Correcting for Heteroscedasticity with Heteroscedasticity Consistent Standard Errors in the Linear Regression Model: Small Sample Considerations.

J. Scott Long and Laurie H. Ervin
Indiana University
Bloomington, IN 47405

September 23, 1998

Abstract

In the presence of heteroscedasticity, OLS estimates are unbiased, but the usual tests of significance are inconsistent. However, tests based on a heteroscedasticity consistent covariance matrix (HCCM) are consistent. While most applications using a HCCM appear to be based on the *asymptotic* version of the HCCM, there are three additional, relatively unknown, small sample versions of the HCCM that were proposed by MacKinnon and White (1985), based on work by Hinkley (1977), Horn, Horn and Duncan (1975), and Efron (1982). Our objective in this paper is to provide more extensive evidence for the superiority of a version of the HCCM known as HC3. Using Monte Carlo simulations, we show that the most commonly used form of HCCM, known as HC0, results in incorrect inferences in small samples. We recommend that the data analyst should: a) correct for heteroscedasticity using HCCM whenever there is reason to suspect heteroscedasticity; b) the decision to correct for heteroscedasticity should *not* be based on a screening test for heteroscedasticity; and c) if the sample is less than 250, a small sample version of the HCCM known as HC3 should be used.

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. If the errors are heteroscedastic, the OLS estimator remains unbiased, but becomes inefficient. More importantly, estimates of the standard errors are inconsistent. The estimated standard errors can be either too large or too small, in either case resulting in incorrect inferences. Given that heteroscedasticity is a common problem in cross-sectional data analysis, methods that correct for heteroscedasticity are important for prudent data analysis.

Standard econometrics texts, such as Judge et al. (1985:422-445), consider a variety of methods that can be used when the form and magnitude is *known* or can be estimated. Essentially, these methods involve weighting each observation by the inverse of the standard deviation of the error for that observation. The resulting coefficient estimates are efficient and unbiased, with unbiased estimates of the standard errors of the coefficients. Unfortunately, the form of heteroscedasticity is rarely known, which makes this solution generally impractical.

When the form of heteroscedasticity is *unknown*, the heteroscedasticity consistent covariance matrix, hereafter HCCM, provides a consistent estimator of the covariance matrix of the slope coefficients in the presence of heteroscedasticity. Theoretically, the use of HCCM allows a researcher to avoid the adverse effects of heteroscedasticity on hypothesis testing even when nothing is known about the form of the heteroscedasticity. This powerful result, which was introduced to econometricians with White's (1980) classic paper, can be traced to the work of Eicker (1963, 1967), Huber (1967), Hartley, Rao and Keifer (1969), Hinkley (1977), and Horn, Horn, and Duncan (1975).

White's (1980) paper presented the asymptotically justified form of the HCCM, referred to hereafter as HC0. In a later paper, 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. To assess the small sample behavior of these alternatives, MacKinnon and White performed Monte Carlo simulations and concluded by recommending that HC3 should be used. MacKinnon and White designed their simulations to keep the $\mathbf{X}'\mathbf{X}/N$ matrix constant in *all* replications, regardless of sample size. While this had the advantage of eliminating one source of variation that might affect the results, subsequent work by Chesher and colleagues (Chesher and Jewitt 1987; Chesher 1989; Chesher and Austin 1991) demonstrated that characteristics of the design matrix critically affect the properties of HCCMs. Chesher and Austin (1991) showed that the data used by MacKinnon and Davidson had one observation with high leverage. When this observation was removed and the simulations were repeated, they found that *all* versions of the HCCM performed well.

As shown in Section 2, researchers and software vendors are either unaware about concerns with the small sample properties of HC0 or are not convinced by the Monte Carlo evidence that has been provided. Our objective in this paper is to provide far more extensive and, hopefully, convincing evidence for the superiority of a computationally simple form of HC3. While no Monte Carlo simulation can cover all variations that might influence the properties of the statistic being studied, our simulations are designed to enhance our understanding in several important ways. First, by sampling from a population of independent variables, we allow substantial variation among samples in the presence of points of high leverage. Second, we have included a greater range of error structures, including skewed and fat-tailed errors, and additional types of heteroscedasticity not considered in earlier simulations. Such errors structures are likely to be common in cross-sectional research. Third, rather than using the computationally more demanding jackknife estimator (HC3) suggested by MacKinnon and White, we consider the properties of a computationally simpler approximation suggested by Davidson and MacKinnon (1993:554). Fourth, our simulations include results with samples that range from 25 through 1000. And finally, we provide Monte Carlo evidence regarding the effects of using a test of heteroscedasticity to determine whether an HCCM correction should be used. Our conclusions suggest that data analyst should change the way in which they use heteroscedasticity consistent standard errors to correct for heteroscedasticity.

We begin in Section 2 by assessing current practice in using HCCMs. Section 3 reviews the effects of heteroscedasticity and presents four versions of the HCCM. Section 4 describes our Monte Carlo simulations, and Section 5 presents the results of our simulations. We conclude by making recommendations for how the HCCM should be used.

2 Assessing Current Practice

MacKinnon and White's recommendation against using HC0 in small samples is either unknown or the justification is unconvincing to most researchers and software vendors. Our conclusion is based on several sources of information.

First, White's (1980) original paper presenting HC0 is highly cited, while the paper by MacKinnon and White (1985) that presents small sample versions receives few citations. For papers published in 1996, *Social Science Citation* Index lists 235 citations to White (1980). Only eight citations were made to MacKinnon and White (1985), with six of these in methodological papers.

Second, we found only two statistics texts that discuss the small sample properties of HC0. The first is MacKinnon and Davidson (1993:554) which strongly recommends against using HC0 ("As a practical matter, one should never use [it]...") and suggest either HC2 or HC3. The second is Greene (1997), which dismisses MacKinnon and Davidson's advice as too strong. Other recent texts that discuss HC0 neither mention the small sample problems with HC0 nor discuss the alternative forms. These include: Amemiya 1994; Fox 1997; Goldberger 1991; Greene 1990, 1993; Gujarati 1995; Judge et al. 1985; and Maddala 1992.

