Comparison of Several Heteroskedasticity-Consistent Covariance Matrix Estimators

Mehmet Orhan[1] Asad Zaman[2]

Abstract

We update MacKinnon and White's (1985) comparison of four HCCME's (Heteroskedasticity Consistent Covariance Matrix Estimators). Several additional estimators from the literature are included in our comparison. We also introduce four new estimators and compare the estimators relative to four different loss functions. For the (indirect) loss functions relating to distribution and inference aspects, an ML based estimator emerges as the top performer. When the loss function focuses directly on the quality of the HCCME as an estimate of the true covariance, then a shrinkage type estimator (JS) appears to be the best choice.
Keywords: Heteroskedasticity-Consistent Covariance Matrix (HCCM) Estimator, Bootstrap, Jackknife, Maximum Likelihood (ML).
JEL classification: C21

1 Introduction

Since the usual estimate of the covariance matrix for OLS estimates in linear regression models is inconsistent in the presence of heteroskedasticity, OLS-based inference procedures go astray. Eicker (1963) showed that it was possible to fix this problem by introducing a *heteroskedasticity consistent covariance matrix estimate* (HCCME) which consistently estimates the covariance matrix of the OLS estimator in presence of arbitrary heteroskedasticity in error variances. White (1980) generalized this to cover many types of dynamic models used by econometricians. Since the original contributions of Eicker and White, many different HCC estimates have been devised. MacKinnon and White (1985) did a Monte-Carlo study comparing four of them and showed that White's (1980) original HCCME could be improved upon.

Our object in this paper is to update the MacKinnon and White (1985) study. We compare most of the HCCME's proposed in the literature so far. In addition, we introduce four new HCCME's in this paper. One of these is based on ML estimates for Random Coefficient (RCMLE) models introduced by Zaman (1997). A second is a James-Stein style estimator based on shrinking the estimated variances towards a common value. A third estimator uses the Chesher-Jewitt formula for the bias of the original Eicker-White HCCME. Our bias-corrected estimator estimates the bias from this formula and subtracts it

[1]Corresponding author, Department of Economics, Fatih University, İstanbul, Turkey, e-mail:morhan@fatih.edu.tr, tel:++90 212 889 08 10, fax:++90 212 889 08 32
[2]Lahore University of Management Science, Lahore, Pakistan, e-mail:asad@lums.edu.pk, fax:++92 42 5722591

from the original estimator. As a fourth new estimator, we formalize the procedure conventionally used by applied econometricians. They conduct a test for heteroskedasticity, using standard OLS covariance formula if the test fails to reject the null of homoskedasticity, and some variant of the Eicker-White if the null is rejected. We call this the pre-test estimator.

These estimators, covering virtually all proposed in the literature, are compared with respect to four different loss criteria. The first two loss criteria assess the quality of inference one can expect to get by replacing the OLS covariance matrix with the HCCME. The ML based HCCME emerges as the overall best estimator in these comparisons. The next two loss criteria assess the accuracy of the HCCME as point estimators of the true covariance matrix. In this category, shrinkage based JS estimators perform very well overall. In cases where the random coefficient model holds, the ML based estimator also performs very well.

2 HCCME's and Bias Corrections

Our basic model is $y = X\beta + \epsilon$ where $\epsilon \sim N(0, \Sigma)$ and $\Sigma = diag(\sigma_1^2, \ldots, \sigma_T^2)$ is the diagonal matrix of variances. The $k \times k$ covariance matrix Ω of the OLS estimate $\hat{\beta} = (X'X)^{-1}X'y$ is easily seen to be

$$\Omega = \mathrm{Cov}(\hat{\beta}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

Since the diagonal matrix Σ contains an increasing number of unknown and unestimable parameters, it did not appear possible to estimate Ω consistently without additional assumptions about Σ. However, Eicker (1963) showed that replacing the diagonal entries σ_i^2 of Σ by squares of the corresponding OLS residuals leads to a consistent estimator of Ω. Explicitly, the **Eicker-White estimator (EW)** for the covariance term is:

$$\hat{\Omega}_{EW} = (X'X)^{-1}X'EX(X'X)^{-1}$$

where $E = \mathrm{diag}(e_1^2, e_2^2, \ldots, e_T^2)$ and e_i is the i-th entry of the residual vector $e = y - X\hat{\beta}$. It is worth noting that e_i^2 is not a consistent estimator of σ_i^2, although it is asymptotically unbiased.

The possibility of making OLS based inference robust to arbitrary forms of heteroskedasticity was appealing to practitioners. After White (1980) taylored the result to econometric applications, substantial use was made of it. It was discovered early that the Eicker-White HCCME has potentially large finite sample biases.

The simulation study of MacKinnon and White (1985) picks up these biases and also shows that a modification of the EW estimator has better finite sample properties. This modification was introduced by Horn, Horn and Duncan (1975) and motivated as follows. Let $H = X(X'X)^{-1}X'$ be the hat matrix (or the projection matrix), and let H_{ii} be the (i, i) diagonal entry of this matrix.

Define $e_i^* = e_i^2/(1-H_{ii})$ to be the squared OLS residual divided by one minus H_{ii}. The expected value of e_i^* is typically closer to σ_i^2 than that of e_i^2, so it is termed an 'almost unbiased' residual by Horn, Horn and Duncan. They suggest replacing E in $\hat{\Omega}_{EW}$ by $E^* =$diag$(e_1^*, e_2^*, \ldots, e_T^*)$. Thus the **Horn, Horn, and Duncan estimator (HD)** is:

$$\hat{\Omega}_{HD} = (X'X)^{-1}X'E^*X(X'X)^{-1}$$

As MacKinnon and White show in their simulation study, this has better finite sample properties than the original Eicker-White estimator.

