Journal of Econometrics 29 (1985) 305–325. North-Holland

SOME HETEROSKEDASTICITY-CONSISTENT COVARIANCE MATRIX ESTIMATORS WITH IMPROVED FINITE SAMPLE PROPERTIES*

James G. MacKINNON

Queen's University, Kingston, Ont., Canada K7L 3N6

Halbert WHITE

University of California at San Diego, La Jolla, CA 92093, USA

Received July 1983, final version received May 1985

We examine several modified versions of the heteroskedasticity-consistent covariance matrix estimator of Hinkley (1977) and White (1980). On the basis of sampling experiments which compare the performance of quasi t-statistics, we find that one estimator, based on the jackknife, performs better in small samples than the rest. We also examine the finite-sample properties of using modified critical values based on Edgeworth approximations, as proposed by Rothenberg (1984). In addition, we compare the power of several tests for heteroskedasticity, and find that it may be wise to employ the jackknife heteroskedasticity-consistent covariance matrix even in the absence of detected heteroskedasticity.

1. Introduction

The linear regression model is extensively used by applied econometricians. Together with its numerous generalizations, it constitutes the foundation of most empirical work in economics. Despite this fact, little is known about the properties of inferences made from this model when standard assumptions are violated. In particular, classical techniques require one to assume that the error terms have a constant variance. This assumption is often not very plausible. Nevertheless, a way of consistently estimating the variance–covariance matrix of ordinary least squares estimates in the face of heteroskedasticity of unknown form is available; see Eicker (1963), Hinkley (1977) and White (1980). This heteroskedasticity-consistent covariance matrix estimator allows one to make valid inferences provided the sample size is sufficiently large.

*We are very grateful to an anonymous referee for a number of very useful suggestions and comments. Earlier versions of this paper appeared as UCSD Department of Economics Discussion Paper No. 82-18, and as Queen's University Institute for Economic Research Discussion Paper No. 537, 1983. Some computer funds were provided by the University of California, but most of the computing was done on an IBM 3081 at Queen's University. MacKinnon's research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada.

Unfortunately, it is not at all obvious what 'sufficiently large' means in practice, and it is well known that statistics with identical large sample properties can perform very differently in samples of small or modest size. In this paper, we examine some estimators which are asymptotically equivalent to the heteroskedasticity-consistent covariance matrix estimator alluded to above, but which may be expected to have superior finite sample properties. Since covariance matrix estimators are most frequently used to construct test statistics, we focus on the behavior of quasi t-statistics constructed using these different estimators. Using sampling experiments, we find that all the new estimators outperform the original one, and that one of them, based on the jackknife, consistently outperforms the other two. These experiments also show that in some circumstances the original estimator can be highly misleading, sometimes even more misleading than the conventional OLS covariance matrix which ignores the possibility of heteroskedasticity.

We next consider an alternative approach due to Rothenberg (1984), in which the original heteroskedasticity-consistent estimator is used in conjunction with modified critical values based on Edgeworth approximations. This approach appears to work well, especially when the sample is reasonably large. Finally, we consider the related question of how well alternative tests for heteroskedasticity perform in the environments studied here. We find that the 'portmanteau' test of White (1980) generally performs well. However, the evidence also suggests that it may be wise to use a heteroskedasticity-consistent covariance matrix estimator even in the absence of detected heteroskedasticity.

The structure of the paper is as follows. In section 2 we describe the problem and the various estimators that will be examined. In sections 3 and 4 we describe the experiments to be performed and present the results of those experiments. In section 5 we discuss the use of modified critical values based on Edgeworth approximations. Finally, in section 6, we examine the performance of alternative tests for heteroskedasticity.

2. Statement of the problem

In this paper we deal exclusively with the linear regression model

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

where y is an $(n \times 1)$ vector of observations on a dependent variable, X is an $(n \times k)$ matrix of observations on independent variables, assumed to be of full rank, and u is an $(n \times 1)$ vector of observations on an error term with mean zero. The ordinary least squares estimator for this model is

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

Inferences about β may be based on the fact that $(\hat{\beta}-\beta)$ has mean zero and covariance matrix

$$(X'X)^{-1}X'\Omega X(X'X)^{-1}, \tag{3}$$

where

$$\mathrm{E}(uu')=\Omega.$$

Conventionally, it is assumed that $\mathrm{E}(uu')=\sigma^2 I_n$. Thus (3) simplifies to $\sigma^2(X'X)^{-1}$, which can be conveniently estimated as

$$\hat{\sigma}^2(X'X)^{-1}, \quad \hat{\sigma}^2=\hat{u}'\hat{u}/(n-k), \quad \hat{u}=\left(I-X(X'X)^{-1}X'\right)y. \tag{4}$$

If X is non-stochastic and u is normally distributed, exact inferences in finite samples can then be based on the t or F distributions. Otherwise, (4) serves as the basis for valid asymptotic inference.