Third, we reviewed 11 statistical packages, including a range of general and specialized packages. For each package, we examined the documentation, the on-line help, information on the vendor's Web site, and sometimes contacted technical support. To verify which versions of the HCCM were included, we compared the output from each package to known results. Our findings are summarized in Table 1. HC0

Package	Version	HC0	HC1	HC2	HC3
BMDP	7	N	N	N	N
GAUSS	3.2	D	P	P	P
GLIM	4	N	N	N	N
LIMDEP	7	D	P	P	P
Microfit	4	N	D	N	N
Minitab	11	N	N	N	N
SAS	6.11	D	P	P	P
SPSS	7.5	N	N	N	N
Systat	7	N	N	N	N
Stata	4	D	Y	Y	Y
Stata	5	Y	D	Y	Y
TSP	4.4	Y	Y	D	Y

Note: Y=yes; D=default option; N=not available; P=can be programmed with matrix language.

Table 1: Types of Heteroscedasticity Consistent Covariance Matrices Estimated by Statistical Packages that Estimate the Linear Regression Model.

is the most common and often the only form of the HCCM estimated. Significantly, the popular SPSS has no correction, and SAS includes only HC0. HC2 and HC3 were available without extra programming only in Stata and TSP. Still, TSP made HC2 the default, while Stata used HC1. HC2 and HC3 could be programmed using matrix capabilities in GLIM, GAUSS, SAS and LIMDEP.[1] Overall, even though it is quite simple to program HC2 and HC3, these methods are not available in most statistical packages, and are the default in only one, fairly specialized package.

[1]LIMDEP includes sample matrix programs for HC2 and HC3. When we tested these programs, we found that the resulting estimates agreed with those from Stata and TSP only to 2 decimal places. This suggests that users who write their own programs for HC2 and HC3 need to be careful that there is not a loss of precision due to rounding error.

Fourth, we used *Social Science Citation Index* to find all 1996 articles in the social sciences that cited either White (1980) or MacKinnon and White (1985). Since the majority of the 240 articles in the list were in business and economics, we selected all of the articles from psychology, political science, and sociology, plus a small number of articles from economics. While our list does not include articles that used HCCM tests but cited other sources, such as a textbook in econometrics, it should be representative of applications of HCCMs. Of the 32 articles reviewed, 26 used HC0, one article used HC2, four articles examined results using HCCM and decided to use tests based on the standard OLS covariance matrix, and in one article the method could not be determined. Based on the conclusions we present in Section 5, nearly half of the papers using HC0 had samples that are small enough to lead to significant size distortion.[2]

These results support our conclusion that current practice in correcting for heteroscedasticity with HCCMs relies heavily on HC0, and consequently may be seriously flawed.

3 The Linear Regression Model

The linear regression model for the ith observation can be written as

$$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \cdots + \beta_K x_{iK} + \varepsilon_i$$

[2] Of the 26 articles using HC0, 15 percent had samples less than 50; 27 percent samples less than 100; 35 percent samples less than 120; and 46 percent had samples less than 250; and 54 percent had samples less than 500.

where y is the dependent variable, the x's are independent variables, and ε is the error in equation. β_2 through β_K are parameters that indicate the effect of a given x on y; β_1 is the intercept. In matrix notation:

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

where $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are $N \times 1$ matrices, $\mathbf{X}$ is $N \times K$, and $\boldsymbol{\beta}$ is $K \times 1$. For the ith row of $\mathbf{X}$, we can write

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$$

The following assumptions complete the model:

1. *Linearity*: y is linearly related to the x's through the β parameters.

2. *Collinearity*: The x's are not linearly dependent.

3. *Expectation of* ε: $E\left(\varepsilon_i \mid \mathbf{x}_i\right) = 0$ for all i.

4. *Homoscedasticity*: For a given $\mathbf{x}_i$, the errors have a constant variance: $Var(\varepsilon_i \mid \mathbf{x}_i) = \sigma^2$ for all i.

5. *Uncorrelated errors*: For two observations i and j, the covariance between ε_i and ε_j is zero.

With these assumptions, the OLS estimator $\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ has the covariance matrix:

$$Var\left(\widehat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Phi}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \qquad (1)$$

where $\mathbf{\Phi}$ is a diagonal matrix with $\phi_{ii} = Var(\varepsilon_i)$. When the errors are homoscedastic, $\mathbf{\Phi}$ can be written as $\mathbf{\Phi} = \sigma^2\mathbf{I}$. With this assumption, Equation 1 can be simplified:

$$\begin{aligned} Var\left(\widehat{\boldsymbol{\beta}}\right) &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\sigma^2\mathbf{I}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \end{aligned} \tag{2}$$

Defining the residuals $e_i = y_i - \mathbf{x}_i\widehat{\boldsymbol{\beta}}$, we can estimate the usual OLS covariance matrix, hereafter referred to as OLSCM, as:

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

If the errors are heteroscedastistic, then OLSCM is biased and the usual tests of statistical significance are inconsistent. When $\mathbf{\Phi}$ is known, Equation 1 can be used to correct for heteroscedasticity. Unfortunately, the form of heteroscedasticity is rarely known. When $\mathbf{\Phi}$ is unknown, we need a consistent estimator of $\mathbf{\Phi}$ in order to apply Equation 1. The HCCM is based on the idea that e_i^2 can be used to estimate ϕ_{ii}. This can be thought of as estimating the variance of the error using a single observation: $\widehat{\phi}_{ii} = (e_i - 0)^2/1 = e_i^2$. Then, let $\widehat{\mathbf{\Phi}} = \text{diag}[e_i^2]$, which results in the estimator:

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

This estimator is referred to variously as the White, Eicker, or Huber estimator. It is also called the sandwich estimator since $\mathbf{X}'\widehat{\mathbf{\Phi}}\mathbf{X}$ is "sandwiched" between a pair of $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$'s. As shown by White (1980) and others, HC0 is a consistent estimator of $Var\left(\widehat{\boldsymbol{\beta}}\right)$ in the presence of heteroscedasticity of an unknown form.

