FACULTY WORKING PAPER 93-0105

On Some Heteroskedasticity-Robust Estimators of Variance-Covariance Matrix

Anil K. Bera
Department of Economics
University of Illinois

Totok Suprayitno
Department of Economics
University of Illinois

Bureau of Economic and Business Research
College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 93-0105

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

January 1993

On Some Heteroskedasticity-Robust Estimators of Variance-Covariance Matrix

Anil K. Bera
Totok Suprayitno

Department of Economics

This version: January 1993

ON SOME HETEROSKEDASTICITY-ROBUST ESTIMATORS OF VARIANCE-COVARIANCE MATRIX

Anil K. Bera and Totok Suprayitno*

University of Illinois at Urbana-Champaign

ABSTRACT

Chesher and Jewitt (1987) demonstrated that White's (1980) consistent estimator of the variance-covariance matrix in heteroskedastic models could be severely *biased* if the design matrix is highly unbalanced. In this paper we, therefore, reconsider the Rao (1970) minimum norm quadratic *unbiased* estimator (MINQUE). We derive the analytical expressions for the mean square errors (MSE) of White's (1980), one of MacKinnon and White's (1985) and MINQU estimators, and perform a numerical comparison. Our analysis shows that although MINQUE is unbiased by construction, it has very large variance particularly for the highly unbalanced design matrices. Since the variance is the dominant factor in our MSE computation, MINQUE is not the preferred estimator in terms of MSE comparison. We also studied the finite sample behavior of the confidence interval of regression coefficients in terms of coverage probabilities based on different variance-covariance matrix estimators. Our results indicate that although MINQUE generally has the largest MSE, it performs relatively well in terms of coverage probabilities. Overall, taking both MSE and coverage probabilities as choice criteria, the 'almost unbiased' estimator suggested in MacKinnon and White (1985) is the winner.

1. Introduction

When the disturbance process in a regression model exhibits heteroskedasticity, the invalidity of standard inference procedures stems from the wrong estimation of the standard errors. A conventional way of overcoming this problem in econometric modeling is to specify the model under an assumed error structure, and apply Aitken's weighted least squares. This method does not seem to be attractive to many practitioners as usually there is very little or no guidance regarding the form of heteroskedasticity. White (1980) proposed an estimator of variance-covariance matrix of the least squares regression coefficients which, under certain conditions, is consistent under heteroskedasticity. Other attractive features of this estimator are that it is obtained without specifying the structural form of heteroskedasticity, and it is very easy to compute. This may explain the reason behind its popularity in applied econometric work.

* We would like to thank Roger Koenker, Paul Newbold for helpful comments, and Sungsup Ra for his research assistance. Financial support from the Research Board of the University of Illinois is gratefully acknowledged.

Recently, several researchers criticized the widespread acceptance of White's procedure, which some people call it the "White washing". Chesher and Jewitt (1987) showed analytically that for certain regression designs the estimator exhibits a large bias even in large sample. In particular, a severe bias arises when there is a large value of point leverage of the regression design, rendering inferences drawn from this estimator uninformative. A Monte Carlo study conducted by Mishkin (1990) also indicated that the use of White's standard errors cannot always correct the inferences, and in some situations can make things even worse.

Alternatives to the heteroskedasticity-consistent variance estimator are available. A close variant of White's estimator is the one suggested by MacKinnon and White (1985). They considered an estimator based on the unreplicated "almost unbiased estimator" of Horn, Horn and Duncan (1975). This estimator is unbiased when there is no heteroskedasticity, but is biased if the homoskedastic assumption is not satisfied. In the special case of balanced regression designs, it reduces to the estimator considered by Hinkley (1977), which differs from White's estimator only by some proportional constant. Other alternatives include those based on minimum norm quadratic estimation (MINQUE) principle of Rao (1970), resampling method of Wu (1986) and maximum likelihood estimation of Hartley and Jayatillake (1973). Some of the extensions to a more general case where the disturbances are also serially correlated are provided by Newey and West (1987), Wooldridge (1989) and Andrews (1991).

In this paper we reconsider the MINQUE principle to obtain an unbiased estimator for variance-covariance matrix under heteroskedasticity. The paper proceeds as follows. In Section 2 we provide some review of White's consistent estimator, highlighting its bias and indicating a simple way to eliminate the bias. In Section 3 we discuss the MINQUE procedure within the framework of a variance component model. The exact expressions for the finite sample variance of different estimators are derived in Section 4, and some numerical and Monte Carlo results are given in Section 5. The last section provides a conclusion.

2. The Bias of White's Heteroskedasticity-Consistent Estimator

We consider the standard regression model

$$\mathbf{y} = X\beta + \varepsilon, \tag{1}$$

where $\mathbf{y}$ is a $(n \times 1)$ vector of dependent variables, X is a $(n \times k)$ matrix of independent variables, β is a $(k \times 1)$ vector of unknown parameters, and ε is a $(n \times 1)$ random vector with mean zero and variance-covariance matrix $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. Under this setup, the ordinary least squares (OLS) estimator of β is given by

$$\hat{\beta} = (X'X)^{-1}X'\mathbf{y}, \tag{2}$$

and its true variance-covariance matrix is

$$\Omega_T = (X'X)^{-1}X'\Sigma X(X'X)^{-1}. \tag{3}$$

Under homoskedasticity assumption $\varepsilon \sim (0, \sigma^2 I)$ this variance-covariance matrix is estimated by

$$\Omega_S = \hat{\sigma}^2(X'X)^{-1}, \tag{4}$$

where $\hat{\sigma}^2$ is the standard OLS estimator of σ^2. This latter estimator is inconsistent if in fact the disturbances are heteroskedastic.

White's (1980) heteroskedastic-consistent estimator is given by [see also Eicker (1963)]

$$\Omega_W = (X'X)^{-1}X'\hat{\Sigma}X(X'X)^{-1} \tag{5}$$

where $\hat{\Sigma} = \text{diag}(\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_n^2)$, with $\hat{\varepsilon}_i$ being the OLS residual. Note that this $\hat{\Sigma}$ is similar to the unreplicated J.N.K Rao's (1972) modified MINQUE or the unreplicated average of squared residuals of Horn, Horn and Duncan (1975). Different from the traditional ways of overcoming the heteroskedasticity problem in econometrics literatures, the Ω_W does not require specification of the particular form of heteroskedasticity.

Under the regularity conditions given in White (1980), Ω_W is a consistent estimator for Ω_T, but it is generally biased under *both homoskedastic and heteroskedastic* disturbances. Following Chesher and Jewitt (1987), let us define $H = X(X'X)^{-1}X'$, $M = I - H$, $\mathbf{h}_i$ is the i-th column of matrix H, $\mathbf{m}_i$ is the i-th column of matrix M, and h_{ij} is the (i,j)-th element of matrix H. We have $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)' = M\varepsilon$ and therefore,

$$\begin{aligned} E(\hat{\varepsilon}_i^2) = \mathbf{m}_i'\Sigma\mathbf{m}_i &= \sigma_i^2 - 2h_{ii}\sigma_i^2 + \sum_{j=1}^{n} h_{ij}^2\sigma_j^2 \\ &= \sigma_i^2 - 2\mathbf{h}_i'\mathbf{h}_i\sigma_i^2 + \mathbf{h}_i'\Sigma\mathbf{h}_i, \end{aligned} \tag{6}$$

since H is an idempotent matrix. The bias of $\hat{\varepsilon}_i^2$ then is given by

$$\begin{aligned} \text{bias}(\hat{\varepsilon}_i^2) &= E(\hat{\varepsilon}_i^2) - \sigma_i^2 \\ &= \mathbf{h}_i'\Sigma\mathbf{h}_i - 2\mathbf{h}_i'\mathbf{h}_i\sigma_i^2, \\ &= \mathbf{h}_i'(\Sigma - 2\sigma_i^2 I)\mathbf{h}_i, \end{aligned} \tag{7}$$

and the bias of White's consistent estimator Ω_W is

$$\text{bias}(\Omega_W) = (X'X)^{-1}X'BX(X'X)^{-1}, \tag{8}$$

where $B = \text{diag}\left(\mathbf{h}_1'(\Sigma - 2\sigma_1^2 I)\mathbf{h}_1, \ldots, \mathbf{h}_n'(\Sigma - 2\sigma_n^2 I)\mathbf{h}_n\right)$, which is not zero under both homoskedastic and heteroskedastic disturbances. Obviously, when $\max_i(\sigma_i^2) < 2\min_i(\sigma_i^2)$ for all i, all elements of B are negative, and therefore the standard error of all elements of $\hat{\beta}$ would be underestimated.

When the disturbances are homoskedastic, the bias of the White's consistent estimator will be $-\sigma^2(X'X)^{-1}X'\left[\text{diag}\left(h_{11}, \ldots, h_{nn}\right)\right]X(X'X)^{-1}$. Horn, Horn and Duncan (1975) proposed $\hat{\varepsilon_i}^2/(1-h_{ii})$ as an almost unbiased estimator (AUE) for σ_i^2, which was then used by MacKinnon and White (1985) to modify the White's consistent estimator. MacKinnon and White's (1985) estimator can be written as

$$\Omega_{MW} = (X'X)^{-1}X'\tilde{\Sigma}X(X'X)^{-1}, \tag{9}$$

where $\tilde{\Sigma} = \text{diag}\left(\hat{\varepsilon}_1^2/(1-h_{11}), \ldots, \hat{\varepsilon}_n^2/(1-h_{nn})\right)$. This estimator is of course unbiased only when the disturbances are homoskedastic. In the special case of a balanced design matrix X, where $h_{ii} = k/n$ for all i, Ω_{MW} reduces to $(n/(n-k))\Omega_W$, which is the variance-covariance matrix estimator suggested by Hinkley (1977). Both MacKinnon and White's (1985) and Hinkley's (1977), however, are biased when the disturbances are heteroskedastic.