The assumption that the errors are homoskedastic is often implausible. Instead, one may assume that $\mathrm{E}(u_t^2)=\sigma_t^2$, where σ_t varies in some unknown fashion over observations. A heteroskedasticity-consistent covariance matrix estimator which allows one to estimate (3) consistently under general conditions is

$$(X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}, \tag{5}$$

where

$$\hat{\Omega}=\mathrm{diag}\left(\hat{u}_1^2,\hat{u}_2^2,\ldots,\hat{u}_n^2\right);$$

see White (1980).

The estimator (5), which we shall refer to henceforth as *HC*, takes no account of the well-known fact that *OLS* residuals tend to be 'too small'. One simple way to modify *HC* is to use a degrees of freedom correction similar to the one conventionally used to obtain unbiased estimates of σ^2. This yields the modified estimator

$$(n/(n-k))(X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}, \tag{6}$$

which was suggested by Hinkley (1977). We shall refer to it as *HC1*.

The degrees of freedom adjustment in *HC1* is not the only way to compensate for the fact that the OLS residuals tend to underestimate the true errors. If there is no heteroskedasticity, it is easily seen that

$$\mathrm{E}\left(\hat{u}_t^2\right)=(1-k_{tt})\sigma^2, \tag{7}$$

where k_{tt} is the tth diagonal element of the matrix $X(X'X)^{-1}X'$. Thus Horn, Horn and Duncan (1975) suggest using

$$\tilde{\sigma}_t^2 = \hat{u}_t^2/(1 - k_{tt}) \tag{8}$$

as an 'almost unbiased' estimator for σ_t^2. Following their approach, we propose the estimator

$$(X'X)^{-1}X'\tilde{\Omega}X(X'X)^{-1}, \tag{9}$$

where

$$\tilde{\Omega} = \mathrm{diag}(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \ldots, \tilde{\sigma}_n^2),$$

as an alternative way to estimate (3) consistently. We shall refer to this estimator as *HC2*. It is immediate from (7) that *HC2* will be unbiased when the u_t's are in fact homoskedastic. In contrast, as Hinkley (1977) points out, *HC1* will only be unbiased in the special case of a 'balanced' experimental design, for which $k_{tt} = k/n$ for all t.

All of these covariance matrix estimators are intimately related to what statisticians refer to as the 'jackknife'. Efron (1982, p. 19) points out that what is essentially *HC* can be obtained by the infinitesimal jackknife method. Hinkley (1977) derived *HC1* as the covariance matrix of what he called the 'weighted jackknife' estimator, and it would have been possible to derive *HC2* using a very similar argument, although Hinkley did not in fact do so. All of this suggests that the ordinary jackknife [see Efron (1982)] might provide another modified heteroskedasticity-consistent covariance matrix estimator, and indeed that turns out to be the case.

The basic idea of the jackknife is to recompute the estimates of a model n times, each time dropping one of the observations, and then to use the variability of the recomputed estimates as an estimate of the variability of the original estimator. For more details, see Efron (1982). Let $\hat{\beta}_{(t)}$ denote the *OLS* estimate of β based on all observations except the tth. It is easily shown that

$$\hat{\beta}_{(t)} = \hat{\beta} - (X^TX)^{-1}X_t'u_t^*, \tag{10}$$

where X_t denotes the tth row of X and $u_t^* = \hat{u}_t/(1 - k_{tt})$. Then from expression (3.13) of Efron (1982), the jackknife estimate of the covariance matrix of $\hat{\beta}$ is

$$((n-1)/n)\sum_{t=1}^{n}\left[\hat{\beta}_{(t)} - (1/n)\sum_{s=1}^{n}\hat{\beta}_{(s)}\right]\left[\hat{\beta}_{(t)} - (1/n)\sum_{s=1}^{n}\hat{\beta}_{(s)}\right]'. \tag{11}$$

After considerable manipulation,[1] it can be shown that (11) reduces to

$$((n-1)/n)(X'X)^{-1}[X'\Omega^*X-(1/n)(X'u^*u^{*\prime}X)](X'X)^{-1}, \quad (12)$$

where Ω^* is an $n \times n$ diagonal matrix with diagonal elements of u_t^{*2} and off-diagonal elements of zero, and u^* is a vector of the u_t^*'s. We shall refer to this covariance matrix estimator as *HC3*. It is evident that *HC3* is asymptotically equivalent to *HC*, *HC1* and *HC2*, since the middle factor (in square brackets) clearly converges to $X'\Omega X$.