MacKinnon and White (1985) consider alternative estimators designed to improve on the small sample properties of HC0. To understand the motivation for these alternatives, we need to consider basic results from the analysis of outliers and influential observations (see, for example, Belsley et al. 1980:13-19; Greene 1997:444-445). Recall that $\widehat{\mathbf{\Phi}}$ in Equation 3 is based on the OLS residuals e, not the errors ε. Even if the errors ε are homoscedastistic, the residuals e will not be. Specifically, if we define $h_{ii} = \mathbf{x}_i \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_i'$, then:

$$Var\left(e_i\right) = \sigma^2 \left(1 - h_{ii}\right) \neq \sigma^2 \tag{4}$$

Since $0 \leq h_{ii} \leq 1$ (Belsley et al. 1980:13-19), $Var(e_i)$ will underestimate σ^2. Further, since $Var(e_i)$ varies across observations when the errors are homoscedastistic, the OLS residuals must be heteroscedastistic.

With these results in hand, it is simple to motivate three variations on HC0. The simplest adjustment, suggested by Hinkley (1977), makes a degrees of freedom correction that inflates each residual by a factor of $\sqrt{N/\left(N-K\right)}$, where K is the number of β's. With this correction, we obtain the version of the HCCM known as HC1:

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

The second variation is based on the idea that while e_i^2 is a biased estimator of σ_i^2, Equation 4 suggests that $e_i^2/\left(1 - h_{ii}\right)$ will be a less biased estimator. This led MacKinnon and White (1985), based on work by Horn, Horn and Duncan (1975), to suggest the estimator:

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

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

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

Since $0 \leq h_{ii} \leq 1$, dividing e_i^2 by $(1-h_{ii})^2$ further inflates e_i^2.

4 Monte Carlo Experiments

We used Monte Carlo simulations to examine the small sample behavior of OLSCM tests and tests based on the four versions of the HCCM presented above. Our experiments simulate a variety of data and error structures that are likely to be encountered in cross-sectional research. To this end, we considered errors that were skewed and fat-tailed, as well as normal errors. The effects of the variance of the errors was examined by considering simulations with R^2's ranging from .2 to .7. The independent variables were constructed with a variety of distributions, including uniform, bell-shaped, skewed, and binary. The effects of the scale of the independent variable were examined by also running simulations using variables normalized to a unit variance. Finally, a variety of different forms and degrees of heteroscedasticity were considered. Each simulation involved the following steps:

1. *Independent variables*: 100,000 observations for five independent variables were constructed and saved to disk. The independent variables were constructed to include a variety of distributions.

2. *Errors*: A variety of error structures were chosen to represent common types of homoscedasticity and heteroscedasticity. 100,000 observations were generated for each error type.

3. *Dependent variables*: The dependent variable was constructed as a linear combination of three of the five independent variables plus the error term. The combination of the independent variables, the error, and the dependent variable made up the population for each structure.

4. *Simulations*: From each population, a random sample without replacement was drawn. Since a different random sample is used for each replication, the design matrix will vary. Regressions were estimated and tests of hypotheses were computed for each sample. This was done 1,000 times each for sample sizes of 25, 50, 100, 250, 500, and $1,000$.

5. *Summary*: The results were summarized across the 1,000 replications for each sample size from each population.

Details of our simulations are now given.

4.1 Data Structures

The first step was to generate $100,000$ observations for five independent random variables with the following distributions:

$$\delta_1 \sim \text{Uniform}\,(0,1) \qquad \delta_2 \sim \text{Normal}\,(0,1)$$

$$\delta_3 \sim \chi^2_{\text{1df}} \qquad \delta_4 \sim \text{Normal}\,(0,1)$$

$$\delta_5 \sim \text{Uniform}\,(0,1)$$

The δ's were combined to construct the four independent variables:

$$x_1 = 1 + \delta_1 \qquad x_2 = 3\delta_1 + .6\delta_2$$

$$x_3 = 2\delta_1 + .6\delta_3 \qquad x_4 = .1\delta_1 + .9\delta_3 - .8\delta_4 + 4\delta_5$$

A dummy variable x_D was then created by splitting x_2:

$$x_D = \begin{cases} 1 \text{ if } x_2 > 1.6 \\ 0 \text{ if } x_2 \leq 1.6 \end{cases}$$

The five x's have the following characteristics:

x	Distribution	Minimum	Maximum	μ	σ
x_1	Uniform.	1.00	2.00	1.50	0.29
x_2	Bell-shaped.	-2.24	5.10	1.50	1.05
x_3	Skewed.	0.00	12.13	1.60	1.03
x_4	Bell-shaped.	-2.28	15.28	3.59	1.69
x_D	Binary	0	1	0.47	0.50

Note that x_3 is skewed, which should increase the likelihood that a sample will include points of moderate to high leverage.

Since the same δ's were combined with different weights to create the x's, there are varying degrees of collinearity among the variables. Specifically, the x's are correlated as follows:

r	x_1	x_2	x_3	x_4	x_D
x_1	1.00				
x_2	0.82	1.00			
x_3	0.56	0.46	1.00		
x_4	0.32	0.27	0.56	1.00	
x_D	0.73	0.82	0.41	0.23	1.00

In addition, standardized variables x_1^s through x_4^s, with mean 0 and variance 1, were constructed from the corresponding x_k's.

A variety of error structures were used to represent different types of homoscedasticity and heteroscedasticity that might be found in real data. First, we created three types of homoscedastistic errors:

$$\begin{aligned} \varepsilon_{Ni} &\sim \text{Normal}\,(0,1) \\ \varepsilon_{Xi} &\sim \chi^2_{5\text{df}} \\ \varepsilon_{ti} &\sim t_{5\text{df}} \end{aligned}$$

ε_{Xi} is slightly skewed, with skewness and excess kurtosis parameters of $\sqrt{8/5}$ and 12/5. ε_{ti} is not skewed, but has excess kurtosis of 6 (i.e., "fat" tails). These errors were combined with the independent variables using:

$$\text{Model 1:} \quad y_i = 1 + 1x_{1i} + 1x_{2i} + 1x_{3i} + 0x_{4i} + \tau\varepsilon_i$$

x_4 is included with a coefficient of 0 to examine significance tests in the presence of inclusion error. The constant τ is used to determine the R^2 for the model.