Given the relation between ε and $\hat{\varepsilon}$, the derivation of unbiased version of the White's estimator for general design matrix X is straightforward. Let us denote by m_{ij} the (i,j)-th element of matrix M. Then, from (6), we have

$$E(\hat{\varepsilon}_i^2) = \sum_{j=1}^{n} m_{ij}^2 \sigma_j^2 \qquad \text{for } i = 1, \ldots, n, \tag{10}$$

which can be expressed as

$$E\begin{bmatrix} \hat{\varepsilon}_1^2 \\ \hat{\varepsilon}_2^2 \\ \vdots \\ \hat{\varepsilon}_n^2 \end{bmatrix} = \begin{bmatrix} m_{11}^2 & m_{12}^2 & \ldots & m_{1n}^2 \\ m_{21}^2 & m_{22}^2 & \ldots & m_{2n}^2 \\ \ldots & \ldots & \ldots & \ldots \\ m_{n1}^2 & m_{n2}^2 & \ldots & m_{nn}^2 \end{bmatrix} \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \vdots \\ \sigma_n^2 \end{bmatrix} \tag{11}$$

$$E(\ddot{\mathrm{e}}) = Q\sigma^{(2)}, \quad \text{say.}$$

Therefore, $\hat{\sigma}^{(2)} = (\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_n^2)' = Q^{-1}\ddot{\mathbf{e}}$ is an unbiased estimator of $\sigma^{(2)}$ if Q is non-singular, and our unbiased version of the White's estimator can be obtained by putting the i-th element of $\hat{\sigma}^{(2)}$ instead of $\hat{\varepsilon}_i^2$ in the expression for Ω_W. It is interesting to see that $\hat{\sigma}_i^2$ turns out to be exactly MINQUE of σ_i^2 proposed by Rao (1970), as demonstrated in the following section.

3. MINQUE of Variance-Covariance Matrix in Heteroskedastic Linear Models

We write the disturbance process ε of (1) in an identity similar to the variance components type model as follows,

$$\varepsilon = \mathbf{u}_1\varepsilon_1 + \cdots + \mathbf{u}_n\,\varepsilon_n, \tag{12}$$

where $\mathbf{u}_i \quad (i = 1, \ldots, n)$ is a $(n \times 1)$ known vector whose i-th element is one and the rest are zero. Since $\text{var}(\varepsilon_i) = \sigma_i^2 \; (i = 1, \ldots, n)$, the variance-covariance matrix of ε is given by $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ which is exactly the heteroskedastic problem we usually consider. Rao (1970) approached precisely this problem in a somewhat different way, and obtained MINQUE of σ_i^2. We will see that variance component representation of the disturbance process above arrives at the same result and provides a more convenient way of obtaining MINQUE of heteroskedastic variances. This variance component framework is also useful for analyzing various forms of heteroskedasticity.

Turning back to our present problem, our interest is to obtain the MINQUE of $k \times k$ variance-covariance matrix $\Omega_T = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$. Denoting by $\mathcal{X}_{ij}$ the (i,j)-th element of $(X'X)^{-1}X'$ for $i = 1, \ldots, k$ and $j = 1, \ldots, n$, the (r,s)-th element of Ω_T may be written as a linear combination of $\sigma_j^2 \; (j = 1, \ldots, n)$

$$\Omega_T(r,s) = \sum_{j=1}^{n} \mathcal{X}_{rj}\mathcal{X}_{sj}\sigma_j^2, \qquad r, s = 1, \ldots, k$$

$$= \mathbf{a}_{rs}'\sigma^{(2)} \tag{13}$$

where $\mathbf{a}_{rs}' = (\mathcal{X}_{r1}\mathcal{X}_{s1}, \ldots, \mathcal{X}_{rn}\mathcal{X}_{sn})$.

It is easy to see that the MINQUE of $\Omega_T(r,s)$ can be obtained directly from Rao (1972). Let us write $V_i = \mathbf{u}_i\mathbf{u}_i'$, then $\sum_{i=1}^{n} V_i = I$. The MINQUE of $\Omega_T(r,s)$ is given by $\mathbf{y}'A\mathbf{y}$, with A satisfying

$$\min_A \text{tr}\,(AA) \text{ subject to } AX = 0 \text{ and } \text{tr}\,(AV_i) = \mathcal{X}_{ri}\mathcal{X}_{si}, \quad i = 1, 2, \ldots, n, \tag{14}$$

where tr$(\cdot)$ denotes a trace of a matrix. The above two restrictions impose invariance and unbiasedness. The solution to (14) is given by

$$A^* = \sum_{i=1}^{n} \lambda_i M V_i M, \tag{15}$$

where $\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ satisfies

$$\lambda' Q = \mathbf{a}'_{rs}, \tag{16}$$

with $Q = [\![\mathrm{tr}(MV_iMV_j)]\!]$. Here we write $[\![a_{ij}]\!]$ to denote a matrix whose (i,j)-th element is a_{ij}. Some simple algebra shows that Q is the Hadamard product of matrix M; explicitly,

$$Q = M * M = \begin{bmatrix} m_{11}^2 & m_{12}^2 & \ldots & m_{1n}^2 \\ m_{21}^2 & m_{22}^2 & \ldots & m_{2n}^2 \\ \ldots & \ldots & \ldots & \ldots \\ m_{n1}^2 & m_{n2}^2 & \ldots & m_{nn}^2 \end{bmatrix}, \tag{17}$$

where m_{ij} is the (i,j)-th element of M.

Now, let us denote by $\Omega_M(r,s)$ the MINQUE of $\Omega_T(r,s) = \mathbf{a}'_{rs}\sigma^{(2)}$. It is given by

$$\Omega_M(r,s) = \mathbf{y}'A^*\mathbf{y} = \sum_{i=1}^{n} \lambda_i \mathbf{y}'MV_iM\mathbf{y} = \sum_{i=1}^{n} \lambda_i \hat{\varepsilon}'V_i\hat{\varepsilon} = \sum_{i=1}^{n} \lambda_i \hat{\varepsilon}_i^2 = \lambda'\ddot{\mathbf{e}}. \tag{18}$$

If Q is non-singular,

$$\Omega_M(r,s) = \mathbf{a}'_{rs}Q^{-1}\ddot{\mathbf{e}}. \tag{19}$$

A contrast between MINQUE and White's consistent estimator of Ω_T can be made in terms of their weighting scheme. Each MINQUE of σ_i^2 is estimated by a linear combination of squares of all the OLS residuals with the weights are related to the design matrix X. In White's approach, on the other hand, all the weights are given to the respective OLS squared residual, i.e, σ_i^2 is estimated by $\hat{\varepsilon}_i^2$ for all i. Clearly, White's estimator is always biased unless matrix $H = X(X'X)^{-1}X' = [\![h_{ij}]\!]$ is zero. Chesher and Jewitt (1987) specifically show that very severe bias may arise when there are large h_{ii}, and becomes extreme as $\max(h_{ii})$ approaches 1, rendering the inferences drawn from White's consistent estimator uninformative. In such situation the MINQUE may be useful as a point estimator.

The obvious problem encountered in using MINQUE is that matrix Q may be singular. An easily verifiable condition for nonsingularity of matrix Q is given by Horn and Horn (1975), namely,

$$\max_i(h_{ii}) < \frac{1}{2} \qquad \text{for} \quad i = 1, \ldots, n, \tag{20}$$

where $h_{ii} = 1 - m_{ii} = \mathbf{x}_i(X'X)^{-1}\mathbf{x}_i'$, which is often regarded as a measure of points of leverage of the regression design. To see the above result note that Q is nonsingular if it is a dominant diagonal matrix, i.e., if $m_{ii}^2 > \sum_{j\neq i}^n m_{ij}^2$ for all $i = 1, 2, \ldots, n$ [see, e.g., Graybill (1983) pp. 250-251]. Since M is idempotent, $\sum_{j=1}^n m_{ij}^2 = m_{ii}$, and the required condition becomes $m_{ii}(1 - 2m_{ii}) < 0$, which is the same as $3h_{ii} - 2h_{ii}^2 < 1$. Note that $3h_{ii} - 2h_{ii}^2$ is a monotonically increasing function for $0 \leq h_{ii} \leq \frac{3}{4}$, and it is less than 1 if $h_{ii} < \frac{1}{2}$. Another set of conditions for nonsingularity of matrix Q is also suggested by Rao (1970), but it is apparently not easy to verify for economic data where we usually have quite complicated regression design.

The condition (20) is sufficient but not necessary. For example, consider the following regression design

$$X' = \begin{bmatrix} 1 & 0 & 0 & 2 & 0 & 1 \\ 0 & 1 & 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 & 1 \end{bmatrix}.$$

The matrix $Q = (I - X(X'X)^{-1}X') * (I - X(X'X)^{-1}X')$ is nonsingular even though $\max_{i=1,\ldots,6} h_{ii} = 0.8$. A set of necessary and sufficient conditions is proposed by Mallela (1972) as follows. Let X_1 be a set of k linearly independent rows of X, and X_2 be the set of X complementary to X_1 in X. Define $Z = X_2X_1^{-1}$, and let $\mathbf{z}_i$ be the i-th row of Z. Further, let R be the $((n-k)(n-k-1)/2) \times k$ whose rows are the Hadamard product $\mathbf{z}_i * \mathbf{z}_j$ for $i < j$, $i, j = 1, 2, \ldots, n-k$. Then Q is nonsingular if and only if the rank of R is equal to k. This set of conditions, however, is neither simple nor easy to interpret and may be computationally burdensome.

Another drawback of MINQUE procedure is the possibility of getting some negative estimates of individual variances σ_i^2. A common suggestion when this problem arises is to apply an *ad hoc* procedure by replacing the non-positive values by a small positive number, resulting a truncated MINQUE

$$\tilde{\sigma}_i^2 = \begin{cases} \hat{\sigma}_i^2 & \text{if } \hat{\sigma}_i^2 > 0 \\ \delta_i & \text{if } \hat{\sigma}_i^2 \leq 0, \end{cases}$$

where δ_i's are some small numbers guaranteed to be greater than zero. This modification, of course, destroys the unbiasedness property of MINQUE and may not be theoretically justified.