MacKinnon and White also study a simpler bias-correction for the EW estimator proposed by Hinkley (1977). Making a simple degrees of freedom correction leads to **Hinkley's estimator (Hi)**:

$$\hat{\Omega}_{Hi}(\hat{\beta}) = \frac{T}{T-k}(X'X)^{-1}X'EX(X'X)^{-1}$$

Note that the only difference between EW and Hi is the $\frac{T}{T-k}$ ratio.

Chesher and Jewitt (1987) developed a formula for the bias of the original Eicker-White HCCME. This formula is given below:

$$E(e_i^2) = \sigma_i^2 - 2\sigma_i^2 H_i'H_i + H_i'\Sigma H_i$$

where H_i is the i^{th} column of H. In the above expression $b_i = -2\sigma_i^2 H_i'H_i + H_i'\Sigma H_i$ is the bias term.

This formula clarifies the nature of the improvement provided by the HD estimator over the original EW estimator. Calculations show that HD removes first order bias terms in the EW HCCME. Another way of using the Chesher-Jewitt formula is to estimate the bias of the EW estimator and simply subtract it. To get the Bias Corrected EW, we first estimate b_i by $\hat{b}_i = H_i'EH_i - 2e_i^2 H_i'H_i$. To the extent that $\hat{b}_i$ is a good estimate of b_i, $\tilde{e}_i^2 = e_i^2 - \hat{b}_i$ will have expectation nearly equal to σ_i^2. Defining $\tilde{E} =$diag$(\tilde{e}_1^2, \ldots, \tilde{e}_T^2)$, we get the **bias-corrected Eicker-White (BC)** estimate by replacing E in the Eicker-White formula by $\tilde{E}$.

3 Bootstrap and Jackknife HCCME's

Efron (1982) has suggested a bootstrapping technique which takes into account heteroskedasticity. **Efron's Bootstrap (BR)** resamples over the observations. More specifically, the observations $(y_i, x_i), i = 1, 2, \ldots, T$ are resampled with replacement to obtain $(y_i^*, x_i^*), i = 1, 2, \ldots, T$. $\hat{\beta}$ is obtained for this resample and is called $\hat{\beta}^*$. The corresponding covariance matrix estimator is:

$$BR = \frac{1}{M} \sum_{j=1}^{M} (\hat{\beta}_j^* - \bar{\beta}^*)(\hat{\beta}_j^* - \bar{\beta}^*)' \quad (1)$$

M is the total number of resamples, and $\bar{\beta}^*$ is the average of all estimated $\hat{\beta}_j^*$'s produced via resampling.

Another bootstrapping technique included in our study is **Wu's (1986) bootstrap (Wu)**. Let $(a_j)_{j=1}^T$ be any sequence of constants satisfying $\sum_{j=1}^T a_j = 0$, and $\frac{1}{T}\sum_{j=1}^T a_j^2 = 1$ Let a_i^* be a resample from this sequence and define

$$y_t^* = x_{t\cdot}\hat{\beta} + \frac{e_t}{1 - H_{tt}} a_i^*.$$

For the simulations we carried out, we used constants a_i defined by:

$$a_j = \frac{e_j - \bar{e}}{\sqrt{\frac{1}{n}\sum_{j=1}^n (e_j - \bar{e})^2}}$$

which satisfy Wu's conditions. The HCCME covariance matrix is calculated by the same formula used for Efron's bootstrap. Wu states that his method leads to results similar to those of Horn, and Horn & Duncan (1975) when the parameter of interest is linear. Our simulations bear this out as the two estimators have similar properties.

Different methods of applying jackkniving and bootstrapping lead to alternative forms for the HCCM estimator. Another jacknife estimator studied by MacKinnon and White has the following formula:

$$\hat{\Omega}_{Ja} = \frac{T-1}{T}(X'X)^{-1}[X'\tilde{E}^* X - \frac{1}{T} X'\tilde{e}^* \tilde{e}^{*'} X)](X'X)^{-1}$$

Here, $\tilde{e}_i^* = e_i/(1 - H_{ii})$, $\tilde{e}^*$ is the column vector of $\tilde{e}_i^*$s and $\tilde{E}^*$ is the diagonal matrix of the squares of $\tilde{e}_i^*$. We include this in our study as the **jacknife estimator (Ja)**; see the original articles for motivation and derivation.

4 Other Methods from the Literature

Rao (1970) has devised the **minimum norm quadratic unbiased estimator (MQ)** where he estimates the Σ matrix by $(M. * M)^{-1}. * E$. Here $M = I - X(X'X)^{-1}X'$ and ".*" stands for entry by entry multiplication of the matrices. The reader may refer to Rao (1970) to see the details of the construction as well as motivation for this estimator.

An interesting suggestion is to use the robust regression techniques to estimate the Σ matrix so that the effect of the outliers will be offset to some extent.

See Furno 1996 and 1997 both of which show that there can be significant improvement in estimation when robust regression residuals are used to estimate HCC matrices. Since we deal with normal errors in this simulation, we did not include these methods. Examination of robustness involves a different set of issues from the ones we deal with in the present study.

Another technique commonly used in practice is pre-testing. This involves first testing for heteroskedasticity. If the test does not reject the null of homoskedasticity, we use the conventional OLS covariance estimate. If the null is rejected in favor of heteroskedasticity, then we use White's HCCME. Although there are several tests for heteroskedasticity we applied White's (1980) standard test. Thus our **Pre-Test Heteroskedastic Covariance Estimator (PO)** simulates common practice among applied econometricians. Specifically, we regress squared OLS errors on all linearly independent cross products of the k regressors. The Breusch-Pagan statistic TR^2 for this regression is asymptotically chi squared with $k(k+1)/2$ degrees of freedom. If TR^2 is less than the 95% critical value, we accept homoskedasticity and use the OLS covariance matrix $\sigma^2(X'X)^{-1}$. If TR^2 is greater than the critical value, we use the Eicker-White HCCME.