Heteroscedasticity was introduced by allowing the variance of the errors to depend on the independent variables as shown in Table 2. This resulted in 11 heteroscedastic error structures which were represent various degrees and types of heteroscedasticity that might be found in practice. The first column of the table gives a simple mnemonic for referring to each structure. Within the mnemonic, the numbers refer to the x_k's that are linked to the error variance; D refers to a link to x_D; N refers to normally distributed errors, and X refers to χ^2 errors. For example, 34N indicates normally distributed heteroscedastistic errors where the skedasticity function includes x_3 and x_4. Since early experiments showed little differences when t-distributed errors were used, results from these simulations are not reported.

Error Structure	Error Distribution	Skedasticity Function	Structural Model
1N	Normal	$\varepsilon_i = \sqrt{x_{i1}}\varepsilon_{Ni}$	1
1X	χ^2_5	$\varepsilon_i = \sqrt{x_{i1}}\varepsilon_{Xi}$	1
3N	Normal	$\varepsilon_i = \sqrt{x_{i3} + 1.6}\varepsilon_{Ni}$	1
3X	χ^2_5	$\varepsilon_i = \sqrt{x_{i3} + 1.6}\varepsilon_{Xi}$	1
3Xeq	χ^2_5	$\varepsilon_i = \sqrt{x_{i3} + 1.6}\varepsilon_{Xi}$	3
34N	Normal	$\varepsilon_i = \sqrt{x_{i3}}\sqrt{x_{i4} + 2.5}\varepsilon_{Ni}$	1
34X	χ^2_5	$\varepsilon_i = \sqrt{x_{i3}}\sqrt{x_{i4} + 2.5}\varepsilon_{Xi}$	1
123N	Normal	$\varepsilon_i = \sqrt{x_{i1}}\sqrt{x_{i2} + 2.5}\sqrt{x_{i3}}\varepsilon_{Ni}$	1
123X	χ^2_5	$\varepsilon_i = \sqrt{x_{i1}}\sqrt{x_{i2} + 2.5}\sqrt{x_{i3}}\varepsilon_{Xi}$	1
DsmN	Normal	$\varepsilon_i = \begin{cases} 1.5\varepsilon_{Ni} & \text{if } x_{iD} = 1 \\ \varepsilon_{Ni} & \text{if } x_{iD} = 0 \end{cases}$	2
DbigN	Normal	$\varepsilon_i = \begin{cases} 4\varepsilon_{Ni} & \text{if } x_{iD} = 1 \\ \varepsilon_{Ni} & \text{if } x_{iD} = 0 \end{cases}$	2

Table 2: Forms of Heteroscedasticity Used in the Simulations.

In most cases, the errors were combined with the independent variables using Model 1. However, in some cases, as noted in Table 2, two additional structural models were used:

$$\text{Model 2:} \quad y_i = 1 + 1x_{1i} + 1x_{Di} + 1x_{3i} + 0x_{4i} + \tau\varepsilon_i$$

$$\text{Model 3:} \quad y_i = 1 + 1x^s_{1i} + 1x^s_{2i} + 1x^s_{3i} + 0x^s_{4i} + \tau\varepsilon_i$$

Model 2 was used to determine if the results were affected by the inclusion of a dummy independent variable. Model 3 was used to assess the effect of the variance of the independent variables. To anticipate our findings, these two variations did not affect our overall conclusions. For each combination of structural model and error, we chose τ so that the R^2 in the population was approximately .4. While simulations with R^2's ranging from .2 to .7 were run, the findings were so similar that we present only results for $R^2 = .4$.

Table 3 reports results of two methods of assessing the degree of heteroscedasticity. First, we sorted the 100,000 observation in the population by a given x_k, and then computed the ratio of the standard deviation of the errors in the $85th$ to $95th$ percentile to the standard deviation of the errors in the $5th$ to $15th$ quantile. We excluded the upper and lower $5th$ percentiles to eliminate the effects of extreme observations. The larger the ratio, the more heteroscedasticity. This measure was computed after sorting by each of the independent variables. In some cases, such as error structure 34N, there is substantial evidence for heteroscedasticity when the data is sorted on x_3, but much less evidence with respect to the other x's. Second, we computed the percent of times out of 1,000 replications that the Breusch-Pagan test for heteroscedasticity was statistically significant at samples sizes of 25, 50 and 100. The larger the percentage, the greater the evidence for heteroscedasticity. Notice that the Breusch-Pagan test is more likely to detect heteroscedasticity when the errors have a χ^2 distribution. As noted by Long and Trivedi (1992), the Breusch-Pagan test is sensitive to departures from normality as well as departures from homoscedasticity.

5 Monte Carlo Results

Each method of computing the covariance matrix is assessed by calculating the empirical size and power of t-tests for the β parameters. For size, we compare the nominal significance level to the proportion of times that the correct H_0 is rejected over the 1,000 replications at a given sample size. The true hypothesis is H_0: $\beta_k = \beta_k^*$, where

Error Structure	Ratio of Standard Deviations for the 85-95 Percentile to the 5-15 Percentile x_1	x_2 or x_D	x_3	x_4	Average Ratio	% Rejected by Breusch-Pagan Test at .05 Level $N = 25$	$N = 50$	$N = 100$
1N	1.3	1.2	1.2	1.1	1.2	.08	.16	.22
1X	1.3	1.2	1.2	1.1	1.2	.20	.33	.41
3N	1.3	1.2	1.5	1.2	1.3	.10	.23	.42
3X	1.3	1.2	1.5	1.3	1.3	.20	.37	.51
34N	1.9	1.6	3.0	2.2	2.2	.34	.75	.97
34X	1.9	1.6	3.1	2.2	2.2	.39	.77	.97
123N	3.1	2.7	3.7	1.7	2.8	.43	.87	.99
123X	3.1	2.7	3.8	1.8	2.8	.50	.86	.99
DsmN	1.5	1.9	1.3	1.2	1.5	.04	.12	.32
DbigN	3.6	5.1	2.6	1.4	3.2	.31	.87	1.00

Table 3: Degree of Heteroscedasticity for Ten Error Structures.