4. The Finite Sample Variance of the Estimators

Given the seriousness of bias in White's estimator it is natural to ask how the unbiased estimator MINQUE will perform in terms of variance. It is expected that MINQUE will have higher variance. As a compromise, we will later use mean square error (MSE), which combines bias and variance, as a criterion to compare different estimators. From Section 2 biases of all variance estimators can be easily obtained. In this section we derive the variances of $\Omega_S(r,s)$, $\Omega_W(r,s)$, $\Omega_A(r,s)$ and $\Omega_M(r,s)$, respectively corresponding to the standard OLS, White's (1980) consistent, almost unbiased MacKinnon and White (1985) and MINQU estimators of the (r,s)-th element of the true OLS variance $\Omega_T(r,s)$. We assume the disturbance vector ε is normally distributed with mean zero and variance $\mathrm{diag}(\sigma_1^2,\ldots,\sigma_n^2)$. To simplify the derivation, we first note, as we mentioned earlier, that the algebra of those estimators differ from each other only in their weighting scheme of the squared OLS residuals. Specifically, all but the standard OLS variance estimator can be written as

$$\widehat{\Omega}(r,s) = \mathbf{a}_{rs}' W \ddot{\mathbf{e}}, \qquad r,s = 1,\ldots,k, \tag{21}$$

where W is a $(n \times n)$ weighting matrix. For the White (1980), MacKinnon and White (1985) and MINQU estimators, W corresponds to $I_{n\times n}$, $\mathrm{diag}(1/m_{11},\ldots,1/m_{nn})$ and Q^{-1}, respectively. The standard OLS variance estimator is the special case for which $\mathbf{a}_{rs}$ is the scalar corresponding to the (r,s)-th element of $(X'X)^{-1}$ and W is $(n\times 1)$ vector of $1/(n-k)$. Under the representation (21), the variance of $\widehat{\Omega}(r,s)$ is given by

$$\mathrm{var}(\widehat{\Omega}(r,s)) = \mathbf{a}_{rs}' W\, \mathcal{V}(\ddot{\mathbf{e}})\, W' \mathbf{a}_{rs}, \tag{22}$$

where $\mathcal{V}(\ddot{\mathbf{e}})$ denotes the variance-covariance matrix of $\ddot{\mathbf{e}}$. It can be shown that $V(\ddot{\mathbf{e}})$ has the following typical elements (see the Appendix for derivation):

$$\mathrm{var}(\hat{\varepsilon}_i^2) = 2\Big(\sum_{t=1}^{n} m_{it}^2 \sigma_t^2\Big)^2 \equiv 2\gamma_{ii}, \tag{23}$$

$$\mathrm{cov}(\hat{\varepsilon}_i^2, \hat{\varepsilon}_j^2) = 2\Big(\sum_{t=1}^{n} m_{it} m_{jt} \sigma_t^2\Big)^2 \equiv 2\gamma_{ij}. \tag{24}$$

Let us denote by q^{ij} the (i,j)-th element of matrix Q^{-1}. Then the variance of MINQUE $\Omega_M(r,s)$ is

$$\mathrm{var}(\Omega_M(r,s)) = 2\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{f=1}^{n}\sum_{g=1}^{n} \mathcal{X}_{ri}\mathcal{X}_{si}\,\mathcal{X}_{rj}\mathcal{X}_{sj}\, q^{if} q^{jg}\ \gamma_{fg} \tag{25}$$

for $r,s = 1,\ldots,k$. If $q^{ii} = 1/m_{ii}$ and $q^{ij} = 0$ for $i \neq j$, $i,j = 1,2,\ldots n$, then it reduces to the variance of $\Omega_A(r,s)$

$$\mathrm{var}(\Omega_A(r,s)) = 2\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\mathcal{X}_{ri}\mathcal{X}_{si}\,\mathcal{X}_{rj}\mathcal{X}_{sj}}{m_{ii}m_{jj}}\ \gamma_{ij}. \tag{26}$$

Variance of White estimator $\Omega_W(r,s)$ is the special case when $q^{ii} = 1$ and $q^{ij} = 0$ for $i \neq j$, $i, j = 1, 2, \ldots, n$; it is simply

$$\mathrm{var}(\Omega_W(r,s)) = 2\sum_{i=1}^{n}\sum_{j=1}^{n} \mathcal{X}_{ri}\mathcal{X}_{si}\,\mathcal{X}_{rj}\mathcal{X}_{sj}\;\gamma_{ij}. \tag{27}$$

More trivially, the variance of the standard OLS estimator $\Omega_S(r,s)$ is given by

$$\mathrm{var}(\Omega_S(r,s)) = a_{rs}^2 \mathrm{var}(\hat{\sigma}^2), \tag{28}$$

where $\hat{\sigma}^2$ is the usual OLS estimator of variance of ε_i under homoskedasticity assumption and a_{rs} is the (r,s)-th element of $(X'X)^{-1}$. Under heteroskedasticity, $\mathrm{var}(\hat{\sigma}^2)$ is of the form

$$\mathrm{var}(\hat{\sigma}^2) = \frac{2}{(n-k)^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij}. \tag{29}$$

Note that because $M = [\![m_{ij}]\!]$ is idempotent, $\sum_{i=1}^{n}\sum_{j=1}^{n}(\sum_{t=1}^{n} m_{it}m_{jt})^2 = \sum_{i=1}^{n} m_{ii} = n-k$. Therefore, when in fact $\sigma_i^2 = \sigma^2$ for all i,

$$\mathrm{var}(\hat{\sigma}^2) = \frac{2\sigma^4}{n-k},$$

which is the standard formula under the homoskedasticity.

The first three variances have very similar algebraic expressions. Obviously, since $0 < m_{ii} < 1$, the variances of White's consistent estimator will never exceed those of MacKinnon and White's. In the special case of a balanced experimental design for which $m_{ii} = (n-k)/n$, they differ only by a proportional constant. An analytical comparison with the MINQUE variance, however, seems to be difficult due to the complicated nature of the weight q^{ij}. For a comparison of those four variances, we will do a simple numerical exercise for given design matrices with different sample sizes and leverage points.

5. A Numerical Exercise

To the best of our knowledge, a numerical comparison on the MSE of those estimators based on their *exact expression* has not been done before. Since we have the exact expression of the variance of those estimators, no sampling experiment is required for the MSE comparison. However, to study the finite sample behavior of the confidence intervals of regression coefficients in terms of coverage probabilities, we carry out a Monte

Carlo study. Related simulation studies have been done previously by MacKinnon and White (1985) and Nanayakkara and Cressie (1991). MacKinnon and White (1985) did not consider MINQUE and concentrated on the behavior of t-statistic based on different variance-covariance matrix estimators. They also did not experiment with different design matrices. Nanayakkara and Cressie (1991) studied mainly the coverage probability of confidence intervals and did not include MINQUE in their study when the regression model has an intercept. And also they used a 'well- behaved' design matrix in their simulation. Here we use less well-behaved design matrices so as to allow an investigation of the performance of each estimator for different nature of such design matrices. Therefore, our numerical exercise could be viewed as a complimentary to the simulation studies of the above two papers. Our experiment is based on a linear regression model specified as follows:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i,$$

where we set parameter $\beta' = (10.0, 3.5, 2.5)$. The first regressor x_1 was obtained from independent log-normal pseudo-random numbers generator with the corresponding normal variable has mean zero and variance unity. That is, $x_{1i} = e^{z_i}$ where z_i is a $\mathcal{N}(0,1)$ variable. The regressor x_2 is simply a $\mathcal{N}(2,1)$ variable. The choice of the distributions here is to generate quite high enough point of leverage in the design matrix, so we could examine its effect on the behavior of estimators under consideration. In this experiment all pseudo-random $\mathcal{N}(0,1)$ variables were generated by the IMSL routine RNNOA, and the log-normal variable were generated directly from the routine RNLNL.

To investigate the effect of heteroskedasticity on the performance of the estimators we assumed a certain structure of the disturbances. In particular, we assumed

$$\varepsilon_i = \sigma_i v_i,$$

where σ_i is a function of non-stochastic variables and v_i is a white noise process. The variance of ε_i then is σ_i^2, which determines the nature of the heteroskedasticity depending on its prespecified functional forms. In our experiment we considered two different models of heteroskedasticity. First, we assumed the heteroskedasticity was induced by the regressors according to

$$\text{Model 1: } \sigma_i^2 = \lambda_0 + \lambda_1 x_{1i} + \lambda_2 x_{2i}^2.$$

The second structure of heteroskedasticity was specified as

$$\text{Model 2: } \sigma_i^2 = \theta_0 + \theta_1 u_i + \theta_2 u_i^2,$$

where u_i was drawn independently from $\mathcal{N}(0,1)$. The experiment was conducted for 1000 replications with x_{1i}, x_{2i} and u_i held fixed in each replication.

For each structure we carried out the experiments by varying degree of heteroskedasticity, which can be easily accomplished by selecting different values of $\lambda' = (\lambda_0, \lambda_1, \lambda_2)$ and $\theta' = (\theta_0, \theta_1, \theta_2)$. Following Chesher and Jewitt (1987), we measure the degree of heteroskedasticity by the ratio $\max_i(\sigma_i^2)/\min_i(\sigma_i^2)$. The value of 1 for this ratio indicates homoskedasticity and the values of greater than 1 correspond to the presence of heteroskedasticity. Four different sets of $\lambda' = (\lambda_0, \lambda_1, \lambda_2)$ and $\theta' = (\theta_0, \theta_1, \theta_2)$ were chosen. For the first set we took $\lambda' = (20.0, 0.0, 0.0)$ and $\theta' = (20.0, 0.0, 0.0)$, corresponding to the homoskedastic case where the standard OLS variance estimator is appropriate. This case was considered to check the cost of using the alternative variance- covariance estimators when in fact there is no heteroskedasticity. In the second, third and fourth set of experiments we took the values of $(20.0, 0.01, 3.5)$, $(20.0, 0.01, 7.0)$, and $(20.0, 0.01, 10.5)$ for λ, and $(20.0, 0.01, 10.0)$, $(20.0, 0.01, 20.0)$, and $(20.0, 0.01, 30.0)$ for θ, producing a relatively moderate to a very severe degree of heteroskedasticity.