As Messer and White (1984) have shown, it is easy to trick a conventional regression package which is capable of IV estimation into producing *HC* or *HC1*. If the k_{tt}'s can be obtained and used to compute the $\tilde{\sigma}_t$'s, their technique can also be used to make a regression package produce *HC2*. Calculating *HC3* will inevitably be a little more difficult. Almost all the calculations can however be performed with a regression package, since (12) can be rewritten as

$$((n-1)/n)(X'X)^{-1}(X'\Omega^*X)(X'X)^{-1}-((n-1)/n^2)\hat{\gamma}\hat{\gamma}', \quad (13)$$

where $\hat{\gamma}=(X'X)^{-1}X'u^*$. It is tempting to omit the factor $(n-1)/n$ from *HC3*. The effect of this omission will normally be very small. Moreover, our experimental results (see below) suggest that this small effect would normally be in the right direction. Since we did not know that when the experiments were designed, however, we retained the factor $(n-1)/n$ in those experiments.

Since covariance matrix estimators are usually used to compute test statistics, we focus our experiments directly on the behavior of such test statistics. In particular, we examine the small-sample performance of quasi t-statistics used to test hypotheses that particular elements of β assume specified values. For related evidence on how well estimators such as *HC* and *HC1* approximate the true covariance matrix directly, see Cragg (1983) and Nicholls and Pagan (1983).

One important property of these quasi t-statistics may be noted immediately. When the hypothesis being tested is true, the numerator of such a statistic does not depend on β, and is homogeneous of degree one in σ. The covariance matrices (4), (6), (9) and (12) also do not depend on β, and are homogeneous of degree two in σ. Thus these test statistics themselves do not depend on either β or σ. They only depend on the X's and the u's, which may be normalized to have arbitrary variance. Since the exact finite sample properties of these test statistics are otherwise quite difficult to obtain analytically, we investigate these properties using sampling experiments.

[1] When we wrote earlier versions of this paper, we were under the false impression that the jackknife covariance estimator was computationally too complicated to be worth studying. We are extremely grateful to an anonymous referee for pointing out that it can be expressed in the form of (12).

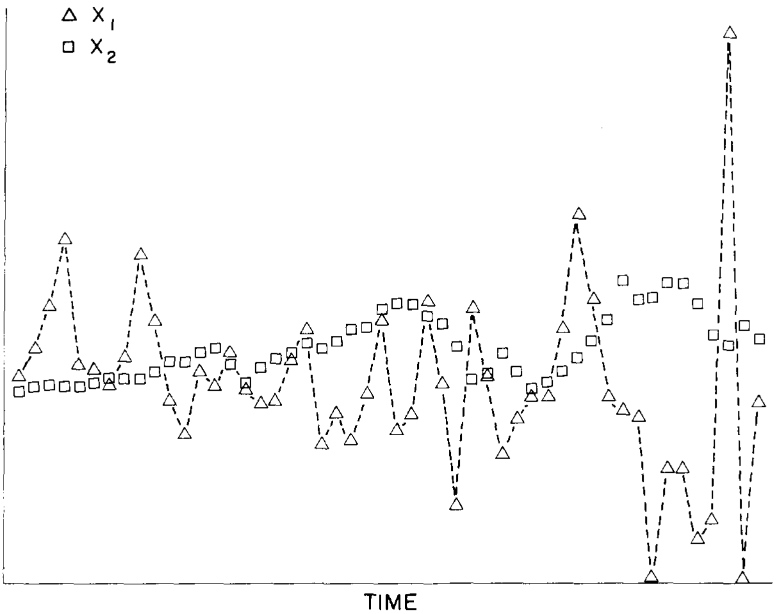

Fig. 1. Regressors used in sampling experiments.

