Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model

Author(s): J. Scott Long and Laurie H. Ervin

Source: *The American Statistician*, Aug., 2000, Vol. 54, No. 3 (Aug., 2000), pp. 217-224

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: https://www.jstor.org/stable/2685594

# Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model

J. Scott LONG and Laurie H. ERVIN

In the presence of heteroscedasticity, ordinary least squares (OLS) estimates are unbiased, but the usual tests of significance are generally inappropriate and their use can lead to incorrect inferences. Tests based on a heteroscedasticity consistent covariance matrix (HCCM), however, are consistent even in the presence of heteroscedasticity of an *unknown* form. Most applications that use a HCCM appear to rely on the asymptotic version known as HC0. Our Monte Carlo simulations show that HC0 often results in incorrect inferences when $N \leq 250$, while three relatively unknown, small sample versions of the HCCM, and especially a version known as HC3, work well even for $N$'s as small as 25. We recommend that: (1) data analysts should correct for heteroscedasticity using a HCCM whenever there is reason to suspect heteroscedasticity; (2) the decision to use HCCM-based tests should *not* be determined by a screening test for heteroscedasticity; and (3) when $N \leq 250$, the HCCM known as HC3 should be used. Since HC3 is simple to compute, we encourage authors of statistical software to add this estimator to their programs.

KEY WORDS: Heteroscedasticity; Heteroscedasticity consistent covariance matrix.

## 1. INTRODUCTION

It is well known that when the assumptions of the linear regression model are correct, ordinary least squares (OLS) provides efficient and unbiased estimates of the parameters. Heteroscedasticity occurs when the variance of the errors varies across observations. When the errors are heteroscedastic, the OLS estimator remains unbiased, but becomes inefficient. More importantly, the usual procedures for hypothesis testing are no longer appropriate. Given that heteroscedasticity is common in cross-sectional data, methods that correct for heteroscedasticity are essential for prudent data analysis.

A variety of statistical methods correct for heteroscedasticity by weighting each observation by the inverse of the standard deviation of the error (see, e.g., Greene 2000, pp. 512–515; Carroll and Ruppert 1988, pp. 9–61). The resulting coefficient estimates are efficient and unbiased, with unbiased estimates of the standard errors. When the form and magnitude of heteroscedasticity are known, using weights to correct for heteroscedasticity is very simple using generalized least squares. If the form of heteroscedasticity involves a small number of unknown parameters, the variance of each residual can be estimated first and these estimates can be used as weights in a second step. In many cases, however, the form of heteroscedasticity is unknown, which makes the weighting approach impractical.

When heteroscedasticity is caused by an incorrect functional form, it can be corrected by making variance-stabilizing transformations of the dependent variable (see, e.g., Weisberg 1980, pp. 123–124) or by transforming both sides (Carroll and Ruppert 1988, pp. 115–173). Although these approaches can provide an efficient and elegant solution to the problems caused by heteroscedasticity, when the results need to be interpreted in the original scale of the variables, nonparametric methods may be necessary (Duan 1983; Carroll and Ruppert 1988, pp. 136–139). As noted by Emerson and Stoto (1983, p. 124), "... re-expression moves us into a scale that is often less familiar." Furthermore, if there are theoretical reasons to believe that errors are heteroscedastic around the correct functional form, transforming the dependent variable is inappropriate.

An alternative approach, which is the focus of our article, is to use tests based on a heteroscedasticity consistent covariance matrix, hereafter HCCM. The HCCM provides a consistent estimator of the covariance matrix of the regression coefficients in the presence of heteroscedasticity of an unknown form. This is particularly useful when the interpretation of nonlinear models that reduce heteroscedasticity is difficult, when a suitable variance-stabilizing transformation cannot be found, or when weights cannot be estimated for use in generalized least squares (GLS). Theoretically, use of the HCCM allows a researcher to easily avoid the adverse effects of heteroscedasticity even when nothing is known about the form of heteroscedasticity.

The development of the HCCM can be traced to the early work of Eicker (1963, 1967) and Huber (1967). White

J. Scott Long is Chancellors' Professor of Sociology, Department of Sociology, Ballantine Hall 744, Indiana University, Bloomington, IN 47405 (E-mail: jslong@indiana.edu). Laurie H. Ervin is a Fellow, Social Psychology Traning Program, Department of Sociology, Indiana University, Bloomington, IN 47405. We thank Paul Allison, Ken Bollen, David Drukker, John Fox, Bill Greene, Lowell Hargens, David James, and the reviewers for their comments. Technical details on the simulations can be found at www.indiana.edu/~jsl650/hccm.htm.

(1980) introduced this idea to econometricians and derived the asymptotically justified form of the HCCM known as HC0. In a later article, MacKinnon and White (1985) raised concerns about the performance of HC0 in small samples and presented three alternative estimators known as HC1, HC2, and HC3. While these estimators are asymptotically equivalent to HC0, they were expected to have superior properties in finite samples. Based on limited Monte Carlo evidence, MacKinnon and White (1985) recommended that in small samples one should use HC3. Based on further simulations, Davidson and MacKinnon (1993, p. 554) later recommended strongly that HC2 or HC3 should be used in favor of HC0.