One way to study the finite sample performance of an estimator is to use the MSE, which in some sense encompasses both the bias and the variance of the estimator. We carried out each experiment with eight different sample sizes $n = 30, 40, 50, \ldots, 100$. In every experiment, for each sample size, we calculated the MSE of each estimator from the *exact expression* of variances we derived in Section 4. Note, these computations does not require any replication, and therefore, are not subject to sampling error.

As we mentioned earlier, the MINQUE procedure may produce negative estimates of σ_i^2 and even negative estimates of the variance of $\hat{\beta}$. Given this situation we also considered a truncated MINQUE with each negative estimate of σ_i^2 being replaced by $\hat{\varepsilon}_i^2/(1-h_{ii})$, where $\hat{\varepsilon}_i$ is the OLS residual and h_{ii} is the diagonal element of the hat matrix $X(X'X)^{-1}X'$. Since the algebraic expression of this estimator is difficult to obtain, we estimated the bias, the variance and the MSE of the estimator from the sampling experiment. For each experiment and sample size, let v_{kr} and $\hat{v}_{kr}$ denote the true and the estimate of the variance of the OLS estimator $\hat{\beta}_k$ in the r-th replication, respectively. The bias of the variance estimator of $\hat{\beta}_k$ was estimated by the average bias; we denote this as

$$\widehat{BIAS} = \sum_{r=1}^{1000} \frac{(\hat{v}_{kr} - v_{kr})}{1000}$$

The variance and the MSE were estimated by

$$\widehat{VAR} = \sum_{r=1}^{1000} \frac{(\hat{v}_{kr} - \bar{v}_k)^2}{1000}$$

and

$$\widehat{MSE} = \sum_{r=1}^{1000} \frac{(\hat{v}_{kr} - v_{kr})^2}{1000},$$

respectively, where $\bar{v}_k$ is the sample average of all $\hat{v}_{kr}$. For comparison, we also carried out the computations for all other procedures.

Since the variance-covariance matrix estimators are usually used for statistical inferences, it is important to study the performance of each estimator in terms of statistical inference. In our experiment we used the estimators to estimate the confidence interval coverage probabilities for the regression parameters. We assumed the distribution of $\hat{\beta}_k/\sqrt{\text{var}(\hat{\beta}_k)}$ is Student's t, where $\hat{\beta}_k$ $(k = 1, 2)$ is the OLS estimator of β_k under the complete model and $\text{var}(\hat{\beta}_k)$ is the variance of $\hat{\beta}_k$ in a given procedure. For each procedure we estimated $\Pr(\hat{\beta}_k - t_{\alpha/2}\sqrt{\text{var}(\hat{\beta}_k)} < \beta_k < \hat{\beta}_k + t_{\alpha/2}\sqrt{\text{var}(\hat{\beta}_k)})$, where $t_{\alpha/2}$ is the $\alpha/2$-th quantile of Student's t distribution. We took the nominal size $\alpha = 5\%$. The estimate of this probability is $\hat{p} = F/R$, where F is the observed frequency of β_k being in its 95% confidence interval and R is the number of replications. Since we have 1000 replications, an estimate of the standard error of $\hat{p}$ is $\sqrt{\hat{p}(1-\hat{p})/1000}$. We also carried out this computation for a procedure which utilizes the true variance of $\hat{\beta}$. In this particular case, $t_{\alpha/2}$ is the $\alpha/2$-th quantile of the standard normal distribution.

Summary of the results of our experiments on the bias, the variance and the mean square errors is presented in Tables I through III. Table I contains some information for homoskedastic case, while Tables II and III present the results for the case of heteroskedasticity of Model 1 for sample sizes 60 and 100, respectively. The rests are not reported as they are generally similar to those presented in the tables. Also, the results for heteroskedasticity of Model 2 have quite similar pattern to those of Model 1. Information contained in these tables is generally self explanatory. Here we mention only some interesting points.

In Tables I through III, the numbers in parentheses in those tables indicate that the results were estimated from the sampling experiment, which evidently are reasonably close to those calculated from their exact expressions. The results for homoskedastic case in Table I shows, as one should expect, that the OLS performed quite well in terms of bias, variance and MSE. As we observe the results in other tables, in each experiment and for a given sample size, the OLS variance estimator generally has the smallest variance, except for a few cases when the heteroskedasticity is very strong. Among the robust procedures, the pattern of their variances is more systematic. As it is expected, the variance of White's estimator is always smaller than that of MacKinnon and White's (MWE). It is also evident that MINQUE consistently have much smaller bias but larger variances than the other three variance estimators.

The MSE results for heteroskedastic cases are quite surprising. In Tables II and III we present only for sample sizes 60 and 100; the other results are very similar. In Table II, except for the coefficient β_0, the OLS estimator generally has smallest MSE, even in the presence of strong heteroskedasticity. Such surprising results are more prominent for sample size 100, in Table III, where we observe that MSE's associated with OLS are the smallest ones for all β's. Based on these results, one might be tempted to suggest that OLS is the estimator of choice even in the presence of heteroskedasticity. Such a conclusion, however, requires cautions, simply because MSE may not be an appropriate criterion to characterize an estimator. As we observe from those tables, the MSE's of the robust procedures are very much dominated by the variance part. Consequently, an MSE comparison is essentially a variance comparison, and whether or not this is an appropriate approach remains questionable. In the subsequence discussion we will make bias comparisons across different design matrices (represented by different sample sizes) and degrees of heteroskedasticity. Also, we will study the behavior of t-statistic associated with each procedure through simulated confidence interval coverage probabilities.

Since the true variance of $\hat{\beta}$ differs across design matrices and degrees of heteroskedasticity (see Table VI), a meaningful comparison requires some adjustment that eliminates the effect of these differences. In Tables IV and V we present the relative bias which we define as the ratio between the bias and the true variance of $\hat{\beta}$. The sample sizes 60 and 100 in Table IV are to represent the design matrices with *small maximum* value of h_{ii}, which is regarded as a measure of point of leverage, while the sample sizes 40 and 90 in Table V correspond to those with *large maximum* value of h_{ii}. The following figures describe the nature of all eight design matrices in terms of h_{ii}:

Sample Size	:	30	40	50	60	70	80	90	100
$\max(h_{ii})$	:	0.609	0.666	0.500	0.266	0.377	0.417	0.749	0.274
$\min(h_{ii})$	:	0.035	0.025	0.021	0.017	0.015	0.013	0.012	0.011
$\frac{\max(h_{ii})}{\min(h_{ii})}$	:	17.604	26.320	24.380	15.952	26.007	32.795	64.603	26.095

Here we should note that to study the effects of different degree of points of leverage, we did not attempt to construct design matrices for various sample sizes with the same measure of points of leverage of the regression design.

General observation on Tables IV and V reveals an obvious result that OLS performs well when disturbances are homoskedastic, but its performance gets worse and worse as the degree of heteroskedasticity increases. White's estimator seems to possess special behavior. The estimator clearly exhibits a large bias if in fact the disturbances are homoskedastic. The downward bias is of course guaranteed in this case since the degree of heteroskedasticity is less than 2. It is quite evident, however, that White's estimator tend to underestimate the true variance even in the presence of strong heteroskedasticity. When the disturbances are indeed heteroskedastic, even though the biases of White's estimator are smaller than those of OLS if the design matrix is relatively balanced, it is not true if the design matrix is unbalanced. Clearly, White's estimator is very sensitive to the presence of high point of leverage. In such a situation the performance of White's estimator is no better than that of OLS, and it is even worse especially for inferences regarding the coefficient β_1.

MacKinnon and White's estimator performs quite well in terms of its bias. Even though there is evidence that it is also sensitive toward the unbalancedness of the design matrix, the effect is not as severe as that in White's estimator. In the extreme case where both $\max(h_{ii})$ and $\max(h_{ii})/\min(h_{ii})$ are high (sample size 90), the relative biases are still quite severe, but they are small enough compared with those of OLS and White's estimators. As in the case of White's estimator, the most severe effect of point of leverage is on the variance of $\hat{\beta}_1$ which is associated with the regressor contributing most to the presence of high point of leverage.

From the sampling experiments, MINQUE is of course the only procedure which consistently produced very small bias, because theoretically its bias is zero irrespective of the nature of the design matrix and the degree of heteroskedasticity. However, to achieve its unbiasedness property, MINQUE seems to have to bear the cost in the form of producing large variance and negative estimates when the design matrix is unbalanced. Our results also indicate that the large maximum value of h_{ii} affects the variance of MINQUE corresponding to $\hat{\beta}_1$. In Table I, for example, the variance of MINQUE for $\hat{\beta}_1$ explodes in sample size 30. The same is also true for sample sizes 40 and 90. Even though the variance of MWE also increases, the effect does not seem as severe as in the MINQUE case. The *ad hoc* truncated MINQUE, denoted by MINQUE1 in the tables, which is obtained by replacing any negative estimate of σ_i^2 by $\hat{\varepsilon}_i^2/(1-h_{ii})$, exhibits quite large bias and variance.

Next, we study the performance of different estimators in terms of coverage probabilities. In Figures I through IV we present the estimates of 95% confidence interval coverage probability for β_2 and β_1 for different sample sizes. They illustrate how the use of different variance estimators of each regression coefficient alters the coverage probabilities in the absence and presence of heteroskedasticity. When the disturbances are homoskedastic, in Figure I, the robust variance estimators cover β_2 quite nicely. Even though they are not as good as OLS, the cost of using them does not seem too high, except White's estimator whose coverage is the farthest away from 95%. When the disturbances are indeed heteroskedastic, in Figure II, all the robust procedures perform much better than the OLS which now cover β_2 far bellow 95% of the time. Interestingly, even though the truncated MINQUE has large MSE, it performs reasonably well in terms of coverage probabilities. The performance of MWE also appears to be good for this case.