Yet another possible HCCME can be based on the maximum likelihood estimator for random coefficient models (RCMLE), discussed in Zaman (1997). As shown there, every heteroskedastic model can, under certain assumptions, be reduced to a random coefficient model. Estimating the random coefficients then provides an estimate for the HCCME. We will call this the **Random Coefficient Model Based HCCME (ML)**. We can summarize the basic technique for computing this estimator as follows. Assume that $y_t = x_t'\beta_t$ where $\beta_t \overset{iid}{\sim} N(\beta, \Lambda)$. This random coefficient model is equivalent to a heteroskedastic model where $y_t = x_t'\beta + \epsilon_t$ and $\text{Var}(\epsilon_t) = x_t'\Lambda x_t = \sigma_t^2$. Now RCML estimates of β, Λ automatically furnish estimates $\hat{\sigma}_t^2 = x_t'\hat{\Lambda}_{RCML}x_t$. These variance estimates can be plugged into the formula for the covariance of OLS under heteroskedasticity to provide us with another type of HCCME.

5 Shrinkage Based HCCME

Given variances σ_t^2 which are being estimated by e_t^2, it seems natural to shrink the estimates towards a common mean. Our **Stein type shrinkage estimator (JS)** for the heteroskedastic variances has the following form:

$$\hat{\sigma}_i^2 = \hat{\sigma}^2 + SF * [(e_i^2/(1 - H_{ii})) - \hat{\sigma}^2].$$

That is, the difference between e_i^2 and the overall mean $\hat{\sigma}^2$ is shrunk towards $\hat{\sigma}^2$, where the shrinkage factor is calculated as follows:

$$SF = \left(1 - \frac{T-3}{\delta'\Sigma_\delta^{-1}\delta}\right)^+$$

The '+' sign indicates that the shrinkage factor is constrained to be positive.

The notation and the justification of this shrinkage factor depends on a few approximations given below. Let $\Delta_i = e_i^2/(1-H_{ii}) = e_i^*$. Assume that $E\Delta_i = \sigma_i^2$ (which is true to first order approximation in large samples) and let σ^2 be the average value of the variances σ^2. Let $\hat{\sigma}^2 = \sum_{t=1}^{T} \Delta_t/(T-k)$ be the usual unbiased estimator for the common variance in a regression model. Let $\delta_i = \Delta_i - \hat{\sigma}^2$ be the differences from the mean. Based on Efron and Morris (1975), the formula for the shrinkage factor, SF, is:

$$SF = 1 - \frac{T-3}{\delta' \mathrm{Cov}(\delta)^{-1}\delta}$$

To calculate the covariance matrix for the differences δ, we use the approximation $\Sigma = \hat{\sigma}^2 \mathbf{I}_T$. This allows us to assume that $e \sim N(0, \hat{\sigma}^2 M)$. Then

$$\begin{aligned} \mathrm{Cov}(\delta_i, \delta_j) &= \frac{E(e_i^2 e_j^2) - Ee_i^2 Ee_j^2}{(1-H_{ii})(1-H_{jj})} \\ &= \frac{(m_{ii}m_{jj} + m_{ij}^2 - m_{ii}m_{jj})}{(1-H_{ii})(1-H_{jj})} = \frac{m_{ij}^2}{(1-H_{ii})(1-H_{jj})} \end{aligned}$$

These formulae are used to compute an approximate Stein type shrinkage estimator for the variances. The estimator is labelled JS in the output to follow. We tried another HCCME based on shrinking the residuals e_i^2 instead of the e_i^* using a similar approximation procedure. The two estimators give very similar results, with the second form being marginally superior, so we have only included the results for the second form.

6 Simulation Design and the Data Set

The data set is kept similar to that of MacKinnon and White (1985) for comparability of results. Specifically,

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \epsilon_t \tag{2}$$

Here X_1 is the growth rate of real U.S. disposable income and X_2 is the U.S. treasury bill rate both of which are seasonally adjusted for 1963-3 to 1975-4. The dependent variable may be thought of as the saving rate. Each column of X is demeaned and multiplied by a factor to make the squares of the entries in the column add up to T. T was fixed at 50 in all simulations. As in MacKinnon and White, we simulated several different types of heteroskedasticity.

First, we assess the behavior of all the estimators under homoskedasticity to establish a baseline. This also permits us to see how damaging it is to allow for heteroskedasticity when there is, in fact, none. We examine two cases: case 0a is *low noise homoskedasticity* characterized by $Var(\epsilon_i) = \sigma_i^2 = 1$ and *high noise homoskedasticity* is characterized by $Var(\epsilon_i) = \sigma_i^2 = 5$.

Case 1 is similar to that of MacKinnon and White (1985). We take different variances for the two equal parts of the data set. That is, we take $\epsilon_t \sim N(0, \sigma^2)$ *for* $i = 1, \ldots, 25$, and $\epsilon_t \sim N(0, \sigma^2\alpha^2)$ *for* $i = 26, \ldots, 50$. Here we took $\alpha = 2$ and 30 for cases 1a, and 1b, respectively. We will call this *structural change* in the variances, referring to 1a as *moderate* and 1b as *drastic* structural change.

We also take different variances for each and every error term, namely we generated random variables from $N(0, \gamma)$, where $\gamma = 49$ to make all of the variances different (Case 2). This case is of interest since it is a heteroskedastic model which can also be interpreted as a homoskedastic model with nonnormal *long-tailed errors*.