In Section 3 we argue that researchers and software vendors are unaware of, or unconvinced by, the limited evidence regarding the small-sample properties of HC0. Our objective in this article is to provide extensive and, we hope, convincing evidence for the superiority of HC3. While no Monte Carlo simulation can cover all variations that might influence the properties of the statistics being studied, our simulations explore a wide range of situations that are common in cross-sectional data. The next section begins by reviewing the effects of heteroscedasticity and defining four versions of the HCCM for the linear regression model. Section 3 assesses current practice in using HCCMs. Section 4 describes the simulations that are presented in Section 5. Overall, our results indicate that data analysts should change the way in which they use heteroscedasticity consistent standard errors. To this end, software vendors need to make simple changes to their software that could result in substantial improvements in the application of the linear regression model.

## 2. HCCM FOR THE LINEAR REGRESSION MODEL

Using standard notation, the linear regression model can be written as

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

where $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Phi}$, a positive definite matrix. Under this specification, the OLS estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is best linear unbiased with

$$\operatorname{var}\left(\widehat{\boldsymbol{\beta}}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Phi}X(\mathbf{X}'\mathbf{X})^{-1}. \quad (1)$$

If the errors are homoscedastic, that is $\boldsymbol{\Phi} = \sigma^2\mathbf{I}$, Equation (1) simplifies to

$$\operatorname{var}\left(\widehat{\boldsymbol{\beta}}\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \quad (2)$$

Defining the residuals $e_i = y_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}$, where $\mathbf{x}_i$ is the $i$th row of $\mathbf{X}$, we can estimate the OLS covariance matrix (OLSCM) of estimates as

$$\text{OLSCM} = \frac{\sum e_i^2}{N-K}(\mathbf{X}'\mathbf{X})^{-1},$$

where $N$ is the sample size and $K$ is the number of elements in $\boldsymbol{\beta}$. The OLSCM is appropriate for hypothesis testing and computing confidence intervals when the standard assumptions of the regression model, including homoscedasticity, hold. When there is heteroscedasticity, tests based on the OLSCM are likely to be misleading since Equation (2) will not generally equal Equation (1).

If the errors are heteroscedastic and $\boldsymbol{\Phi}$ is known, Equation (1) can be used to correct for heteroscedasticity. More often, the form of heteroscedasticity is unknown and a HCCM should be used. The basic idea behind the HCCM estimator is to use $e_i^2$ to estimate $\phi_{ii}$. This can be thought of as estimating the variance of $\varepsilon_i$ with a single observation: $\widehat{\phi}_{ii} = (e_i - 0)^2/1 = e_i^2$. Then, let $\widehat{\boldsymbol{\Phi}} = \operatorname{diag}[e_i^2]$, which results in

$$\begin{aligned}\text{HC0} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\boldsymbol{\Phi}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\operatorname{diag}\left[e_i^2\right]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \quad (3)\end{aligned}$$

HC0 is the most commonly used form of the HCCM and is referred to variously as the White, Eicker, or Huber estimator. As shown by White (1980) and others, HC0 is a consistent estimator of $\operatorname{var}\left(\widehat{\boldsymbol{\beta}}\right)$ in the presence of heteroscedasticity of an unknown form.

MacKinnon and White (1985) considered three alternative estimators designed to improve the small sample properties of HC0. The simplest adjustment, suggested by Hinkley (1977), makes a degrees of freedom correction that inflates each residual by the factor $\sqrt{N/(N-K)}$. With this correction, we obtain the version known as HC1:

$$\begin{aligned}\text{HC1} &= \frac{N}{N-K}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\operatorname{diag}\left[e_i^2\right]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \frac{N}{N-K}\text{HC0}.\end{aligned}$$

To understand the motivation for the second alternative, we need some basic results from the analysis of outliers and influential observations (Belsley, Kuh, and Welsch 1980, 13–19). Recall that $\widehat{\boldsymbol{\Phi}}$ in Equation (3) is based on the OLS residuals $e$, not the errors $\varepsilon$. Even if the errors are homoscedastic, the residuals will not be. Specifically, if we define $h_{ii} = \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'$, then

$$\operatorname{var}(e_i) = \sigma^2(1 - h_{ii}) \neq \sigma^2. \quad (4)$$

Since $1/N \leq h_{ii} \leq 1$, var($e_i$) underestimates $\sigma^2$. Equation (4) suggests that although $e_i^2$ is a biased estimator of $\sigma_i^2$, $e_i^2/(1-h_{ii})$ will be less biased. This led MacKinnon and White (1985), based on work by Horn, Horn, and Duncan (1975), to propose

$$\text{HC2} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\operatorname{diag}\left[\frac{e_i^2}{1-h_{ii}}\right]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

A third variation approximates a more complicated jackknife estimator of Efron (1982, as cited by MacKinnon and White 1985):

$$\text{HC3} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\operatorname{diag}\left[\frac{e_i^2}{(1-h_{ii})^2}\right]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

Dividing $e_i^2$ by $(1-h_{ii})^2$ further inflates $e_i^2$, which is thought to adjust for the "over-influence" of observations with large variances.