β_k^* is the population value determined from the regression based on the entire 100,000 observations. For power, the empirical significance level is the proportion of times a false hypothesis was rejected over 1,000 replications at a given sample size. For β_1, β_2, and β_3, we assess power by testing the false hypothesis: H_0: $\beta_k = \#$. For the tables and figures below, we consider H_0: $\beta_k = 0$. Power curves for values from two below to two above the population value were also computed and are summarized where appropriate. While size and power were examined at the .05 and .10 nominal levels, the findings were similar so only results for the .05 level are presented.

5.1 Results for Homoscedastistic Errors

Figure 1 illustrates the problems associated with using heteroscedasticity adjustments with *homoscedastic* χ^2_5 errors. The horizontal axis indicates the size of the sample

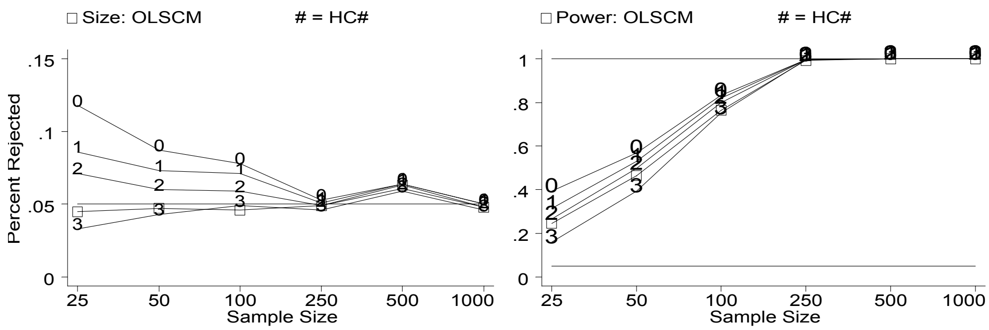

Figure 1: Size and Power of t-tests of β_3 for Homoscedastistic χ^2_5 Errors.

used in the simulation; the vertical axis indicates the proportion of times that the true null hypothesis H_0: $\beta_3 = \beta_3^*$ 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 are indicated by □'s; the proportion rejected by each type of HCCM are indicated by numbers: 0 for HC0, 1 for HC1, and so on.

There are several key results regarding size that also hold for other forms of homoscedasticity, as discussed below):

1. Tests based on OLSCM have the best size properties, as would be expected given that the errors are homoscedastistic. Note that this holds for homoscedastistic errors that are not normal.

2. The size properties of tests based on HC3 are nearly as good as those for OLSCM tests, even at $N = 25$.

3. Tests based on HC0, HC1 and HC2 show significant size distortion for $N \leq 100$,

with distortion increasing from HC2 to HC1 to HC0. All three version have size properties very similar to OLSCM tests for $N \geq 250$.

The right panel of Figure 1 shows the power of the various tests for the false hypothesis H_0: $\beta_3 = 0$. OLSCM and HC3 have similar power, but their ability to reject a false hypothesis is somewhat less than tests based on HC0, HC1 and HC2. However, if the power for HC0, HC1 and HC2 is adjusted for their tendency to over-reject (as shown in the size results), their power advantage is reduced by about half. Power curves (not shown) for tests that β_3 has values ranging from two below to two above the true value show that the results in Figure 1 hold for other hypotheses.

Table 4 summarizes the size results from the homoscedastistic data structures that we examined. The three panels correspond to χ^2_5, normal, and t_5 distributed errors. Within each panel, the rows indicate the type of covariance matrix used, while the columns indicate the sample size used in the simulations. The cells of the table were constructed as follows. For each combination of method, sample size, and error distribution, the nominal alpha of .05 was subtracted from the percent of times out of 1,000 replications that the true hypothesis was rejected. The mean deviation across the four β's is given in the table. For example, .051 (shown in bold) for HC0 at $N = 25$ with χ^2_5 errors indicates that on average tests based on HC0 rejected the true hypothesis 51 times more per $1,000$ tests than indicated by the nominal significance level.

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

	Sample Size					
Panel 1: χ^2_5	25	50	100	250	500	$1,000$
OLSCM	-.005	-.006	-.006	-.001	.008	.002
HC0	**.051**	.022	.009	.003	.011	.003
HC1	.023	.011	.003	.002	.010	.002
HC2	.014	.005	-.002	-.001	.009	.001
HC3	-.019	-.010	-.009	-.004	.008	.000
Panel 2: N	25	50	100	250	500	$1,000$
OLSCM	.004	.002	.000	.004	-.003	-.002
HC0	.057	.031	.016	.011	.001	.000
HC1	.027	.020	.009	.009	.000	-.001
HC2	.019	.014	.005	.007	-.002	-.001
HC3	-.013	-.003	-.002	.004	-.003	-.001
Panel 3: t_5	25	50	100	250	500	$1,000$
OLSCM	-.001	-.003	-.001	.002	-.003	-.005
HC0	.058	.022	.013	.005	.001	-.004
HC1	.025	.008	.006	.004	-.001	-.005
HC2	.015	.003	.003	.002	-.001	-.006
HC3	-.018	-.011	-.006	-.001	-.002	-.007

Table 4: Mean Difference between Nominal and Empirical Size for Homoscedastistic Errors.

the properties of the 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 nearly 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 homoscedastistic.