Cases 3a and 3b are also patterned on MacKinnon and White (1985). Here the variances of the error terms depend on their corresponding regressors: $\sigma_i^2 = 1 + \gamma_1^2 X_{1i}^2 + \gamma_2^2 X_{2i}^2$. Case 3a takes both γ_1 and γ_2 equal to 3, whereas Case 3b sets both of them equal to 10. Since both of these cases are special cases of case 4 below where Λ is a diagonal matrix, we will refer to these as *diagonal RC* cases. Case 3a has *low correlations* with the regressors, while 3b has *high correlations*.

Lastly, we have case 4, which is *the general random coefficients model.* Since the error terms are related to the regressors, this model is a generalization of Case 3. More specifically, the variances are generated according to the equation $\sigma_t^2 = x_t \Lambda x_t'$ where Λ is equal to LL'. L is a fixed lower triangular matrix which we varied as part of the design of our Monte Carlo. This is equivalent to a random coefficients model. Set $y_i = X_i \beta_i$ and generate the random coefficients β_i from $\beta_i \overset{iid}{\sim} N(\beta, LL')$. This reduces to a heteroskedastic model with $\sigma_t^2 = x_t LL' x_t'$. In order to allow for a wide range of possibilities, we generated several different matrices L from the Cauchy distribution. The Monte Carlo sample size is set equal to 2500 and the simulation is run for 33 different L matrices. The Monte Carlo sample size for all bootstraps is 1000.

7 Performance Criteria

Different criteria are used to assess aspects of performance of the estimators. We label the first criterion 'CHI' as an abbreviation for *chi-squared loss.* Define $CHI = (\hat{\beta} - \beta)'(\hat{C})^{-1}(\hat{\beta} - \beta)$. If $\hat{C}$ equals C, the true covariance matrix of $\hat{\beta}$, then CHI will have a χ^2 distribution with k degrees of freedom. Allowing for the estimation of one parameter within C will reduce the degrees of freedom to $k-1$. To the extent that $\hat{C}$ is a good estimator of C, this distribution should still hold approximately. The percentage of times that CHI exceeds the 99 % χ_{k-1}^2 critical values are computed. The closer this is to 1%, the better the estimator $\hat{C}$ is.

The second criterion is due to White and MacKinnon (1985). Their quasi-t statistic is obtained by dividing the OLS estimate by its HCC estimated Standard Error, parallel to the procedure for obtaining the usual statistic. An HC-CME is judged successful to the extent that successful inference can be done based on this quasi-t statistic. We used the t-table for the 99 % critical values as our benchmark to compare the following statistics for each entry of the covariance matrix:

$$t_{loss} = \frac{\hat{\beta}_i - \beta_i}{\hat{C}_{ii}}, i = 1, 2, \ldots, k \quad (3)$$

Our criteria concerning the t-losses are given in (3). This statistic would preferably be 1 % for the perfect estimator, and the further this number is away from 1% the worse the estimator is.

The first two criteria relate to distributional and inference aspects of the HCCME. Our second set of two criteria directly measure the accuracy of the HCCME as an estimate of the true covariance. A mathematically elegant and convenient measure of the difference between a covariance estimate and the true matrix is the entropy-loss (E_{loss}). This was introduced in James and Stein (1961), and has since been used in many theoretical studies. See Haff (1980), Dey and Srinivasan (1985), and Fang and Li (1999) for example. Our third criterion is the entropy loss defined by:

$$E_{loss} = \text{ trace } (\hat{C} - C^{-1}) - \ln(\text{abs}(\det(\hat{C}C^{-1}))) - k \quad (4)$$

where 'trace is' the sum of the diagonal entries of a matrix, 'abs' stands for absolute value, 'det' denotes the determinant and 'ln' is the natural logarithm. If the estimator is perfect in the sense that it hits the true value then, the first term of the summation, the trace, returns k, the logarithm component returns 0, and when k is subtracted from the previous two components of the E_{loss}, one gets 0.

A simpler way to measure the difference is the standard quadratic-loss:

$$Q_l = tr((\hat{C} - C)^2) = \sum_{i=1}^{k} \sum_{j=1}^{k} (\hat{C}_{ij} - C_{ij})^2 \quad (5)$$

This is our fourth and final criterion for measuring the relative accuracy of different estimators.

The above criteria are reported directly for cases 0 to 3. But for Case 4 we displayed the average of the simulation statistics for the 33 different L matrices.

8 Simulation Results for Chi Squared Loss

We will now compare the estimators in the study on the basis of Chi-Squared Loss. Table 1 gives 10^4 times the estimated probabilities of exceeding the critical value for each estimator. Thus 100 is the ideal value. The Monte Carlo standard error of the estimated probability is also indicated as ± 19, for example.

HCCME	CASE 0a	0b	1a	1b	2
ML	99±19	184±27	76±17	100±20	44±13
OL	108±21	176±26	76±17	120±22	40±13
PO	108±21	184±27	76±17	112±21	40±13
JS	116±21	188±27	92±19	76±17	60±15
Ja	144±24	248±31	148±24	52±14	112±21
MQ	160±25	260±32	152±24	64±16	124±22
BR	184±26	280±33	160±25	128±22	116±21
Wu	192±27	284±33	176±26	100±20	144±24
HD	200±28	292±34	176±22	104±20	152±24
Hi	212±29	312±35	192±27	132±23	168±26
BC	216±29	320±35	200±28	108±21	172±26
EW	244±31	352±37	232±30	196±28	216±29

Table 1: χ^2 cases 0a, 0b, 1a, 1b, and 2

Generally speaking, we found that there was a rank ordering of the probabilities of the estimators, and we have listed them in order of increasing probability.