All four HCCM estimators are easy to program since they are functions of statistics routinely computed by standard regression packages. Regardless of the number of observations or variables in the model, HCCM tests should less than double the computing time required. For example, in Stata 6 it took 90% longer to compute HC2 or HC3 compared to OLSCM and 10% longer compared to HC1. In a test version of LIMDEP (Greene 1999, personal communication), it took 66% longer to compute any of the HCCMs compared to OLSCM. Clearly, the added time to compute these estimates is feasible given current computing power.

## 3. CURRENT PRACTICE

An increasing number of applications use HCCMs, especially applications in the social sciences. However, MacKinnon and White's recommendation against using HC0 in small samples appears to be unknown or unconvincing to most researchers and software vendors. Our conclusion is based on several sources of information. First, Table 1 shows that HC0 is the most common form of the HCCM estimated by 12 statistical packages, including a range of general and specialized packages. HC2 and HC3 are available only in Stata and TSP. While all forms of the HCCM can be programmed in packages such as LIMDEP, S-Plus, or GAUSS, it is unlikely that many users will do this. Second, in the 1996 edition of *Social Science Citation Index*, White's (1980) paper which discusses only HC0 was cited 235 times compared to only 8 citations to MacKinnon and White (1985) where HC1, HC2, and HC3 are presented. Third, while many recent texts discuss HC0 (e.g., Amemiya 1994; Fox 1997; Goldberger 1991; Gujarati 1995; Judge et al. 1988; Maddala 1992), we found only two that discuss the small sample properties of HC0 (Davidson and MacKinnon 1993; Greene 2000). Finally, we reviewed 32 of the 240 articles in the 1996 *Social Science Citation Index* that cited either White (1980) or MacKinnon and White (1985). Only one article used a small sample version of the HCCM, even though half of the articles had samples that our simulations suggest are too small to justify the use of HC0. We now turn to those simulations.

*Table 1. Types of heteroscedasticity consistent covariance matrices estimated by statistical packages that estimate the linear regression model*

| Package | Version | HC0 | HC1 | HC2 | HC3 |
|---|---|---|---|---|---|
| BMDP | 7 | ⊘ | ⊘ | ⊘ | ⊘ |
| GAUSS | 3.2 | ■ | □ | □ | □ |
| GLIM | 4 | ⊘ | ⊘ | ⊘ | ⊘ |
| LIMDEP | 7 | ■ | □ | □ | □ |
| Microfit | 4 | ⊘ | ■ | ⊘ | ⊘ |
| Minitab | 11 | ⊘ | ⊘ | ⊘ | ⊘ |
| SAS | 6.11 | ■ | □ | □ | □ |
| SPSS | 10 | ⊘ | ⊘ | ⊘ | ⊘ |
| Systat | 8 | ■ | ⊘ | ⊘ | ⊘ |
| S-Plus | 2000 | □ | □ | □ | □ |
| Stata | 6 | ⊠ | ■ | ⊠ | ⊠ |
| TSP | 4.4 | ⊠ | ⊠ | ■ | ⊠ |

■ = default option; ⊠ = yes; □ = can be programmed; ⊘ = not available.

NOTE: This table is based on our familiarity with these packages as well as a review of documentation, Web sites for the packages, and, in some cases, calls to technical support. We apologize for missing any features of the software. Please see the Appendix for more information on the software packages and documentation used here.

*Table 2. Error structures used in the simulations*

| Structure | Scedasticity function | Average ratio of standard deviations of $\varepsilon$ for 85–95 and 5–15 percentiles |
|---|---|---|
| 0 | $\varepsilon_i = \varepsilon_i^*$ (no heteroscedasticity) | 1.0 |
| 1 | $\varepsilon_i = \sqrt{x_{i1}}\varepsilon_i^*$ | 1.2 |
| 2 | $\varepsilon_i = \sqrt{x_{i3} + 1.6}\varepsilon_i^*$ | 1.3 |
| 3 | $\varepsilon_i = \sqrt{x_{i3}}\sqrt{x_{i4} + 2.5}\varepsilon_i^*$ | 2.2 |
| 4 | $\varepsilon_i = \sqrt{x_{i1}}\sqrt{x_{i2} + 2.5}\sqrt{x_{i3}}\varepsilon_i^*$ | 2.8 |
| 5 | $\varepsilon_i = \begin{cases} 1.5\varepsilon_i^* & \text{if } x_{iD} = 1 \\ \varepsilon_i^* & \text{if } x_{iD} = 0 \end{cases}$ | 1.5 |
| 6 | $\varepsilon_i = \begin{cases} 4\varepsilon_i^* & \text{if } x_{iD} = 1 \\ \varepsilon_i^* & \text{if } x_{iD} = 0 \end{cases}$ | 3.2 |

NOTE: $\varepsilon^*$ has a $z$, $t_5$, or $\chi_5^2$ distribution.