The behavior of the variance estimators for $\hat{\beta}_1$, in Figures III and IV, is much different. In both homoskedastic and heteroskedastic cases, all the robust estimators perform very poorly. In the heteroskedastic case, even though there is slight improvement in the performance of the robust estimators and some deterioration in the performance of OLS, the robust estimators are still worse than the OLS. This illustration shows once again that the use of the robust procedures, especially White's, can lead to a serious inferential problem when the design matrix exhibits high points of leverage.

6. Conclusion

In this paper we have reconsidered MINQUE as an alternative procedure for estimating variance-covariance matrix in a general heteroskedastic model. We showed that the problem can be approached very easily within the framework of variance components models, where the heteroskedastic model is a special case. By construction, MINQUE incorporates all the information on the design matrix in order to eliminate the bias which is known to be the leading cause of inferential problems of White's estimator. Our Monte Carlo study, however, indicated that high points of leverage in the design matrix can lead to negative estimates of MINQUE and dramatically increase its variance. We also considered an *ad hoc* truncation of MINQUE by replacing the negative estimates by some small positive values. In terms of coverage probabilities, this truncated MINQUE performs reasonably well compared to the other robust procedures. It is, however, quite desirable to truncate those negative estimates in a more systematic way. Overall, based on the previous two simulation studies of MacKinnon and White (1985) and Nanayakkara and Cressie (1991) and our results on the bias, variance and coverage probabilities, the simple almost unbiased estimator suggested by MacKinnon and White seems to be preferable for practical purposes.

MINQUE is developed within the framework of variance components model. This framework is very rich and it encompasses many econometric models. One interesting extension of our approach is the estimation of the autoregressive conditional heteroskedasticity (ARCH) models. Suprayitno (1992) applied MINQUE for various ARCH models to provide an alternative method to the maximum likelihood estimation procedure. This approach might be promising since MINQUE does not require the explicit distributional assumption. Also in this case, MINQUE might have better finite sample properties since here we need to estimate only a few ARCH coefficients.

Table I
Biases, Variances and Mean Square Errors for Homoskedastic Case

Sample Size	Coef.	Estimation Procedure †								
		OLS		White		MWE		MINQUE		MINQUE1
						Bias×10				
30	β_0	0.00	(0.43)	-5.97	(-5.87)	0.00	(0.05)	0.00	(-0.06)	(3.26)
	β_1	0.00	(0.05)	-1.95	(-2.00)	0.00	(-0.15)	0.00	(-0.45)	(1.90)
	β_2	0.00	(0.10)	-1.40	(-1.34)	0.00	(0.08)	0.00	(0.10)	(0.66)
						Variance×100				
30	β_0	118.93	(115.75)	213.35	(214.01)	301.50	(300.33)	347.81	(341.01)	(366.86)
	β_1	1.83	(1.78)	3.55	(3.04)	18.33	(15.38)	107.00	(90.52)	(70.23)
	β_2	6.77	(6.59)	13.36	(13.89)	19.34	(20.09)	23.17	(24.08)	(24.70)
						Mean Square Error×100				
30	β_0	118.93	(115.94)	248.96	(248.50)	301.50	(300.33)	347.81	(341.01)	(377.50)
	β_1	1.83	(1.78)	7.36	(7.06)	18.33	(15.40)	107.00	(90.72)	(73.87)
	β_2	6.77	(6.60)	15.32	(15.67)	19.34	(20.10)	23.17	(24.09)	(25.15)
						Bias×10				
60	β_0	0.00	(0.02)	-1.54	(-1.53)	0.00	(0.01)	0.00	(0.00)	(0.39)
	β_1	0.00	(0.00)	-0.29	(-0.26)	0.00	(0.04)	0.00	(0.04)	(0.15)
	β_2	0.00	(0.00)	-0.34	(-0.34)	0.00	(0.01)	0.00	(0.02)	(0.08)
						Variance×100				
60	β_0	20.20	(18.86)	47.84	(46.54)	57.58	(55.51)	63.74	(60.74)	(60.96)
	β_1	0.20	(0.18)	0.72	(0.74)	1:07	(1.10)	1.34	(1.35)	(1.40)
	β_2	0.76	(0.71)	1.55	(1.53)	1.88	(1.86)	2.09	(2.05)	(2.06)
						Mean Square Error×100				
60	β_0	20.20	(18.86)	50.22	(48.88)	57.58	(55.51)	63.74	(60.74)	(61.12)
	β_1	0.20	(0.18)	0.80	(0.81)	1.07	(1.10)	1.34	(1.35)	(1.43)
	β_2	0.76	(0.71)	1.66	(1.66)	1.88	(1.86)	2.09	(2.05)	(2.06)
						Bias×10				
100	β_0	0.00	(-0.02)	-0.62	(-0.65)	0.00	(-0.04)	0.00	(-0.05)	(0.13)
	β_1	0.00	(0.00)	-0.09	(-0.10)	0.00	(-0.02)	0.00	(-0.02)	(0.01)
	β_2	0.00	(0.00)	-0.11	(-0.11)	0.00	(0.00)	0.00	(0.00)	(0.02)
						Variance×100				
100	β_0	2.96	(2.98)	6.63	(6.35)	7.78	(7.36)	8.95	(8.36)	(8.26)
	β_1	0.01	(0.01)	0.08	(0.07)	0.11	(0.11)	0.17	(0.15)	(0.15)
	β_2	0.10	(0.10)	0.26	(0.26)	0.30	(0.29)	0.31	(0.31)	(0.31)
						Mean Square Error×100				
100	β_0	2.96	(2.98)	7.01	(6.78)	7.78	(7.36)	8.95	(8.36)	(8.26)
	β_1	0.01	(0.01)	0.08	(0.08)	0.11	(0.11)	0.17	(0.15)	(0.15)
	β_2	0.10	(0.10)	0.28	(0.27)	0.30	(0.29)	0.31	(0.31)	(0.31)

† Numbers in parentheses were calculated from sampling experiment.

Table II
Biases, Variances and Mean Square Errors for Heteroskedasticity of Model 1
(Sample Size=60, Max(h_{ii})=.266, Min(h_{ii})=.017)

$\frac{\max(\sigma_i^2)}{\min(\sigma_i^2)}$	Coef.	Estimation Procedure †								
		OLS		White		MWE		MINQUE		MINQUE1
						Bias×10				
3.25	β_0	7.33	(7.42)	-1.97	(-1.85)	0.17	(0.30)	0.00	(0.13)	(0.74)
	β_1	0.55	(0.56)	-0.32	(-0.25)	0.07	(0.16)	0.00	(-0.09)	(0.28)
	β_2	-0.72	(-0.71)	-0.61	(-0.60)	-0.01	(0.00)	0.00	(0.01)	(0.17)
						Variance×100				
3.25	β_0	62.75	(58.43)	74.88	(75.78)	89.62	(90.39)	97.39	(97.56)	(98.09)
	β_1	0.61	(0.57)	1.44	(1.63)	2.04	(2.30)	2.39	(2.65)	(2.81)
	β_2	2.37	(2.21)	6.48	(6.71)	7.70	(8.00)	8.33	(8.67)	(8.78)
						Mean Square Error×100				
3.25	β_0	116.45	(113.44)	78.77	(79.19)	89.65	(90.48)	97.39	(97.57)	(98.64)
	β_1	0.91	(0.88)	1.55	(1.70)	2.05	(2.32)	2.39	(2.66)	(2.89)
	β_2	2.90	(2.71)	6.85	(7.07)	7.70	(8.00)	8.33	(8.67)	(8.81)
						Bias×10				
5.42	β_0	14.66	(14.80)	-2.40	(-2.16)	0.34	(0.59)	0.00	(0.25)	(1.10)
	β_1	1.11	(1.12)	-0.36	(-0.25)	0.15	(0.27)	0.00	(0.13)	(0.41)
	β_2	-1.45	(-1.42)	-0.88	(-0.85)	-0.02	(-0.01)	0.00	(0.03)	(0.26)
						Variance×100				
5.42	β_0	135.49	(126.59)	118.27	(123.24)	141.11	(147.10)	151.14	(157.17)	(159.25)
	β_1	1.32	(1.22)	2.51	(2.96)	3.46	(4.07)	3.89	(4.52)	(4.84)
	β_2	5.12	(4.78)	16.02	(16.65)	18.98	(19.79)	20.44	(21.38)	(21.75)
						Mean Square Error×100				
5.42	β_0	350.38	(345.80)	124.03	(127.92)	141.22	(147.45)	151.14	(157.23)	(160.46)
	β_1	2.54	(2.49)	2.64	(3.02)	3.49	(4.15)	3.89	(4.54)	(5.01)
	β_2	7.21	(6.80)	16.79	(17.37)	18.98	(19.79)	20.44	(21.38)	(21.82)
						Bias×10				
7.50	β_0	21.99	(22.19)	-2.83	(-2.48)	0.50	(0.88)	0.00	(0.38)	(1.47)
	β_1	1.66	(1.68)	-0.39	(-0.24)	0.22	(0.39)	0.00	(0.18)	(0.54)
	β_2	-2.17	(-2.13)	-1.15	(-1.10)	-0.03	(-0.02)	0.00	(0.05)	(0.36)
						Variance×100				
7.50	β_0	238.45	(223.30)	178.02	(188.83)	212.07	(225.53)	225.05	(239.43)	(243.52)
	β_1	2.32	(2.17)	3.93	(4.73)	5.33	(6.42)	5.84	(6.98)	(7.51)
	β_2	9.01	(8.44)	30.19	(31.33)	35.72	(37.20)	38.43	(40.16)	(40.91)
						Mean Square Error×100				
7.50	β_0	722.04	(715.78)	186.02	(194.98)	212.33	(226.32)	225.05	(239.58)	(245.69)
	β_1	5.08	(5.00)	4.08	(4.79)	5.38	(6.57)	5.84	(7.01)	(7.81)
	β_2	13.72	(12.98)	31.50	(32.54)	35.72	(37.20)	38.43	(40.17)	(41.04)

† Numbers in parentheses were calculated from sampling experiment.