In the case of homoskedasticity with low noise (0a), we found that OLS, PO, JS, and ML had estimated probabilities within one Monte Carlo standard error (MCSE) of the target level 100. This is not surprising as OLS under homoskedasticity is expected to have the right probability. The Pre-Test estimator PO should fail to reject the null 95% of the time and hence reduce to OLS. The JS estimator shrinks towards the null of homoskedasticity and performs similar to the OLS. The strong performance of the ML was a little surprising. It shows that the attempt to estimate the additional heteroskedastic parameters does not cause much deterioration in performance under the null. The Jacknife (Ja), Minque (MQ), Efron's Bootstrap (BR) are somewhat oversized, falling more than two MCSE's above the target level 100. In the third category, Wu, HD, Hi and BC are about three MCSE's above target level, and hence more biased. Finally, the original Eicker-White estimator is in a class by itself as being the most biased.

The same categorization emerges in the case of homoskedasticity with high noise (0b). The statistic CHI has an exact chi-squared distribution only when the covariance estimate $\hat{C}$ exactly equals the true covariance. In the case of OLS with homoskedasticity, the variance σ^2 is being estimated, and hence the distribution will not be exactly chi-squared except asymptotically. This leads to distortions in OLS performance in the high noise case, with an estimated probability of 1.76% as compared to the target 1%. Nonetheless, the OLS has the best performance among all the estimators, and the relative rankings of the estimators remains the same as in the low noise case.

In the case of moderate structural change in variances (1a), the ordering of the estimators in terms of probabilities of rejecting the null remains the same as in the cases 0a and 0b. In case 1a, where the change in variance is less pronounced, the four top estimators ML,OLS,PO and JS appear to be a little undersized. However all are within two standard errors of the target 100, so that this could be just due to the Monte Carlo error. The ranking of the other estimators is the same as in the homoskedastic case. However, drastic structural change (1b) does lead to some interesting variations in the performance of the estimators. Rao's MINQUE (MQ) and Efron's Bootstrap (BR) appear somewhat under the target probability, as is also the shrinkage estimator JS. With the exception of Hinkley (Hi), which is slightly over target, and Eicker-White (EW) which is severely over target, all other estimators are more or less exactly on target in this case.

In the case of the long-tailed errors (case 2 – equivalent to normals with heteroskedasticity), our four top estimators (ML, OLS, PO, JS) have probabilities of rejection significantly below the target 100. The next group (Ja, MQ, BR) are all within one Monte Carlo Standard Error (MCSE) of the target. The third group (Wu, HD, Hi, BC) are all over two MCSE's over target. As usual, EW is in a class by itself as being the most biased. It is interesting that the overall ranking of the estimators is preserved in this case. It appears that ML, OLS, PO and JS are sensitive to normality assumption. This is reasonable as all four are more or less explicitly derived under assumptions of normality.

Table 2 displays the performance of the estimators with respect to chi-squared loss in cases 3a, 3b, and 4. Case 3a is diagonal RC with low correlations. Case 3b is diagonal RC with high correlations. Case 4 is a general RC model (33 cases summarized). All three are special types of random coefficient models, and thus we expect some similarity in the behaviors of the estimators in these cases. Since the ML is especially designed for the diagonal RC models, we expect it to do well here, as indeed it does (case 3a and 3b). It also does well in the nondiagonal RC models, being the top estimator in this set of cases. The jacknife and MINQUE estimators follow closely, showing that they are better adapted to the RC situation. It is immediately evident that the OLS is the worst of the estimators in all three cases, taking over this dubious distinction from EW. Efron's bootstrap (BR), Wu's bootstrap (Wu), and the Horn, Horn and Duncan (HD) estimator fall into the second category in this set of models, being between 1 and 2 MCSD's worse than the ML. The third category consists of Bias Correction (BC) and Hinkley (Hi) which are about 1 MCSD worse than the second category estimators. The fourth category, being distinctly inferior to the third, consists of JS, PO, and EW. Finally, the OLS falls to being the worst estimator in the RC case.

Overall we learn the following lessons from these simulations. We consider

HCCME	CASE 3a	3b	4
ML	224±30	220±29	197±54
Ja	260±32	292±34	242±39
MQ	276±33	300±34	254±40
BR	332±36	348±37	313±56
Wu	340±36	404±39	321±51
HD	340±36	412±40	317±51
BC	364±37	492±43	342±55
Hi	388±39	492±43	360±61
JS	436±40	576±47	439±127
PO	456±42	624±48	491±129
EW	472±42	616±48	425±69
OL	568±46	764±53	878±390

Table 2: χ^2 cases 3a, 3b, and 4

the chi-squared loss function as a proxy for the F-statistic, F-tests for linear restrictions, and similar statistics in regression models. How are rejection probabilities for such tests affected by the use of heteroskedasticity corrections? The answer depends on the type of heteroskedasticity, with estimator rankings differing in the case of random coefficients and in other types of heteroskedasticity. Over all the cases we studied, the ML emerged as a strong performer, being at the top in the RC cases for which it is designed, and being in the top group of estimators in other cases as well. The heavy biases in the original EW turn it into the worst or near-worst estimator in all the cases studied. All other estimators have inconsistent performances, being good in some cases, and bad in others. In particular, the Pre-Test Estimator (representing conventional practice of applied econometricians) has acceptable performance in cases 1 and 2, but is very poor in cases 3 and 4 (the random coefficient models). The jacknife and MINQUE estimators are consistently second rank relative to this criterion. The shrinkage (JS) estimator also performs poorly for this loss function.