## 4. MONTE CARLO EXPERIMENTS

Monte Carlo simulations were used to examine the small sample behavior of tests using the OLSCM and the four versions of the HCCM presented earlier. Each simulation was based on the model

$$y_i = 1 + 1x_{1i} + 1x_{2i} + 1x_{3i} + 0x_{4i} + \tau\varepsilon_i. \tag{5}$$

Characteristics of the $x$'s and $\varepsilon$'s were varied to simulate data typically found in cross-sectional research. The independent variables had a variety of distributions, including uniform, bell-shaped, skewed (to increase the likelihood of sampling points of high leverage), bimodal, and binary. Correlations among the $x$'s ranged from .2 to .8. The effect of the variance of the errors was examined by running simulations with $R^2$'s ranging from .2 to .7.

Table 2 lists the error structures that were considered. Structure 0 includes three distributions of homoscedastic errors: skewed $(\chi_5^2)$, fat-tailed $(t_5)$, and normal $(z)$. Heteroscedasticity was introduced by allowing the variance of the errors to depend on the independent variables in six ways, corresponding to Structures 1 through 6. As a simple measure of the extent of heteroscedasticity, we sorted the data by each independent variable and computed the ratio of the standard deviations of the errors within the 5th to 15th percentiles of the sorted $x$'s to those within the 85th to 95th percentiles. The upper and lower 5th percentiles were dropped to eliminate the effects of extreme observations. The average ratio across all $x$'s appears in the last column. We also computed the percent of times out of 1,000 replications that the Breusch-Pagan test for heteroscedasticity was statistically significant at $N$ = 25, 50, and 100. These percentages, which are not reported, tracked closely with our simple measures of the ratio of the standard deviations.

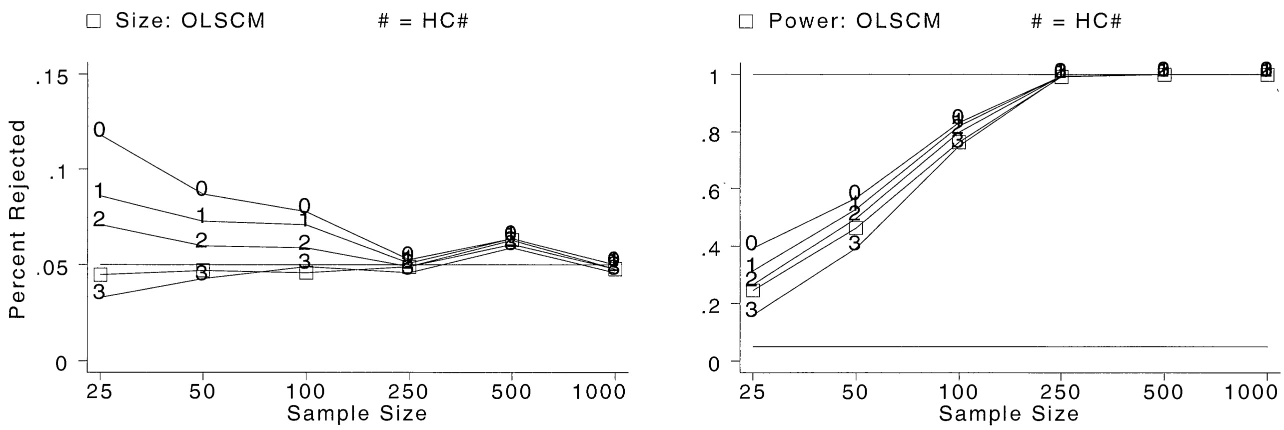


*Figure 1. Size and power of t tests of $\beta_3$ for homoscedastistic $\chi^2_5$ errors.*

For each error structure and combination of types of variables, we ran simulations as follows:

*Data Generation of the Population.* 100,000 observations for the independent variables ($x$'s) were constructed and saved to disk. Random errors $\varepsilon$ were generated according to the error structure being evaluated. These were used to construct the dependent variable $y$ according to Equation (5).

*Simulations.* 1,000 random samples without replacement were drawn for $N = 25$, 50, 100, 250, 500, and 1,000. Regressions were estimated and hypothesis tests were computed for each sample at each sample size. The estimates of the $\beta$'s and $t$ statistics using the OLSCM and the four HCCMs were saved.

*Evaluating Size and Power.* To evaluate size, the null hypothesis was $H_0$: $\beta_k = \beta_k^*$, where $\beta_k^*$ is the population value determined by a regression using all 100,000 observations. We compared the nominal significance level to the proportion of times that the correct $H_0$ was rejected over the 1,000 replications at a given sample size. For power, the empirical significance level is the proportion of times the false hypothesis $H_0$: $\beta_k = 0$ was rejected over 1,000 replications. Power curves for values from two below to two above the population value were also computed. Although size and power were examined at the .05 and .10 nominal levels, the findings were similar so only results for the .05 level will be presented.