3. Design of the experiments

In all of our experiments we utilized the following model:

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t, \tag{14}$$

where $n = 50$, 100 or 200. For the regressors X_1 and X_2 we used the rate of growth of real U.S. disposable income and the U.S. treasury bill rate, respectively, seasonally adjusted, for 1963-3 to 1975-4. The dependent variable can be thought of as a savings rate. These fifty observations were then replicated the required number of times when more than fifty observations were used. We chose the regressors in this way because we wanted them to be representative of real data, and so that the matrix $X'X/n$ would not change as the sample size n was changed. Plots of X_1 and X_2 are shown in fig. 1.

There were six sets of experiments, in each of which the u_t's were chosen differently. In the first set, referred to as case 1, the u_t's were NID$(0, \sigma^2)$, so that the *OLS* t-statistics are appropriate. The object here is to see how costly it is to use the various heteroskedasticity-consistent estimators when there is in fact no heteroskedasticity.

In the next two sets of experiments (cases 2 and 3), the variance of u_t changed abruptly, as if due to some sort of structural change. The errors u_t

were chosen as $N(0,\sigma^2)$ for $t=1,\dots,25$, $t=51,\dots,75$, $t=101,\dots,125$ and $t=151,\dots,175$, and as $N(0,\alpha^2\sigma^2)$ for the remaining observations. The structural change parameter α was chosen to be 2 in case 2 and 4 in case 3. Notice that the pattern of structural change we used is equivalent to replicating the first 25 observations as many times as necessary (1, 2 or 4 depending on whether $n=50$, 100 or 200), with the u_t's having variance σ^2, and also replicating the second 25 observations as many times as necessary, with the u_t's having variance $\alpha^2\sigma^2$. This pattern was chosen so that increasing the sample size from 50 to 100 or 200 would not change the relationship between the u_t's and the X_t's.

In the final three sets of experiments (cases 4, 5 and 6), the variance of u_t varied because the β_j's varied randomly. Specifically, the model (14) was modified by assuming that

$$\beta_j=\bar{\beta}_j+v_{jt},\qquad v_{jt}\sim N\left(0,\omega_j^2\right),\qquad j=0,1,2. \tag{15}$$

Together with (14), (15) implies that

$$y_t=\bar{\beta}_0+\bar{\beta}_1X_{1t}+\bar{\beta}_2X_{2t}+u_t+v_{0t}+v_{1t}X_{1t}+v_{2t}X_{2t}. \tag{16}$$

Assuming that u_t and the v_{it}'s are independent of each other, the variance of the error term in (16) is

$$\sigma^2+\omega_0^2+X_{1t}^2\omega_1^2+X_{2t}^2\omega_2^2=\sigma_0^2\left(1+X_{1t}^2\gamma_1^2+X_{2t}^2\gamma_2^2\right). \tag{17}$$

Without loss of generality (since the statistics we will be studying are independent of β) we normalized X_{1t} and X_{2t} so that $\Sigma X_{it}^2/n=1$, $i=1,2$. Then for case 4 we chose ($\gamma_1=1,\gamma_2=1$), for case 5 we chose ($\gamma_1=3,\gamma_2=1$) and for case 6 we chose ($\gamma_1=1,\gamma_2=3$).

Each experiment involved 2000 replications, and there were eighteen experiments in all (six cases for each of $n=50$, $n=100$ and $n=200$).[2] For each of the β_i's we calculated four test statistics of the hypothesis that β_i equals its true value. These statistics, denoted *OLS*, *HC1*, *HC2* and *HC3*, utilize the covariance matrices after which they are named. In addition, we calculated a control variate which utilizes the true covariance matrix (3) and is thus exactly $N(0,1)$.

For each experiment we calculated the sample mean, standard deviation, skewness and kurtosis (over the 2000 replications) of each of these test

[2] In fact, we conducted six additional experiments in which $n=150$, but the results were predictable given those for $n=100$ and $n=200$ and are therefore not reported.

statistics. There was nothing in the experimental results to suggest that any of them had a non-zero mean, or that their distributions were not symmetric. In the tables, therefore, we only report the standard deviation (under 'S.D.') and the kurtosis (under 'Kurt.'), which should be one and three, respectively, if the test statistic in question is N(0, 1). If the standard deviation were in fact unity, the sample standard deviation would, assuming normality, have a variance of $1/4000$. The number reported under 'Kurt.' is a standard test statistic for kurtosis, namely the estimated fourth moment about the mean, divided by the square of the estimated second moment.