5.2 Results for Heteroscedastistic Errors

While OLSCM is superior to the HCCM based tests when errors are homoscedastis-

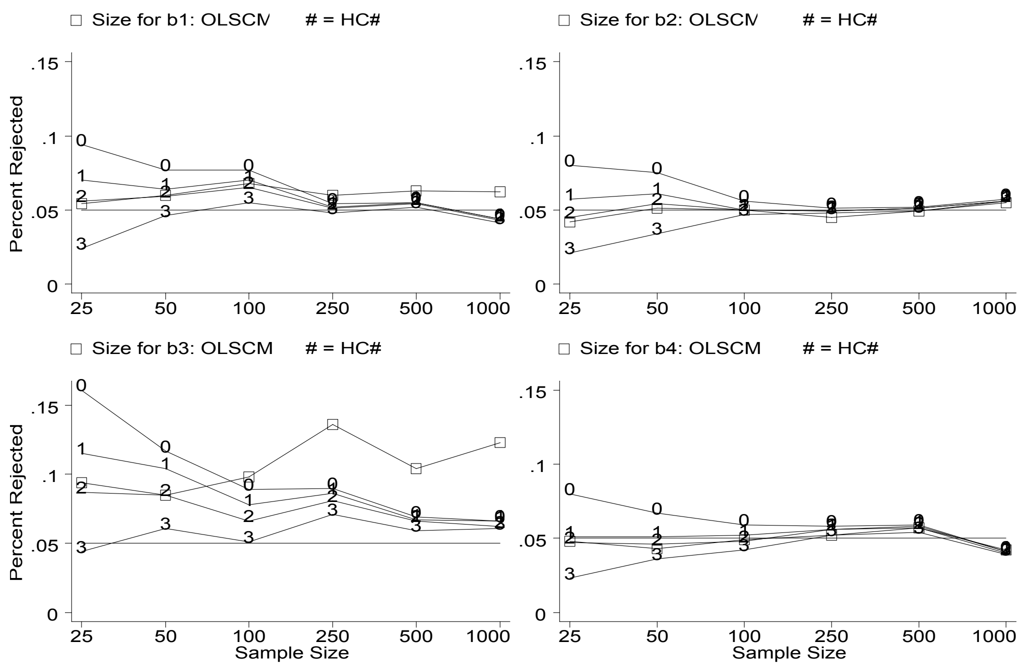

Figure 2: Size of t-test for χ^2_5 Errors with Heteroscedasticity Associated with x_3.

tic, in the presence of heteroscedasticity OLSCM tests are biased. This is illustrated in Figure 2 which shows the results of testing a true hypothesis when the errors are χ^2_5 with the scedasticity function: $\varepsilon_i = \tau\sqrt{x_{i3} + 1.6}\,\varepsilon_{Xi}$. This error structure, referred to as 3X in Table 2, has a moderate amount of heteroscedasticity that is associated most strongly with x_3.

The four panels of Figure 2 correspond to tests of the four β_k's. While the findings in this figure are for a single error structure, they are representative of the results for other heteroscedastistic structures as shown later in the paper. The key points are:

1. The degree to which OLSCM tests are biased varies across parameters. Thus, the OLSCM tests show a great deal of distortion with respect to β_3, limited

distortion for β_1, and no problems for β_2 and β_4.

2. When size distortion is found for OLSCM tests, it does not decrease as the sample size increases. For example, the empirical size for the OLSCM test of β_3 increases to nearly .15 as the sample increases.

3. For $N \leq 100$, OLSCM tests do as well or better than HC0 and HC1 tests.

4. HC3 is superior for tests of coefficients that are most affected by heteroscedasticity (e.g., β_3).

Table 5 summarizes the size results for seven error structures. In Panels 1 through 5 the corresponding results for normal errors are very similar to those presented for χ^2_5 errors, and have not been shown. Again, each cell in the table contains the deviation from the nominal significance level .05 averaged across the four β's. We conclude:

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

2. With milder forms of heteroscedasticity (e.g., 1X, 3X, DsmN), tests using OLSCM work quite well for all sample sizes.

3. With more extreme forms of heteroscedasticity, OLSCM tests have size distortion that increases with sample size (e.g., 3Xeq, 34X, 123X).

4. For $N \geq 500$, there is little difference among tests using different forms of the HCCM.

Error Structure	Method	Sample Size					
		25	50	100	250	500	1,000
Panel 1: 1X	OLSCM	.003	.003	.000	.006	-.002	.001
	HC0	.063	.032	.011	.014	.003	.003
	HC1	.027	.018	.007	.011	.001	.003
	HC2	.016	.011	.004	.010	-.001	.003
	HC3	-.016	-.004	-.007	.006	-.003	.002
Panel 2: 3X	OLSCM	.010	.010	.016	.023	.018	.021
	HC0	.054	.034	.020	.013	.009	.002
	HC1	.023	.020	.013	.011	.008	.002
	HC2	.009	.011	.007	.009	.007	.000
	HC3	-.022	-.006	-.001	.005	.004	-.001
Panel 3: 3Xeq	OLSCM	.015	.030	.036	.044	.036	.044
	HC0	.070	.046	.033	.018	.008	.010
	HC1	.037	.033	.025	.016	.007	.010
	HC2	.025	.025	.020	.014	.006	.008
	HC3	-.010	.003	.008	.008	.001	.007
Panel 4: 34X	OLSCM	.026	.030	.046	.047	.047	.053
	HC0	.078	.038	.034	.012	.007	.004
	HC1	.043	.024	.028	.009	.006	.003
	HC2	.024	.013	.020	.006	.004	.001
	HC3	-.014	-.007	.005	.001	.001	-.001
Panel 5: 123X	OLSCM	.020	.023	.027	.030	.028	.036
	HC0	.070	.045	.022	.008	.003	.008
	HC1	.032	.032	.016	.004	.002	.007
	HC2	.016	.021	.011	.003	.000	.005
	HC3	-.015	.001	.000	-.004	-.002	.005
Panel 6: DsmN	OLS	.002	-.006	.001	-.007	-.001	.000
	HC0	.071	.023	.018	.003	-.001	.003
	HC1	.033	.009	.012	.000	-.002	.002
	HC2	.024	.002	.007	-.003	-.003	.002
	HC3	-.007	-.012	-.003	-.008	-.005	.001
Panel 7: DbigN	OLS	.006	-.003	-.001	-.005	-.002	-.002
	HC0	.053	.023	.015	.006	.002	-.004
	HC1	.020	.010	.008	.004	.001	-.004
	HC2	.011	.003	.004	.002	.000	-.004
	HC3	-.019	-.016	-.007	-.003	-.002	-.005

Table 5: Mean Deviations from Nominal Significance for Various Forms of Heteroscedastistic Errors.