9 Rankings for Quasi-t Loss

This loss function is meant to assess the behavior of the t-statistic which is obtained by using HCCME instead of conventional OLS estimates. Since model selection, and inclusion/exclusion of regressors is frequently done on the basis of these t-statistics, it is important to assess the relative performance of the estimators in this area.

A general finding was that in all of the cases considered, the quasi-t statis-

	β_1			β_2	
HCCME	Case 1a	Case 1b	HCCME	Case 1a	Case 1b
ML	52±14	72±17	MQ	72±15	20±10
OLS	56±15	104±20	Ja	72±15	24±10
PO	56±15	104±20	JS	72±17	56±15
JS	68±16	100±20	BR	80±18	60±15
Ja	124±22	80±18	ML	60±18	68±16
BR	124±22	108±21	PO	68±16	80±18
MQ	124±22	84±18	HD	76±17	80±18
Wu	144±24	96±20	OL	72±17	80±18
HD	156±25	104±20	BC	80±18	84±18
BC	160±25	112±21	Wu	72±15	80±18
Hi	164±25	108±21	Hi	84±18	116±21
EW	180±27	136±23	EW	96±20	136±23

Table 3: Quasi-t losses cases 1a, 1b

tic for the constant term β_0 always gave an estimated probability within two MCSE's of the target level 1%. Thus *all* estimators in the study provide reasonable quasi-t statistics for the constant terms. To save space, we will not tabulate these results (which have no further information content).

Another general finding was that in the case of homoskedasticity (cases 0a and 0b), *all* the estimators were within two MCSE of the target level 1%. Since the tables furnished no further information, we do not report these simulations. We conclude that the quasi-t statistic formed by all these estimators is reliable in the case of homoskedasticity. The only exception to this is the original Eicker-White estimator (EW) which appears significantly oversized for some of the coefficients in some of the cases. Thus, except for the EW, there is no cost to using the heteroskedasticity corrections in terms of this criterion; in the null case, they all work fine.

Table 3 displays the performance of the HCCME in the case of structural change (1a,1b). For β_1, the coefficient of disposable income, the radical structural change (1b) gives no trouble to any of the estimators. All come within 2 MCSD of target values. In the case of moderate structural change (1a), OLS, PO, JS, and ML appear significantly below target, with probabilities at roughly 1/2 of 1%. Also, Bias Correction (BC), Hinkley (Hi), and Eicker-White (EW) appear significantly over target with probability of exceeding critical value between 1.6% and 1.8%. For β_2, the coefficient of the interest rate, the outcomes are different. Moderate structural change (1a) poses no difficulty for any of the estimators, all being within 2 MCSD of the target value. Radical structural change damages Jacknife (Ja) and MINQUE (MQ) the most, these having

HCCME	β_1	β_2
EW	68±16	136±23
Wu	44±13	104±20
Hi	40±13	112±21
HD	40±13	104±20
BC	40±13	116±21
MQ	36±12	96±20
Ja	32±11	96±20
BR	32±11	96±20
JS	28±11	88±19
OLS	24±10	56±15
PO	24±10	56±15
ML	16±8	68±16

Table 4: Quasi-t losses Case 2

probabilities of only 1/4 of 1%, substantially below target. Next James-Stein (JS) and Efron's Bootstrap (BR) are somewhat below target at about 1/2 of 1%. The rest of the estimators are within 2 MCSD's of target.

Table 4 displays the performance of the HCCME for the coefficients β_1, β_2 in the case of long-tailed errors (case 2). All estimators face difficulty with the quasi-t statistic for β_1, the coefficient of disposable income. All are below target with EW being the best and the ML the worst. The coefficient β_2 of the interest rate causes difficulty only for OLS and PO which are slightly below target at about 1/2 of 1%. All others are within 2 MCSD's of the target. The ML fares surprisingly well, given that the errors are not normal in this setup. Although it is bad for β_1, all of the other estimators also fare poorly in this case.

Table 5 deals with the random coefficients cases 3a, 3b, and 4. For the coefficient β_1, ML is at or near the top, although a bit higher than target in case 3b. EW, JS, PO, and OLS are all at the bottom in terms of distance from target. Similar patterns hold for the third coefficient β_2. As expected, ML comes out on top in this random coefficients case. It does fine even in the non-diagonal RC case, for which it is not especially adapted.

10 Entropy Loss

The first two loss functions assess the quality of inference based on using HCCME's instead of the usual OLS covariance. The next two, entropy and quadratic loss, directly measure the accuracy of the HCCME as an estimate of the true covariance matrix of the OLS. These two are qualitatively different types of as-

	Quasi t-stat for β_1			Quasi t-stat for β_2		
HCCME	Case 3a	Case 3b	Case 4	Case 3a	Case 3b	Case 4
ML	156±25	184±27	115±38	132±23	104±20	121±50
Ja	128±22	192±27	113±40	112±21	108±21	134±40
MQ	128±22	188±27	117±42	112±21	116±21	139±41
Wu	152±24	256±32	140±45	148±24	144±24	170±48
HD	156±25	252±31	140±45	148±24	136±23	168±49
BC	156±25	260±32	148±40	160±25	160±25	173±41
BR	168±26	268±32	143±54	148±24	116±21	177±59
Hi	176±26	288±33	156±49	164±25	180±27	191±56
EW	212±29	304±34	179±53	188±27	208±29	221±61
JS	288±33	424±40	180±117	124±22	156±25	260±137
PO	268±32	412±40	191±90	144±24	160±25	265±120
OL	356±37	596±47	276±257	108±21	172±26	491±364

Table 5: Quasi-t losses Cases 3a, 3b, and 4

sessments, and correspondingly, we get different rankings for estimators based on these latter two loss functions.