Table III
Biases, Variances and Mean Square Errors for Heteroskedasticity of Model 1
(Sample Size=100, Max(h_{ii})=.274, Min(h_{ii})=.011)

$\frac{\max(\sigma_i^2)}{\min(\sigma_i^2)}$	Coef.	Estimation Procedure †								
		OLS		White		MWE		MINQUE		MINQUE1
						Bias×10				
4.42	β_0	1.25	(1.25)	-1.69	(-1.72)	-0.30	(-0.36)	0.00	(-0.08)	(0.23)
	β_1	-0.23	(-0.23)	-0.28	(-0.31)	-0.06	(0.10)	0.00	(-0.05)	(0.01)
	β_2	-1.04	(-1.04)	-0.33	(-0.32)	-0.05	(-0.04)	0.00	(0.01)	(0.06)
						Variance×100				
4.42	β_0	11.71	(11.92)	24.61	(22.67)	32.19	(28.89)	41.56	(36.25)	(36.00)
	β_1	0.05	(0.05)	0.45	(0.38)	0.78	(0.65)	1.26	(1.04)	(1.03)
	β_2	0.41	(0.41)	2.43	(2.35)	2.80	(2.71)	3.03	(2.92)	(2.95)
						Mean Square Error×100				
4.42	β_0	13.28	(13.48)	27.46	(25.64)	32.28	(29.02)	41.56	(36.25)	(36.06)
	β_1	0.11	(0.11)	0.53	(0.48)	0.78	(0.66)	1.26	(1.04)	(1.03)
	β_2	1.49	(1.50)	2.53	(2.45)	2.80	(2.71)	3.03	(2.92)	(2.95)
						Bias×10				
7.71	β_0	2.49	(2.52)	-2.75	(-2.79)	-0.60	(-0.67)	0.00	(-0.11)	(0.35)
	β_1	-0.46	(-0.46)	-0.47	(-0.52)	-0.13	(0.19)	0.00	(-0.08)	(0.01)
	β_2	-2.08	(-2.07)	-0.54	(-0.52)	-0.09	(-0.08)	0.00	(0.02)	(0.09)
						Variance×100				
7.71	β_0	27.80	(28.32)	61.47	(55.75)	82.09	(72.49)	107.84	(92.42)	(92.19)
	β_1	0.13	(0.13)	1.18	(0.97)	2.06	(1.68)	3.39	(2.75)	(2.73)
	β_2	0.97	(0.99)	7.11	(6.87)	8.23	(7.93)	8.93	(8.58)	(8.67)
						Mean Square Error×100				
7.71	β_0	34.02	(34.66)	69.05	(63.55)	82.45	(72.95)	107.84	(92.43)	(92.31)
	β_1	0.34	(0.34)	1.40	(1.24)	2.08	(1.72)	3.39	(2.76)	(2.73)
	β_2	5.29	(5.29)	7.41	(7.14)	8.24	(7.93)	8.93	(8.58)	(8.68)
						Bias×10				
10.87	β_0	3.74	(3.79)	-3.82	(-3.86)	-0.90	(-0.99)	0.00	(-0.14)	(0.47)
	β_1	-0.70	(-0.69)	-0.65	(-0.73)	-0.19	(-0.27)	0.00	(-0.10)	(0.01)
	β_2	-3.12	(-3.11)	-0.76	(-0.73)	-0.14	(-0.11)	0.00	(0.02)	(0.13)
						Variance×100				
10.87	β_0	51.25	(52.19)	117.21	(105.62)	157.49	(138.18)	207.83	(176.90)	(177.11)
	β_1	0.23	(0.24)	2.25	(1.84)	3.96	(3.21)	6.56	(5.29)	(5.25)
	β_2	1.78	(1.82)	14.32	(13.81)	16.59	(15.95)	18.01	(17.28)	(17.49)
						Mean Square Error×100				
10.87	β_0	65.21	(66.54)	131.83	(120.52)	158.31	(139.16)	207.83	(176.92)	(177.33)
	β_1	0.72	(0.72)	2.68	(2.36)	4.00	(3.29)	6.56	(5.30)	(5.25)
	β_2	11.51	(11.48)	14.90	(14.35)	16.61	(15.96)	18.02	(17.28)	(17.49)

† Numbers in parentheses were calculated from sampling experiment.

Table IV
Relative Bias †
for Homoskedastic Case and Heteroskedasticity of Model 1

Sample Size	$\frac{\max(\sigma_i^2)}{\min(\sigma_i^2)}$	Coef.	OLS	White	MWE	MINQUE	MINQUE1
		β_0	0.472	-11.338	0.270	0.202	7.154
40	1.00	β_1	0.000	-47.468	-2.110	-3.165	79.114
		β_2	0.000	-9.958	1.062	1.195	4.249
		β_0	26.658	-10.647	1.099	0.523	9.104
40	3.00	β_1	1.971	-47.306	-0.394	-2.628	80.815
		β_2	-2.612	-9.813	-0.158	1.662	4.669
		β_0	43.554	-10.215	1.817	0.641	10.344
40	4.94	β_1	2.869	-47.346	-0.478	-2.869	81.779
		β_2	-3.721	-9.755	-0.226	1.804	4.849
		β_0	55.239	-9.899	2.312	0.723	11.182
40	6.82	β_1	3.383	-47.368	-0.376	-3.008	81.842
		β_2	-4.335	-9.678	-0.307	1.883	4.905
		β_0	-0.085	-5.527	-0.680	-0.595	3.486
90	1.00	β_1	0.000	-56.548	0.000	5.952	107.143
		β_2	-0.390	-4.677	-0.779	-0.779	0.390
		β_0	8.539	-7.210	-0.716	-0.102	4.500
90	4.42	β_1	-25.735	-63.725	-15.931	6.127	84.559
		β_2	-21.647	-6.088	-0.676	0.507	0.676
		β_0	12.252	-7.936	-0.987	0.146	5.120
90	7.69	β_1	-32.483	-64.965	-20.108	6.961	77.340
		β_2	-27.652	-6.481	-0.972	-0.324	0.864
		β_0	14.262	-8.341	-1.167	0.256	5.323
90	10.82	β_1	-35.008	-66.064	-22.021	6.776	76.228
		β_2	-30.546	-6.665	-1.031	-0.238	1.031

† The figures below each procedure are the percentage of the ratio between the bias and the true variance. For homoskedastic case, the numbers below OLS and MWE were calculated from sampling experiment. All the numbers below MINQUE were calculated from sampling experiment.

Table V

Relative Bias †

for Homoskedastic Case and Heteroskedasticity of Model 1

Sample Size	$\frac{\max(\sigma_i^2)}{\min(\sigma_i^2)}$	Coef.	OLS	White	MWE	MINQUE	MINQUE1
		β_0	0.083	-6.419	0.0417	0.000	1.626
60	1.00	β_1	0.000	-12.262	1.691	1.691	6.342
		β_2	0.000	-7.290	0.214	0.429	1.715
		β_0	22.249	-5.979	0.516	0.395	2.246
60	3.25	β_1	16.091	-9.362	2.048	-2.633	8.192
		β_2	-8.418	-7.132	-0.117	0.117	1.988
		β_0	34.996	-5.729	0.812	0.597	2.626
60	5.42	β_1	24.849	-8.059	3.358	2.910	9.178
		β_2	-11.656	-7.074	-0.161	0.241	2.090
		β_0	43.258	-5.567	0.984	0.748	2.892
60	7.50	β_1	30.094	-7.070	3.988	3.263	9.790
		β_2	-13.291	-7.044	-0.184	0.306	2.205
		β_0	-0.167	-5.174	-0.334	-0.417	1.085
100	1.00	β_1	0.000	-11.097	-2.466	-2.466	1.233
		β_2	0.000	-4.919	0.000	0.000	0.894
		β_0	5.961	-8.059	-1.431	-0.381	1.097
100	4.42	β_1	-13.249	-16.129	-3.456	-2.880	0.576
		β_2	-20.062	-6.366	-0.965	0.193	1.157
		β_0	8.315	-9.184	-2.004	-0.367	1.169
100	7.71	β_1	-17.300	-17.676	-4.889	-3.009	0.376
		β_2	-25.584	-6.642	-1.107	0.246	1.107
		β_0	9.609	-9.815	-2.312	-0.360	1.208
100	10.90	β_1	-19.542	-18.146	-5.304	-2.792	0.279
		β_2	-28.169	-6.862	-1.264	0.181	1.174

† See note on Table IV.