These simulations were used to evaluate three situations in which the HCCM might be used. First, we examined the cost of using a HCCM-based test when errors were homoscedastic. Second, we compared OLSCM tests and the HCCM tests when there was heteroscedasticity. Finally, we examined the consequences of using a test for heteroscedasticity to determine whether HCCM tests should be used. For each application, we present a few results that highlight our major findings. Full details, including computer code, are available at www.indiana.edu/~jsl650/hccm.htm

### 4.1 Homoscedastic Errors

First, we consider the consequences of correcting unnecessarily for heteroscedasticity when the errors are homoscedastic. Figure 1 uses results from a population with homoscedastic $\chi^2_5$ errors to illustrate our findings. The horizontal axis indicates the size of the sample used in the simulation; the vertical axis indicates the proportion of times that $H_0$ was rejected out of 1,000 replications. The nominal significance level is indicated by a horizontal line at .05. The proportion of times that the null hypotheses is rejected using tests based on the standard OLSCM is indicated by a □; the proportion rejected by each type of HCCM is indicated by a number: 0 for HC0, 1 for HC1, and so on.

*Size.* The left side of Figure 1 illustrates our key findings regarding size. These findings are:

1. OLSCM-based tests have the best size properties, as would be expected since the errors are homoscedastic.
2. The size properties of HC3 tests are nearly as good as those for OLSCM tests, even at $N = 25$.
3. Tests based on HC2, HC1, and HC0 have increasingly large size distortion for $N \leq 100$.
4. For $N \geq 250$, all tests have nearly identical size properties.

These conclusions were also supported in simulations using homoscedastic errors with $z$ or $t_5$ distributions.

*Power.* The right panel of Figure 1 shows the power of the various tests for the false hypothesis $H_0$: $\beta_3 = 0$. This example illustrates our overall findings from all simulations:

1. HC3 and OLSCM have less power than HC0, HC1, and HC2 until $N \geq 250$.
2. If the power estimates for HC0, HC1, and HC2 are adjusted for the tendency of these tests to over-reject, their power advantage is reduced by about half.
3. Power curves (not shown) for tests that $\beta_k$ has values ranging from two below to two above the population value show that the results in Figure 1 hold for other false hypotheses.

*Summary.* Overall, the greatest size distortion is seen for HC0 with small samples. At $N = 25$, HC0 rejects the true null hypothesis more than twice as often as it should. HC1 cuts the size distortion in half, and HC2 and HC3 have distortion of less than .02. By $N = 100$, the properties of the

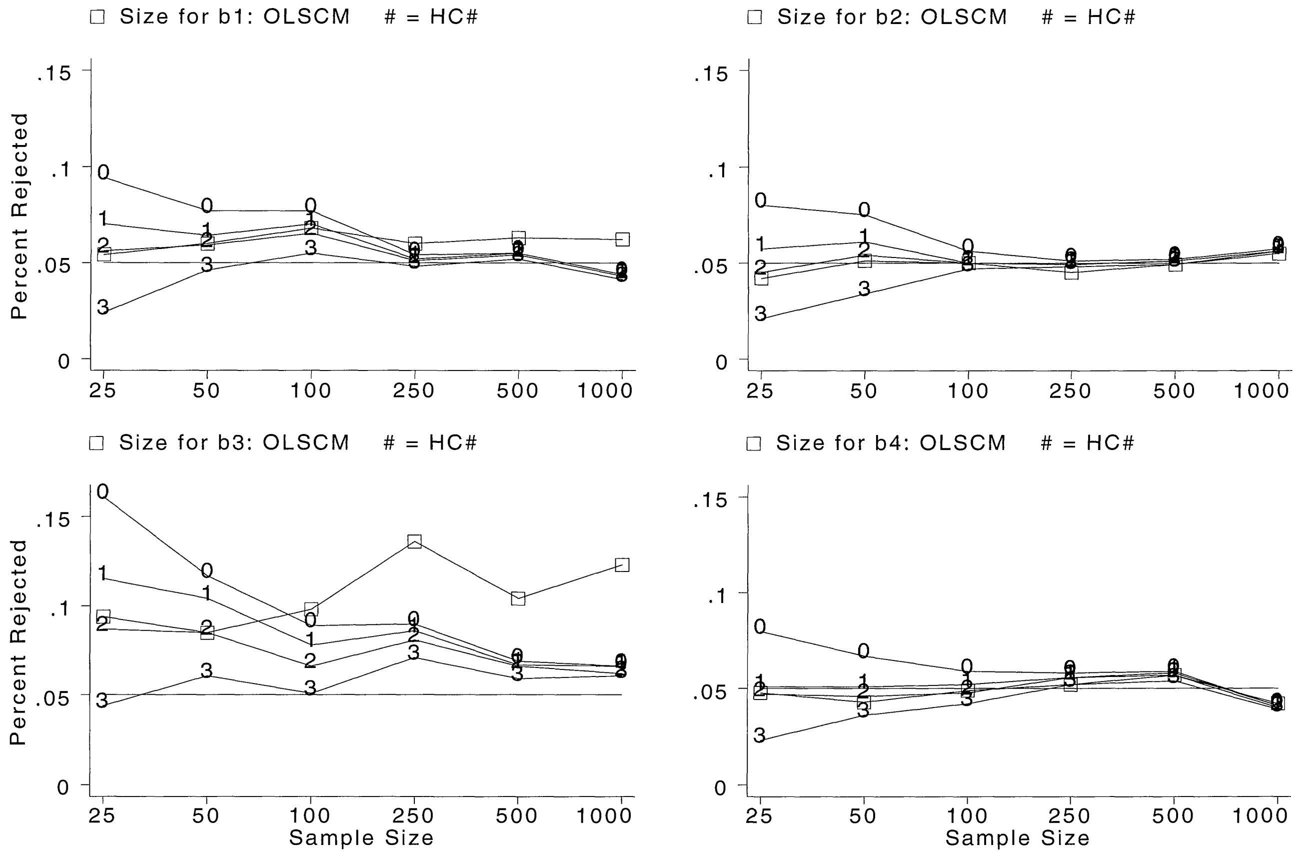


*Figure 2. Size of t test for $\chi^2_5$ errors with heteroscedasticity associated with $x_3$.*