Table 6 displays the performance of the HCCME with respect to entropy loss. As expected, OLS is best for homegeneity (cases 0a,0b). Quite surprisingly, it also emerges as the top estimator in cases 1 and 2. This must be a feature of the type of heteroskedasticity chosen, since OLS is inconsistent and can be made to have very high losses by selecting the type of heteroskedasticity suitably. Apparently cases 1a, 1b, and 2 do not deviate sufficiently from the OLS case to cause trouble in point estimation of the true covariance. This also shows the great difference between inferences based on the HCCME and the quality of the HCCME as a point estimate of the true covariance. Despite the strong performance of the OLS as a point estimator in these cases, it does poorly in the inference based losses discussed earlier.

After OLS, PO and JS are close, with PO being slightly better. The strong performance of the PO must be correlated with the strong performance of the OLS and cannot be trusted over large and varied patterns of heteroskedasticity. The Shrinkage Estimator emerges as the top reliable point estimator after excluding the inconsistent OLS and PO. Somewhat behind the JS, we have the ML. All other HCCME do quite poorly as point estimators relative to entropy loss.

Table 7 displays the performance of the HCCME relative to the entropy loss in the random coefficient cases. The ML is adapted to be efficient for this

HCCME	CASES				
	0a	0b	1a	1b	2
OL	648±18	653±17	14.0±0.3	13.8±0.3	1818±50
PO	784±25	800±24	14.5±0.3	16.0±0.1	1847±51
JS	704±19	722±19	13.9±0.3	23.3±0.5	2176±50
ML	1094±28	1107±40	24.6±8.6	23.9±1.1	3170±115
BR	2119±30	2157±30	26.5±3.9	42.1±0.6	3637±59
Hi	2276±32	2363±32	28.4±4.1	48.4±0.7	3855±58
HD	2322±33	2424±34	28.9±4.3	49.5±0.7	3875±56
Wu	2369±33	2488±34	29.5±4.4	50.0±0.7	3941±59
MQ	2391±35	2527±37	29.8±4.7	52.3±0.9	3934±61
EW	2372±32	2448±33	29.2±4.1	49.6±0.7	3932±57
Ja	2431±37	2581±39	30.4±4.9	52.7±0.9	3981±79
BC	2482±35	2605±36	31.1±4.5	54.2±0.8	4137±61

Table 6: Entropy Losses in Cases 0, 1 and 2

HCCME	CASE		
	3a	3b	4
ML	1722±135	2098±106	582±511
JS	2847±43	3617±102	741±90
BR	3349±48	3741±104	726±75
OL	2946±31	3896±86	1032±282
PO	3507±47	4308±102	896±80
Hi	3819±55	4465±112	810±89
EW	3970±55	4670±113	839±95
HD	4006±51	4638±137	832±90
Wu	4056±62	4684±120	843±91
MQ	4307±74	5071±135	873±197
Ja	4388±112	5120±162	885±98
BC	4365±67	5115±124	941±198

Table 7: Entropy losses Cases 3a, 3b, and 4

HCCME	CASE 0a	0b	1a	1b	2
OL	51±2	1308±40	59.8±2.0	22.0±0.6	21±0.8
JS	57±2	1472±40	65.2±2.0	42.3±1.4	26±1.1
PO	66±3	1691±60	65.1±2.0	28.4±1.3	21±0.8
ML	103±4	2913±234	136.6±9.4	72.0±10.1	45±3.0
EW	166±3	4335±80	115.8±2.8	69.0±1.6	34±1.0
BR	172±4	4485±90	125.0±2.8	71.3±1.4	39±1.3
Hi	180±4	4783±90	126.6±2.8	75.8±1.9	38±1.3
HD	192±4	5143±110	135.3±4.0	87.2±2.6	39±1.3
Wu	195±4	5263±113	137.7±4.0	87.7±2.5	39±1.2
BC	201±5	5415±120	141.4±4.0	92.1±2.8	40±1.3
MQ	213±5	5883±141	152.3±4.9	112.0±4.0	42±1.4
Ja	229±4	6302±150	162.2±5.3	115.3±2.3	44±1.5

Table 8: Quadratic loss cases 0, 1, and 2

case and regains its top position, with JS close behind. These results confirm the basic intuition that shrinkage estimators are strong with respect to distance criteria for judging point estimates such as entropy and quadratic loss. They do not have nice distributional properties and therefore are weak in the earlier two criteria which assess inference quality based on distribution. Since the OLS and PO do not do well in these random coefficient cases, we confirm our intuition that the performance of these two is strongly design dependent and unreliable over a large range of heteroskedastic behaviors.

11 Quadratic Loss

We now turn to the assessment of the HCC estimates via the quadratic loss function. In general, the picture which emerges is quite similar to the case of entropy loss. Since the two losses measure the performance of the HCCME as point estimators of the true covariance, this similarity is to be expected. As in the case of entropy loss, OLS performs best not only for cases 0a and 0b, but also in the case of structural change (1a and 1b) as well as long tailed errors (2). Indeed, estimator rankings as a whole are fairly stable in all of these cases (0a,0b,1a,1b,2) relative to quadratic loss. OLS, JS, and PO rank at the top in all of these cases. ML is somewhat worse. Other estimators are ranked roughly in the order they are listed, although the differences between any two next to each other in the ranking are slight.

Table 9 displays the results for the random coefficient cases (3a, 3b, and 4).