5. For $N \leq 250$, tests based on HC3 perform the best, although HC2 tests work nearly as well.

It is important to keep in mind that the results in Table 5 summarize the size properties across all four β's. As illustrated in Figure 2, however, the effects of heteroscedasticity can vary greatly across coefficients. This point is demonstrated more fully in Table 6 which presents size distortion for tests of β_3. Recall that the data structures were designed so that heteroscedasticity was most strongly associated with x_3, and x_3 is most highly skewed which may increase the chance of having points of high leverage. Several key results emerge:

1. Size distortion for OLSCM tests is much worse for tests of β_3 than for the other β's.

2. The poor size properties for HC0 and HC1 tests are greater for tests of β_3 than for other coefficients.

3. HC3 tests are superior to HC2 tests in nearly all cases.

Figure 3 presents the power curves for tests of H_0: $\beta_1 = 0$ and H_0: $\beta_3 = 0$ for heteroscedastistic χ^2_5 errors associated with x_3 and x_4. This figure reflects several key results that are also found with other data structures and for testing other hypotheses (results not included):

1. Tests based on the OLSCM are most powerful, but this is because of the significant size distortion of these tests. For $N \geq 250$, the size adjusted power is smaller for the OLSCM tests than for other tests.

Error Structure	Method	Sample Size					
		25	50	100	250	500	1,000
Panel 1: 1X	OLSCM	.012	.007	.004	.007	.002	.008
	HC0	.086	.062	.018	.027	.010	.007
	HC1	.044	.048	.011	.025	.007	.006
	HC2	.030	.029	.009	.022	.006	.006
	HC3	-.007	.003	-.006	.016	.005	.005
Panel 2: 3X	OLSCM	.044	.035	.048	.086	.054	.073
	HC0	.111	.067	.039	.040	.019	.016
	HC1	.065	.054	.028	.036	.017	.016
	HC2	.037	.035	.016	.031	.016	.012
	HC3	-.006	.011	.001	.021	.009	.011
Panel 3: 3Xeq	OLSCM	.069	.105	.114	.130	.131	.152
	HC0	.144	.101	.077	.038	.021	.033
	HC1	.105	.085	.069	.036	.020	.033
	HC2	.079	.067	.061	.031	.019	.028
	HC3	.020	.030	.042	.021	.014	.023
Panel 4: 34X	OLSCM	.096	.111	.153	.173	.167	.188
	HC0	.160	.095	.081	.043	.018	.019
	HC1	.121	.076	.074	.039	.017	.018
	HC2	.075	.048	.054	.034	.014	.015
	HC3	.016	.014	.033	.025	.010	.011
Panel 5: 123X	OLSCM	.074	.065	.095	.101	.111	.118
	HC0	.148	.091	.063	.023	.014	.010
	HC1	.100	.078	.055	.020	.013	.009
	HC2	.072	.057	.040	.015	.009	.004
	HC3	.019	.035	.026	.004	.007	.003
Panel 6: DsmN	OLSCM	.004	-.009	-.012	-.001	.005	.000
	HC0	.076	.038	.021	.024	.007	.005
	HC1	.039	.021	.018	.021	.006	.005
	HC2	.021	.008	.010	.015	.003	.005
	HC3	-.007	-.008	-.009	.002	-.001	.003
Panel 7: DbigN	OLSCM	.014	.003	.003	-.002	-.006	.002
	HC0	.084	.018	.017	.016	-.007	-.002
	HC1	.045	.004	.008	.015	-.007	-.003
	HC2	.022	-.003	-.003	.012	-.010	-.003
	HC3	-.018	-.026	-.013	.004	-.010	-.004

Table 6: Deviation from Nominal Significance for Tests of β_3 for Various Forms of Heteroscedastistic Errors.

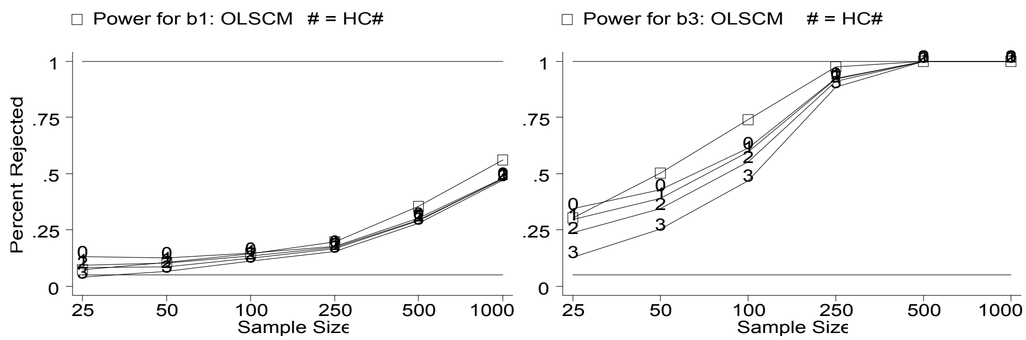

Figure 3: Power of t-test of β_1 and β_3 for χ^2_5 Errors with Heteroscedasticity Associated with x_3 and x_4.

2. HC3 tests are the least powerful of the HCCM tests, followed by HC2 and HC1. These differences are largest for tests of β_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.

5.3 Screening for Heteroscedasticity

Before making our recommendations on how the data analyst should correct for heteroscedasticity, we review what happens if one begins by screening for heteroscedasticity. Applied papers often state that since a model failed to pass a test for heteroscedasticity, HCCMs were used. Our review in Section 2 found that 37 percent of the articles stated that they used a test for heteroscedasticity to determine whether HCCM tests should be used. And, it is likely that other authors used screening tests but did not report them. To determine the consequences of this procedure, we ran the following simulations:

1. Compute a White test for heteroscedasticity.

2. If the test is significant at the .05 level, use a HCCM based test; if the test is not significant, use the OLSCM test.

The White (1980) test is computed by regressing the squared residual, e_i^2, on a constant plus the original x's, their squares, and the cross-products. The White statistic is $W = NR^2$, where R^2 is the coefficient of determination. If the errors are homoscedastistic, W is distributed as χ^2 with degrees of freedom equal to the number of regressors in the auxiliary regression, excluding the constant. A significant value of W leads to the rejection of the null hypothesis of homoscedasticity. We chose the White test since we found it referred to most frequently in applied papers, but we obtained similar results using the Glejser (1969) and Breush and Pagan (1979) tests.