HCCM tests are nearly identical to those using OLSCM, with the exception of HC0. By $N = 1{,}000$, the results from all types of tests are indistinguishable. Thus, for tests with samples of 250 or more, there is very little distortion introduced by using any of the HCCM-based tests when the errors are homoscedastic. For smaller samples, HC3 performs nearly as well as OLSCM.

### 4.2 Heteroscedastic Errors

Although OLSCM tests are superior to the HCCM-based tests when errors are homoscedastic, for the types of heteroscedasticity that we consider, OLSCM tests are biased. Figure 2 plots the size properties of each test when the errors have a $\chi^2_5$ distribution with the scedasticity function $\varepsilon_i = \tau\sqrt{x_{i3} + 1.6}\ \varepsilon^*$. This error structure (Structure 2 in Table 2) has a moderate amount of heteroscedasticity.

*Size.* The four panels of Figure 2 correspond to tests of the four $\beta_k$'s from Equation (5). Although the findings in this figure are for a single error structure, they are representative of the results for other heteroscedastic structures. The key findings are:

1. Heteroscedasticity does not affect tests of each coefficients to the same degree. In general, *HC3 is superior for tests of coefficients that are most affected by heteroscedasticity* (e.g., $\beta_3$). HC2 is somewhat better for tests of coefficients that are least affected by heteroscedasticity (e.g., $\beta_1$).

2. When size distortion is found in OLSCM tests, it does not decrease as the sample size increases. For example, the empirical size of the OLSCM test of $\beta_3$ increases to nearly .15 as the sample increases.

3. For $N \leq 50$, OLSCM tests always do better than HC0-based tests and generally do as well or better than tests using HC1.

By comparing the results from simulations with each of the heteroscedasticity structures described in Table 2, we conclude:

1. With milder forms of heteroscedasticity (e.g., Structures 1 and 2), tests using OLSCM work quite well for all sample sizes.

2. With more extreme forms of heteroscedasticity (e.g., Structures 3 and 4), OLSCM tests have size distortion that increases with sample size.

*Power.* Figure 3 plots power curves for tests of $H_0$: $\beta_1 = 0$ and $H_0$: $\beta_3 = 0$ for heteroscedastic $\chi^2_5$ errors associated with $x_3$ and $x_4$ (Structure 3). This figure illustrates several key results that are also found with other data structures and for testing other hypotheses:

1. OLSCM tests are most powerful, but this is because of the size distortion of these tests. For $N \geq 250$, the size-

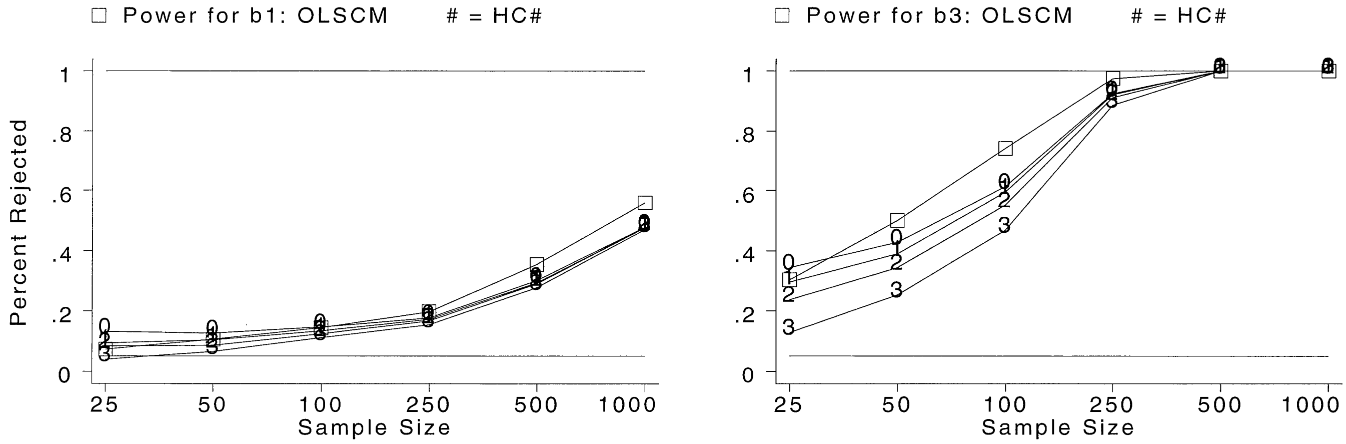


*Figure 3. Power of t test of $\beta_1$ and $\beta_3$ for $\chi_5^2$ errors with heteroscedasticity associated with $x_3$ and $x_4$.*

adjusted power is smaller for the OLSCM tests than for other tests.