HCCME	CASE 3a	3b	4
ML	1387±41	25.6±0.9	468±384
OL	1784±16	39.8±0.3	804±186
JS	2070±61	46.8±2.1	708±87
BR	2494±76	44.7±1.8	677±52
PO	2628±73	52.6±1.9	790±59
EW	2678±74	49.8±1.9	715±44
Hi	2873±89	53.1±1.9	769±51
HD	3425±121	66.8±3.0	892±49
Wu	3443±120	66.7±3.3	904±51
BC	3719±131	73.1±3.7	964±57
MQ	4383±169	98.5±5.9	1129±141
Ja	4675±202	97.9±5.7	1179±125

Table 9: Quadratic losses cases 3a, 3b, and 4

These are the domain of natural superiority of the ML and it emerges as the top performer by far in this set of cases. None of the other estimators even come close to the ML. The JS estimator is second rank but viable in this context. Unlike the entropy loss case, the OLS continues to perform well in the random coefficients case.

Based on these simulations, we reach the following conclusions. If heteroskedasticity comes from random coefficients, then we should use the ML based HCCME – this holds for all four loss functions tried. If we cannot be sure about the nature of the heteroskedasticity, and point estimation of the true covariance is of interest, then the JS recommends itself as providing robust performance over the types of heteroskedasticity included in this study.

12 Concluding Remarks

The newly developed ML-based HCCME emerges as a very strong performer in the set of simulations described above. Since the space of possible heteroskedastic sequences is very high dimensional, it is not possible to say that we have covered all possibilities, and not easy to generalize the results of the Monte Carlo study. Nonetheless, it does seem clear that the ML-based HCCME has good performance over a wide range of cases, and does especially well in the random coefficient situations. Surprisingly, Eicker-White and its modifications, especially Horn, Horn, and Duncan, do not do well in many cases. Since these are the most widely used estimates in practice, our results suggest that greater attention should be paid to ML-based HCCME.

Typically the covariance of the OLS is calculated as an intermediate product,*en route* to inference. If for some reason the covariance is directly of interest then the shrinkage estimators introduced are worthy of attention. They have good performance, especially in cases not corresponding to random coefficients. Even considered as a direct estimator, the ML-based HCCME is the best for random coefficient cases.

One point that deserves attention is the computation time of the estimators. The main drawback of the ML is that its computation time is very high. We run simulations adapted from Case 4 to have an idea about the computation time of some of the estimators for which there is a problem. 100 iterations lasted for 22 seconds for ML, 1.7 seconds for JS, 46 seconds for BR, and 18 seconds for Wu, whereas this time was ignorable for the others.

Although we have learned a lot about the relative performance of different types of HCCME estimators, there remain many unanswered questions. Some of these are listed below as being potential topics for future research.

- We have not studied the effects of sample size. It may be possible to get theoretical results in this direction using higher order asymptotics for intermediate to large sample sizes.

- The techniques developed by Chesher and Jewitt can be applied to develop theoretical formulae and/or approximations for the performance of the different estimators. This would be useful in determining types of heteroskedasticity which are favorable and unfavorable for the different estimators.

- An important design factor is the balance (or lack of it) of the design matrix. We have generally chosen fairly balanced designs (where the diagonals of the hat matrix are not too large). Making the design matrix unbalanced could have a big impact on the relative performance of the estimators.

- When inference about OLS is of interest, it may be possible to bypass the estimation of an HCCM and directly focus on the desired inference. This may yield better results than the two stage procedure of estimating an HCCM and then basing inference on it.

References

[1] Chesher, A. and Jewitt, I. *The bias of a heteroskedasticity-consistent covariance matrix estimator* Econometrica, 55, 1217-1222, 1987

[2] Dey, D.K. and Srinivasan, C. *Estimation of a covariance matrix under Stein's loss* The Annals of Statistics, 13, 1581-1591, 1985

[3] Efron, B. *The jacknife, the bootstrap and other resampling plans* Society for Industrial and Applied Mathematics, Philadelphia, PA, 1982

[4] Eicker, F. *Asymptotic normality and consistency of the least squares estimators for families of linear regressions* Annals of Mathematical Statistics, Vol 34, 447-456, 1963

[5] Fang, K. T. and Li, R. Z. *Elliptical Matrix Distributions* Journal of Multivariate Analysis, 70, 66-85, 1999

[6] Furno, M. *Small sample behavior of a robust heteroskedasticity consistent covariance-matrix estimator* Journal of Statistical Computation and Simulation, Vol 54, 115-128, 1996

[7] Furno, M. *A robust heteroskedasticity consistent covariance-matrix estimator* Statistics, Vol 30, 201-219, 1997

[8] Haff, L.R. *Empirical Bayes estimation of the multivariate normal covariance matrix* The Annals of Statistics, Vol 8, 586-597, 1980

[9] Hinkley, D. V. *Jacknifing in unbalanced situations* Technometrics, Vol 19, 285-292, 1977

[10] Horn, S. D, Horn, R. A. and Duncan, D. B. *Estimating heteroskedastic variances in linear models* Journal of the American Statistical Association, Vol 70, 380-385, 1975

[11] James, W. and Stein C. *Estimation with quadratic loss* In Fourth Berkeley Symp. Math. Statist. Prob., Univ. of California Press, 1, 361-380, 1961

[12] MacKinnon, J. G. and White H. *Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties* Journal of Econometrics, Vol 29, 305-325, 1985

[13] Rao, J. R. *Estimation of heteroskedastic variances in linear models* Journal of the American Statistical Association, Vol 65, 161-172, 1970

[14] White, H. *A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity* Econometrica, Vol 48, 817-838, 1980

[15] Wu, C. F. J. *Jacknife, bootstrap and other resampling methods in regression analysis* Annals of Statistics, Vol 14, No 4, 1261-1295, 1986

[16] Zaman, A. *Random coefficients and heteroskedasticity* Technical paper, Bilkent University, 1997