Although the moments of the sample distributions of the test statistics are of interest, they do not directly tell us how often we will be led to make invalid inferences by using test statistics whose distributions differ from N(0, 1). It is more interesting to ask what proportion of the time each of the test statistics exceeds certain critical values. The critical values we chose were the 5% and 1% levels; absolute critical values for the standard normal at these levels are 1.960 and 2.576.

The obvious way to estimate these rejection frequencies is to use the estimator $\hat{q} = R/N$, where R is the observed number of rejections and N is the number of replications (here 2000). A consistent estimate of the variance of this estimator is $\hat{q}(1-\hat{q})/N$. Since all of the test statistics have the same numerator as the control variate, they should all be highly correlated with it, and it should therefore be possible to obtain more accurate estimates than $\hat{q}$. Davidson and MacKinnon (1981) have proposed a simple technique for doing so, which we utilize here. If the control variate has exceeded its critical value more than the expected number of times, the estimated rejection frequency for the statistic in question will be reduced by an amount that depends on how closely it and the control variate are correlated; the reverse will be true if the control variate has exceeded its critical value less than the expected number of times. The variance of the resulting estimate will depend on the amount of correlation between the control variate and the other statistic, and will never exceed $q(1-q)/N$, asymptotically. For details, see Davidson and MacKinnon (1981).

The fact that we estimated rejection frequencies in this way should be borne in mind when reading the tables. The same estimated rejection frequency may have quite different standard errors in different cases, because the correlation between the control variate and the test statistic may be different. This means that the gain from utilizing this technique varies from case to case. In some cases, the standard errors reported in the tables are more than sixty percent below what they would have been if $\hat{q}$ had been used, equivalent to a replication number of 12,000 or more; in others, the standard errors are only about twenty percent lower, equivalent to a replication number of less than 3000. These are asymptotic standard errors, but the very large number of replications should endure their validity.

4. Results of the experiments

The results of twelve of the eighteen experiments just discussed are presented in tables 1 through 4. Cases 2 and 4 are omitted to save space; the results for case 2 were similar to those for case 3, but not as pronounced, while the results for case 4 were reasonably similar to those for cases 5 and 6. An asterisk indicates that the quantity in question differs significantly at the one percent level from what it should be if the test statistic were really N(0,1). The tables largely speak for themselves, but we will discuss a few points of interest.

The most obvious result in tables 1 through 4 is that almost all the quasi t-statistics have standard deviations greater than unity, so that rejection frequencies of tests based on them almost always exceed the nominal sizes of the tests. As one would expect, these standard deviations tend to decline as the sample size increases. They also vary systematically with the coefficient being estimated, the quasi t-statistics for β_1 tending to have much larger variances than those for β_0 or β_2. The pattern of heteroskedasticity has a major impact on the distributions of the quasi t-statistics. They tend to be closest to their asymptotic N(0,1) distribution when there is no heteroskedasticity, in table 1.

In every single case, the standard deviation of the quasi t-statistic based on *HC1* exceeded that for *HC2*, which in turn exceeded that for *HC3*. Since there was certainly no tendency for *HC3* to have too small a variance, this implies that *HC3* is the covariance matrix estimator of choice. The difference between *HC1* and *HC3* is often striking, and the difference between *HC* and *HC3* would of course be even more striking. From table 1 it is clear that using *HC* or *HC1* when there is in fact no heteroskedasticity and the sample size is small could easily lead to serious errors of inference, while using *HC3* is almost as reliable as using *OLS*.

Even *HC3* did not always perform well when the sample size was small and there was substantial heteroskedasticity. Its worst performance was in case 5 (table 3) for β_1 when $n = 50$. The standard deviation of the *HC3* t-statistic is 1.177 here, and it would incorrectly reject the null hypothesis 3.1% of the time at the nominal 1% level. But although *HC3* performs poorly here, it performs much better than its competitors, since *HC2* would reject the null 4.7% of the time, *HC1* would reject it 6.8% of the time, and the usual *OLS* t-statistic would reject it 27.2% of the time.

Thus, subject to the usual qualifications about results of sampling experiments, those in tables 1 to 4 suggest the following conclusions:

1. Among the heteroskedasticity-consistent estimators, *HC3* is clearly the procedure of choice.
2. The usual *OLS* covariance estimator can be very seriously misleading in the presence of heteroskedasticity. When it is, *HC3* is also likely to be misleading if the sample size is small, but much less so than *OLS*.