2. HC3 tests are the least powerful of the HCCM tests, followed by HC2 and HC1. These differences are largest for tests of $\beta_3$. However, after adjusting the power for size distortion, these differences are greatly reduced.

3. For $N \geq 250$, there are no significant differences in the power of tests based on different forms of the HCCM.

*Summary*. Overall, for $N \geq 500$, there is little difference among tests using different forms of the HCCM. For $N \leq 250$, tests based on HC2 and HC3 perform much better than those using HC0 or HC1. In tests of those coefficients which are most affected by heteroscedasticity, HC3-based tests were almost always superior to those based on HC2, HC1, and HC0. This is the major advantage of HC3.

### 4.3 Screening for Heteroscedasticity

Before making our final recommendations on how the data analyst should correct for heteroscedasticity, we consider the consequences of using a test for heteroscedasticity to determine whether HCCMs should be used. To this end, we ran the following two-step simulation that models the process we have seen frequently in applied papers. First, we computed a White test (White 1980) for heteroscedasticity. (We obtained similar results using the Glejser (1969) and Breusch and Pagan (1979) tests.) Second, if the White test was significant at the .05 level, we used a HCCM-based test; if the White test was not significant, we used the OLSCM test.

Figure 4 shows the effects of screening when there is moderate heteroscedasticity (Structure 3). The left panel shows that the White test has low power to detect heteroscedasticity in small samples. The right panel shows the size properties for a test of $\beta_3$. The □'s plot the results of the standard OLSCM test; the △'s plot results of an HC3 test applied regardless of the outcome of the screening test. The numbers correspond to results from a two-step procedure. For example, 3's plot the results when an HC3 test was used if the White test detected heteroscedasticity, otherwise an OLSCM tests was used. Since the White test has less power in small samples, the two-step process uses the OLSCM test more frequently when $N$ is smaller. Consequently, for small $N$'s the two-step tests have similar size

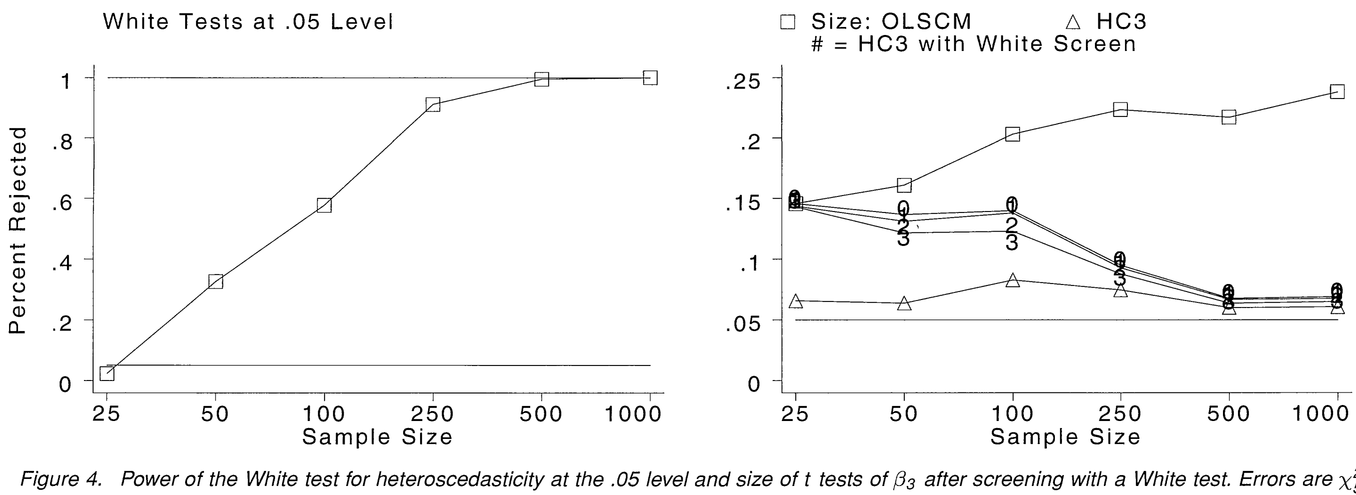


*Figure 4. Power of the White test for heteroscedasticity at the .05 level and size of t tests of $\beta_3$ after screening with a White test. Errors are $\chi_5^2$ with heteroscedasticity associated with $x_3$ and $x_4$.*

properties to the standard OLSCM test. As the power of the screening test increases along with the sample size, the size of the two-step tests converge to those of HC3 tests. The overall conclusion is clear: *a test for heteroscedasticity should not be used to determine whether HCCM-based tests should used.* Far better results are obtained by using HC3 all of the time.