Table VI
True Variances of $\hat{\beta}$ †

Sample Size	Coef.	H(0,0)	H(1,1)	H(1,2)	H(1,3)	H(2,1)	H(2,2)	H(2,3)
	β_0	4.0068	5.4141	6.8189	8.2237	8.0799	12.1552	16.2304
30	β_1	0.4971	0.7170	0.9360	1.1550	0.6692	0.8415	1.0137
	β_2	0.9563	1.6766	2.3964	3.1162	1.8673	2.7788	3.6903
	β_0	2.9634	3.8225	4.6792	5.5359	6.1454	9.3290	12.5126
40	β_1	0.0948	0.1522	0.2091	0.2660	0.1027	0.1105	0.1184
	β_2	0.7532	1.2636	1.7735	2.2835	1.5338	2.3146	3.0954
	β_0	2.5372	3.3862	4.2340	5.0817	4.1211	5.7060	7.2909
50	β_1	0.0833	0.1229	0.1623	0.2017	0.0950	0.1067	0.1183
	β_2	0.6340	1.0885	1.5425	1.9965	0.9636	1.2934	1.6231
	β_0	2.3991	3.2946	4.1890	5.0835	4.5405	6.6830	8.8255
60	β_1	0.2365	0.3418	0.4467	0.5516	0.3679	0.4993	0.6308
	β_2	0.4664	0.8553	1.2440	1.6327	0.8938	1.3212	1.7487
	β_0	1.5109	2.4239	3.3359	4.2478	2.9763	4.4422	5.9080
70	β_1	0.0756	0.1297	0.1834	0.2372	0.1010	0.1264	0.1519
	β_2	0.3416	0.7395	1.1371	1.5347	0.6914	1.0413	1.3911
	β_0	1.4829	2.3171	3.1502	3.9834	2.8315	4.1807	5.5298
80	β_1	0.1411	0.2527	0.3640	0.4753	0.1745	0.2079	0.2414
	β_2	0.2684	0.6096	0.9507	1.2918	0.6030	0.9376	1.2723
	β_0	1.1761	1.9557	2.7343	3.5129	2.3632	3.5507	4.7382
90	β_1	0.0336	0.0816	0.1293	0.1771	0.0522	0.0708	0.0894
	β_2	0.2566	0.5913	0.9258	1.2604	0.5835	0.9103	1.2372
	β_0	1.1984	2.0970	2.9945	3.8920	2.2594	3.3205	4.3816
100	β_1	0.0811	0.1736	0.2659	0.3582	0.1267	0.1726	0.2184
	β_2	0.2236	0.5184	0.8130	1.1076	0.4735	0.7234	0.9733

† Numbers under H(i,j) are the true variances in the heteroskedasticity of Model i level j; H(0,0) denotes homoskedastic case.

Figure I: Homoskedastic Case
Estimated 95% C.I. Coverage Prob. of B2

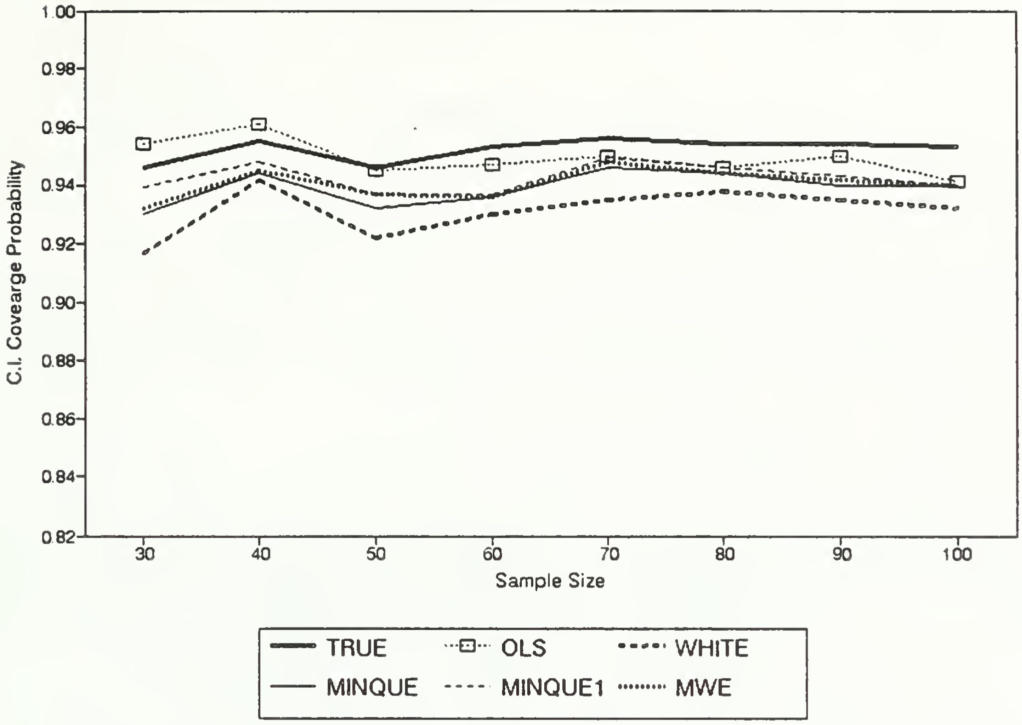


Figure II: Heteroskedastic Case
Estimated 95% C.I. Coverage Prob. of B2

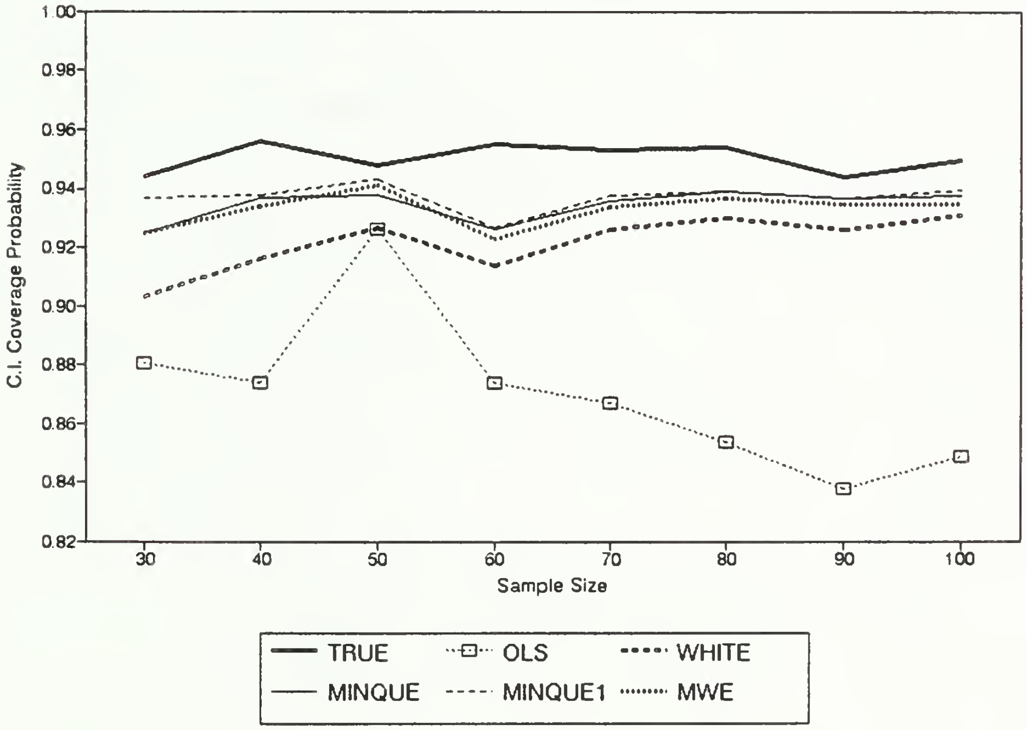


Figure III: Homoskedastic Case
Estimated 95% C.I. Coverage Prob. of B1

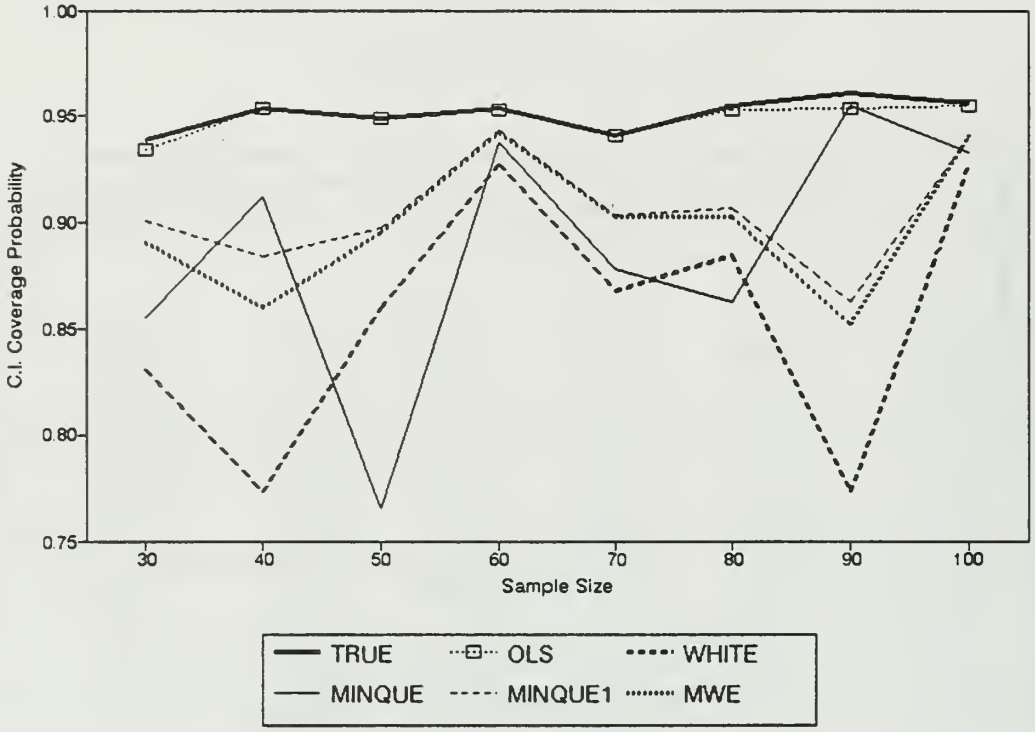
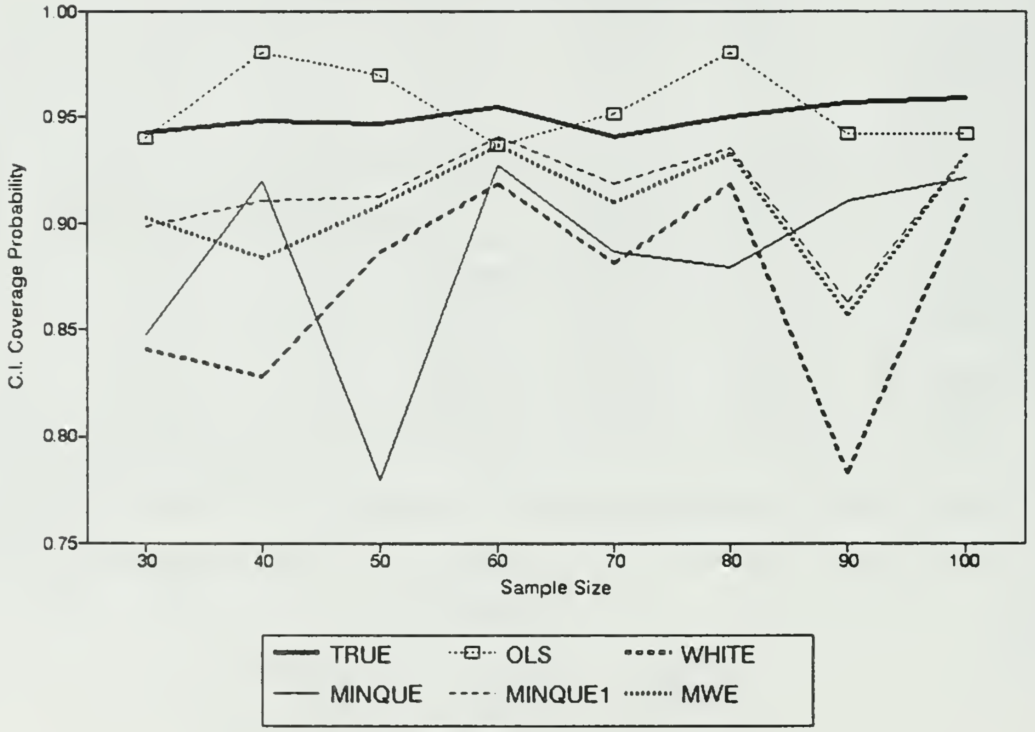


