Regression," *American Statistician*, 20(1966), 30–31.

[4] Glejser, H.: "A New Test for Heteroskedasticity," *Journal of the American Statistical Association*," 64(1969), 316–323.

[5] Goldfeld, S. M., and R. E. Quandt: *Nonlinear Methods in Econometrics*. Amsterdam: North-Holland, 1972, Chapter 3.

[6] Hammersley, J. M., and D. C. Handscomb: *Monte Carlo Methods*. London: Methuen, 1964.

[7] Hartley, H. O., J. N. K. Rao, and G. Keifer: "Variance Estimation With One Unit Per Stratum," *Journal of the American Statistical Association*, 64(1969), 841–851.

[8] HARVEY, A. C.: "Estimating Regression Models with Multiplicative Heteroscedasticity," *Econometrica*, 44(1976), 461–466.

[9] HINKLEY, DAVID, V.: "Jackknifing in Unbalanced Situations," *Technometrics*, 19(1977), 285–292.

[10] RAO, C. R.: "Estimation of Heteroscedastic Variances in Linear Models," *Journal of the American Statistical Association*, 65(1970), 161–172.

[11] UNIVERSITY OF BRITISH COLUMBIA COMPUTING CENTER: "UBC Random," mimeo., 1978.

[12] WHITE, HALBERT: "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(1980), 817–838.

[13] ———: "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50(1982), 483–499.

[14] ZELLNER, A.: "Estimators of Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results," *Journal of the American Statistical Association*, 58(1963), 977–992.