Figure 4 shows the effects of screening when the heteroscedastistic errors are χ_5^2 and associated with x_3 and x_4. The left panel shows the results of the White test that was used to screen for heteroscedasticity. Notice that the test has low power for small samples. The right panel shows the size properties of various tests of H_0: $\beta_3 = \beta_3^*$, where β_3^* is the population value. The results of the standard OLSCM test are shown with $\Box$'s; the results of HC3 tests applied regardless of the result of the screening test are shown with $\triangle$'s. The numbers correspond to the type of HCCM used in the two-step procedure. For example, 3's plot the results when HC3 tests were used when the White test detected heteroscedasticity, otherwise OLSCM tests were used.

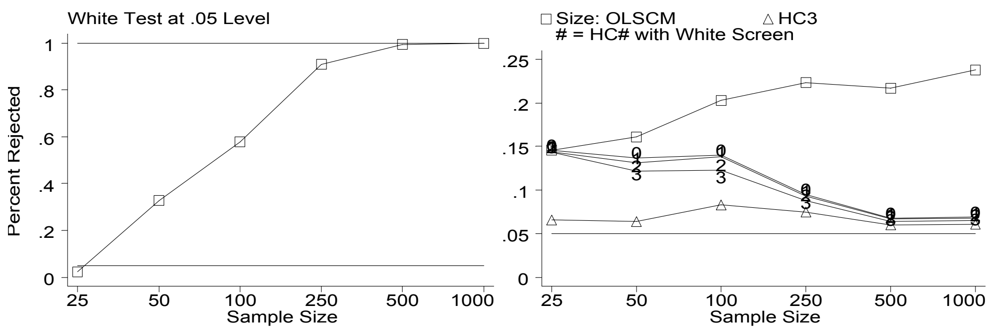

Figure 4: Size and Power of t-tests of β_3 after Screening with a White Tests at the .05 Level, Using Heteroscedastistic χ^2_5 Errors Associated with x_3 and x_4.

Since the White test has less power in small samples, the two-step process will use OLSCM tests more frequently when N is smaller. Consequently, for small N's tests based on screening will have similar size properties to the standard OLSCM test. As N increases and the power of the screening test increases, 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.

6 Summary and conclusions

In this paper, we explored the small sample properties of four versions of the HCCM in the linear regression model. While no Monte Carlo can represent all possible structures that can be encountered in practice, the consistency of our results across a wide variety of structures adds credence to our suggestions for the correction of heteroscedasticity:

1. If there is any reason to suspect that there is heteroscedasticity, tests using HCCMs should be used.

2. If the sample is less than 250, the form of HCCM known as HC3 should be used; when samples are 500 or larger, other versions of the HCCM can be used. The superiority of HC3 over HC2 lies in its better properties in the most extreme cases of heteroscedasticity.

3. The decision to correct for heteroscedasticity should *not* be based on the results of a screening test for heteroscedasticity.

Given the trade-off between correcting for heteroscedasticity with HC3 when there is homoscedasticity and the size distortion of tests based on the OLSCM when there is heteroscedasticity, we recommend that tests based on HC3 should be used tests of individual coefficients in the linear regression model. Given this advice, we hope that software vendors will add HC2 and HC3 to their programs, ideally making HC3 the default option.

In White's classic paper in 1980, he commented on the HCCM by saying that "It is somewhat surprising that these very useful facts have remained unfamiliar to practicing econometricians for so long." We would add that it is unfortunate that authors of statistical texts and software packages seem unfamiliar with the problematic small sample properties of the original HCCM estimator, and that consequently it continues to be used in applied work.

Acknowledgments: We would like to thank Paul Allison, Ken Bollen, Lowell Hargens, and David James for comments on an earlier draft of this paper. Any remaining errors are, of course, our own.

References

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

Aptech Systems, Inc. (1992). *Gauss Version 3.0 Applications: Linear Regression.* Maple Valley, WA: Aptech Systems, Inc.

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 A.R. Pagan. (1979) A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47, 1287-1294.

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

Chesher, A. (1989) Hájek inequalities, measures of leverage and the size of heteroskedasticity robust Wald tests. *Econometrica*, 57, 971-977.

Chesher, A. and G. Austin. (1991) The finite-sample distributions of heteroscedasticity robust Wald statistics. *Journal of Econometrics*, 47, 153-173.

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

Dixon, W.J. (editor). (1992). *BMDP Statistical Software Manual, Version 7.* Berkeley, CA: University of California Press.

Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. (Society for Industrial and Applied Mathematics, Philadelphia, PA).

Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimator for families of linear regressions. *Annals of Mathematical Statistics, 34, 447-456.*

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Volume 1, 59-82). Berkeley: University of California Press.

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.

Greene, W.H. (1990). *Econometric Analysis* (1st Ed.). New York: MacMillan.

Greene, W.H. (1993). *Econometric Analysis* (2nd Ed.). New York: MacMillan.

Greene, William H. (1995). *LIMDEP Version 7.0 User's Manual.* Bellport, NY: Econometric Software, Inc.

Greene, W.H. (1997). *Econometric Analysis* (3rd Ed.). Upper Saddle River, NJ: Prentice Hall.

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

Hall, Bronwyn H., Clint Cummins and Rebecca Schnake. (1995) *TSP Reference Manual, Version 4.3*. Palo Alto, CA: TSP International.

Hartley, H.O., J.N.K. Rao, and G. Keifer. (1969). Variance estimation with on unit per stratum. *Journal of the American Statistical Association*, 64, 841-851.

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 model. *Journal of the American Statistical Association*, 70, 380-385.

Huber, P.J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 221-233.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl, and T-C. Lee. (1985). *The Theory and Practice of Econometrics* (2nd Ed.). New York: Wiley.

Long, J.S., and P.K. Trivedi. (1992). Some specification tests for the linear regression model. *Sociological Methods & Research*, 21, 161-204.

MacKinnon, J.G. and H. White. (1985). Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29, 53-7.

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

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. Hashem and Bahram Pesaran. (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. (1997). *SPSS BASE 7.5 for Windows. User's Guide.* Chicago, IL: SPSS, Inc.

StataCorp. (1997). *Stata Statistical Software: Release 5..* College Station, TX: Stata Corporation.

White, H. (1980). A heteroskedastic-consistent covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica*, 48, 817-838.