## 5. SUMMARY AND CONCLUSIONS

This article explores the small sample properties of tests using four versions of the HCCM in the linear regression model. Our results lead us to the following conclusions:

1. If there is an a priori reason to suspect that there is heteroscedasticity, HCCM-based tests should be used.
2. For samples less than 250, HC3 should be used; when samples are 500 or larger, other versions of the HCCM can also be used. The superiority of HC3 over HC2 lies in its better properties when testing coefficients that are most strongly affected by heteroscedasticity.
3. The decision to correct for heteroscedasticity should *not* be based on the results of a screening test for heteroscedasticity.

Given the relative costs of correcting for heteroscedasticity using HC3 when there is homoscedasticity and using OLSCM tests when there is heteroscedasticity, we recommend that HC3-based tests should be used routinely for testing individual coefficients in the linear regression model. Although no Monte Carlo study can include all possible situations that can be encountered in practice, the consistency of our results across a wide variety of structures adds credence to our recommendations. The biggest practical obstacle to following our advice is the lack of software that estimates HC3. Even though it is quite simple to program HC3, it is not available in most statistical packages and is the default in only one package. We hope that our results will encourage authors of statistical software to add this estimator to their programs.

## APPENDIX

Following are the software packages and documentation used in this article.

- Aptech Systems, Inc. (1992), *Gauss Version 3.0 Applications: Linear Regression*, Maple Valley, WA: Aptech Systems.
- Dixon, W. J. (ed.) (1992), *BMDP Statistical Software Manual, Version 7*, Berkeley, CA: University of California Press.
- Greene, W. H. (1995), *LIMDEP Version 7.0 User's Manual*. Bellport, NY: Econometric Software, Inc.
- Hall, B. H., Cummins, C., and Schnake, R. (1995), *TSP Reference Manual, Version 4.3*, Palo Alto, CA: TSP International.
- Mathsoft, Inc. (2000), *S-PLUS 2000 User's Guide*, Seattle: Mathsoft.
- Minitab, Inc. (1996), *Minitab Reference Manual, Release 11 for Windows*, State College, PA: Minitab, Inc.
- NAG, Inc. (1993), *GLIM4: The Statistical System for Generalized Linear Interactive Modelling*, Oxford: Clarendon Press.
- Pesaran, M. H., and Pesaran, B. (1997), *Working with Microfit 4.0: Interactive Econometric Analysis*, Cambridge: Camfit Data LTD.
- SAS Institute, Inc. (1989), *SAS/STAT User's Guide: Version 6, Fourth Edition, Volume 2*, Cary, NC: SAS Institute, Inc.
- SPSS, Inc. (1998), *SYSTAT 8.0 Statistics*, Chicago, IL: SPSS, Inc.
- SPSS, Inc. (1999), *SPSS 10 for Windows*, Chicago, IL: SPSS, Inc.
- StataCorp (1999), *Stata Statistical Software: Release 6*, College Station, TX: Stata Corporation.

*[Received September 1998. Revised September 1999.]*

## REFERENCES

Amemiya, T. (1994), *Introduction to Statistics and Econometrics*, Cambridge, MA: Harvard University Press.

Belsley, D.A., Kuh, E., and Welsch, R.E. (1980), *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, New York: Wiley.

Breusch, T.S., and Pagan, A. R. (1979), "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287–1294.

Carroll, R. J., and Ruppert, D. (1988), *Transformation and Weighting in Regression*, New York: Chapman and Hall.

Chesher, A., and Austin, G. (1991), "The Finite-Sample Distributions of Heteroscedasticity Robust Wald Statistics," *Journal of Econometrics*, 47, 153–173.

Davidson, R., and MacKinnon, J. G. (1993), *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Duan, N. (1983), "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association*, 78, 605–610.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia, PA: Society for Industrial and Applied Mathematics.

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

——— (1967), "Limit Theorems for Regressions With Unequal and Dependent Errors," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, pp. 59–82.

Emerson, J. D., and Stoto, M. A. (1983), "Transforming Data," in *Understanding Robust and Explanatory Data Analysis*, eds. D. C. Hoaglin, F. Mosteller, and J. W. Tukey, New York: Wiley, pp. 97–128.

Fox, J. (1997), *Applied Regression Analysis, Linear Models, and Related Methods*, Thousand Oaks, CA: Sage Publications.

Glejser, H. (1969), "A New Test of Heteroskedasticity," *Journal of American Statistical Association*, 64, 314–323.

Goldberger, A. S. (1991), *A Course in Econometrics*, Cambridge, MA: Harvard University Press.

——— (2000), *Econometric Analysis* (4th Ed.), Upper Saddle River, NJ: Prentice Hall.

Gujarati, D. N. (1995), *Basic Econometrics* (3rd ed.), New York: McGraw-Hill.

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

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

Huber, P. J. (1967), "The Behavior of Maximum Likelihood Estimates

Under Non-standard Conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press, pp. 221–233.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H., and Lee, T-C (1988), *Introduction to the Theory and Practice of Econometrics* (2nd ed.), New York: Wiley.

MacKinnon, J. G., and White, H. (1985), "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29, 53–57.

Maddala, G. S. (1992), *Introduction to Econometrics* (2nd ed.), New York: Macmillan.

Weisberg, S. (1980), *Applied Linear Regression*, New York: Wiley.

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