Figure IV: Heteroskedastic Case
Estimated 95% C.I. Coverage Prob. of B1



Appendix

We derive the variance-covariance matrix $\mathcal{V}(\ddot{\mathrm{e}})$ of vector of the OLS residuals squared $\ddot{\mathrm{e}}' = (\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_n^2)$. We assume the disturbance process ε is normally distributed with mean 0 and variance $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, and define $M = [\![m_{ij}]\!] = I - X(X'X)^{-1}X'$. Then $\hat{\varepsilon} = \mathrm{M}\varepsilon$ is also normally distributed with mean 0 and variance-covariance matrix $\Gamma = M\Sigma M$. Explicitly,

$$\Gamma = \begin{bmatrix} \sum_{\ell=1}^n m_{1\ell}^2\sigma_\ell^2 & \sum_{\ell=1}^n m_{1\ell}m_{2\ell}\sigma_\ell^2 & \cdots & \sum_{\ell=1}^n m_{1\ell}m_{n\ell}\sigma_\ell^2 \\ \sum_{\ell=1}^n m_{1\ell}m_{2\ell}\sigma_\ell^2 & \sum_{\ell=1}^n m_{2\ell}^2\sigma_\ell^2 & \cdots & \sum_{\ell=1}^n m_{2\ell}m_{n\ell}\sigma_\ell^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{\ell=1}^n m_{1\ell}m_{n\ell}\sigma_\ell^2 & \sum_{\ell=1}^n m_{2\ell}m_{n\ell}\sigma_\ell^2 & \cdots & \sum_{\ell=1}^n m_{n\ell}^2\sigma_\ell^2 \end{bmatrix} = [\![\gamma_{ij}]\!], \quad \text{say.} \tag{A1}$$

Variance-covariance matrix $\mathcal{V}(\ddot{\mathrm{e}})$ can be derived following the standard procedure. Define a real valued vector $\mathbf{t}' = (t_1, \ldots, t_n)$ and let $\lambda = \sqrt{-1}$, then the characteristic function of $\hat{\varepsilon}' = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)$ is given by

$$G(\mathbf{t}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(\lambda \mathbf{t}'\hat{\varepsilon})\, dF, \tag{A2}$$

where dF is the multivariate normal density function of $\hat{\varepsilon}$. It can be shown that the integral in (A2) factorizes into $(n-k)$ single integrals, each of which is bounded above for all real values of $\hat{\varepsilon}_i$ $(i = 1, \ldots, n)$. Thus (A2) is bounded, and therefore we can differentiate it under the integral signs. Taking partial derivatives of (A2) with respect to $(\lambda t_i, \lambda t_j)$, r and s times respectively, and putting $t_1 = t_2 = \cdots = t_n = 0$, we have

$$\left.\frac{\partial^{rs} G(t_1, \ldots, t_n)}{\partial(\lambda t_i)^r\, \partial(\lambda t_j)^s}\right|_{\mathbf{t}=0} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \hat{\varepsilon}_i^r \hat{\varepsilon}_j^s\, dF = E(\hat{\varepsilon}_i^r \hat{\varepsilon}_j^s). \tag{A3}$$

Evaluation of the right hand side of (A2) leads to [Anderson (1984, pp. 45-46)]

$$\begin{aligned} G(t_1, \ldots, t_n) &= \exp(\frac{1}{2}\lambda \mathbf{t}'\, \Gamma\, \lambda \mathbf{t}) \\ &= \exp(\frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \gamma_{ij}(\lambda t_i)(\lambda t_j)\,), \end{aligned} \tag{A4}$$

and the following immediately follows from (A3):

$$\begin{aligned} E(\hat{\varepsilon}_i^2) &= \gamma_{ii}, & i &= 1, \ldots, n, \\ E(\hat{\varepsilon}_i^4) &= 3\gamma_{ii}^2, & i &= 1, \ldots, n, \\ E(\hat{\varepsilon}_i^2\hat{\varepsilon}_j^2) &= 2\gamma_{ij}^2 + \gamma_{ii}\gamma_{jj}, & i, j &= 1, \ldots, n; \quad i \neq j. \end{aligned}$$

Consequently, the variances of $\ddot{\mathrm{e}}$ are given by

$$\mathrm{var}(\hat{\varepsilon}_i^2) = E(\hat{\varepsilon}_i^4) - (E(\hat{\varepsilon}_i^2))^2 = 2\gamma_{ii}^2 = 2\left(\sum_{\ell=1}^n m_{i\ell}^2\sigma_\ell^2\right)^2, \tag{A5}$$

for $i = 1, \ldots, n$, and the covariances,

$$\mathrm{cov}(\hat{\varepsilon}_i^2, \hat{\varepsilon}_j^2) = E(\hat{\varepsilon}_i^2\hat{\varepsilon}_j^2) - E(\hat{\varepsilon}_i^2)\, E(\hat{\varepsilon}_j^2) = 2\gamma_{ij}^2 = 2\left(\sum_{\ell=1}^n m_{i\ell}m_{j\ell}\sigma_\ell^2\right)^2 \tag{A6}$$

for $i, j = 1, \ldots, n$; $i \neq j$.

References

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, 2nd edition, New York: John Wiley & Sons.

Andrews, D. W. K (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, **59**, 817-858.

Chesher, A. and I. Jewitt (1987), " The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator," *Econometrica*, **55**, 1217-1222.

Eicker, F. (1963), "Asymptotic Normality and Consistency of the Least Squaresfo Families of Linear Regression," *Annals of Mathematical Statistics*, **34**, 447-456.

Graybill, A. F. (1983), *Matrices with Applications in Statistics*, 2nd edition, Belmont: Wadsworth, Inc.

Hartley, H. O. and K. S. E. Jayatillake (1973), "Estimation for Linear Models with Unequal Variances," *Journal of the American Statistical Association*, **68**, 189-192.

Hinkley, D. V. (1977), "Jackknifing in Unbalanced Situations," *Technometrics*, **19**, 285-292.

Horn, S.D., R.A. Horn and D.B. Duncan (1975), "Estimating Heteroscedastic Variances in Linear Models," *Journal of American Statistical Association*, **70**, 380-385.

Horn, S.D. and R.A. Horn (1975), "Comparison of Estimators of Heteroscedastic Variances in Linear Models," *Journal of American Statistical Association*,**70**, 872-875.

MacKinnon, J.G. and H. White (1985), "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, **29**, 305-325.

Mallela, P. (1972), "Necessary and Sufficient Conditions for MINQU-Estimation of Heteroskedastic Variances in Linear Models," *Journal of American Statistical Association*, **67**, 486-487.

Mishkin, F.S. (1990), "Does Correcting for Heteroskedasticity Help ?" *Economics Letters*, **34**, 351-356.

Nanayakkara, N. and N. Cressie (1991), "Robustness to Unequal Scale and Other Departures from the Classical Linear Model", in Stahel, W. and S. Weisberg (eds.), *Direction in Robust Statistics and Diagnostics Part II*, New York: Springer Verlag, pp. 65-113.

Newey, W.K. and K.D. West (1987), "A Simple, Positive Semi- Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, **55**, 703-708.

Rao, C.R. (1970), "Estimation of Heteroscedastic Variances in Linear Model," *Journal of American Statistical Association*, **65**, 161-172.

Rao, C.R. (1972), "Estimation of Variance and Covariance Components in Linear Models," *Journal of American statistical Association*, **67**, 112-115.

Rao, J.N.K. (1972), "On the Estimation of Heteroskedastic Variances," *Biometrics*, **29**, 11-24.

Suprayitno, T. (1992), "Some Applications of Minimum Norm Quadratic Estimation and Eigenvalue-Based Test for Heteroskedasticity", unpublished Ph.D. dissertation, Department of Economics, University of Illinois at Urbana-Champaign.

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, **48**, 817-838.

Wooldridge, J.M. (1989), "A Computationally Simple Heteroskedasticity and Serial Correlation Robust Standard Error for the Linear Regression Models," *Economics Letters*, **31**, 239-243.

Wu, C.J.F. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *The Annals of Statistics*, **14**, 1261-1295.