Table 1

Case 1: No heteroskedasticity.[a]

Coef.	No. of obs.	C.V.	Stat.	S.D.	Kurt.	5%	1%
β_0	50	1.030	*OLS*	1.053*	3.06	0.059*(0.0032)	0.016*(0.0019)
		2.89	*HC1*	1.099*	3.21	0.066*(0.0037)	0.020*(0.0021)
			HC2	1.081*	3.22	0.063*(0.0037)	0.018*(0.0020)
			HC3	1.035	3.24	0.058 (0.0037)	0.013 (0.0020)
β_0	100	0.998	*OLS*	1.007	2.89	0.052 (0.0024)	0.011 (0.0014)
		2.85	*HC1*	1.023	2.91	0.055 (0.0027)	0.013 (0.0018)
			HC2	1.014	2.91	0.053 (0.0028)	0.013 (0.0018)
			HC3	0.994	2.92	0.048 (0.0025)	0.012 (0.0018)
β_0	200	1.011	*OLS*	1.016	2.82	0.054 (0.0019)	0.009 (0.0012)
		2.80	*HC1*	1.024	2.85	0.057*(0.0026)	0.012 (0.0014)
			HC2	1.020	2.85	0.056 (0.0026)	0.011 (0.0015)
			HC3	1.010	2.85	0.052 (0.0022)	0.011 (0.0015)
β_1	50	1.018	*OLS*	1.037	3.16	0.057 (0.0033)	0.013 (0.0019)
		3.02	*HC1*	1.217*	3.47*	0.094*(0.0048)	0.038*(0.0033)
			HC2	1.159*	3.55*	0.082*(0.0046)	0.030*(0.0031)
			HC3	1.074*	3.69*	0.067*(0.0046)	0.023*(0.0029)
β_1	100	0.998	*OLS*	1.010	3.11	0.051 (0.0024)	0.013 (0.0015)
		2.99	*HC1*	1.100*	3.42*	0.072*(0.0039)	0.026*(0.0029)
			HC2	1.071*	3.46*	0.067*(0.0037)	0.024*(0.0027)
			HC3	1.030	3.52*	0.055 (0.0036)	0.018*(0.0024)
β_1	200	1.006	*OLS*	1.009	3.16	0.052 (0.0025)	0.011 (0.0013)
		3.04	*HC1*	1.059*	3.38*	0.065*(0.0033)	0.014 (0.0018)
			HC2	1.043*	3.39*	0.060*(0.0034)	0.014 (0.0018)
			HC3	1.022	3.41*	0.056 (0.0034)	0.011 (0.0017)
β_2	50	1.029	*OLS*	1.051*	3.06	0.062*(0.0033)	0.016*(0.0021)
		2.94	*HC1*	1.100*	3.16	0.073*(0.0036)	0.022*(0.0024)
			HC2	1.083*	3.16	0.071*(0.0036)	0.022*(0.0024)
			HC3	1.040	3.17	0.059 (0.0039)	0.017*(0.0024)
β_2	100	0.994	*OLS*	1.002	2.95	0.051 (0.0025)	0.010 (0.0016)
		2.92	*HC1*	1.021	2.97	0.050 (0.0030)	0.012 (0.0018)
			HC2	1.013	2.97	0.049 (0.0030)	0.011 (0.0018)
			HC3	0.994	2.97	0.045 (0.0028)	0.009 (0.0017)
β_2	200	1.003	*OLS*	1.008	2.84	0.052 (0.0019)	0.010 (0.0015)
		2.81	*HC1*	1.015	2.87	0.054 (0.0025)	0.013 (0.0016)
			HC2	1.011	2.87	0.053 (0.0025)	0.013 (0.0016)
			HC3	1.002	2.87	0.051 (0.0024)	0.011 (0.0016)

[a]Numbers under 'C.V.' are the standard deviation and kurtosis of the control variate.

Numbers under '5%' and '1%' are the estimated rejection probabilities at those nominal levels. The standard errors of these estimates, which incorporate the information in the control variate, are in brackets.

An asterisk indicates that a quantity is significantly different at the 1% level from what it should be if the t-statistic were N(0,1).

Number of replications = 2000.

Table 2

Case 3: Structural change in variance, $\alpha = 4$.[a]

Coef.	No. of obs.	C.V.	Stat.	S.D.	Kurt.	5%	1%
β_0	50	0.979	*OLS*	1.095*	2.99	0.084*(0.0041)	0.019*(0.0027)
		2.97	*HC1*	1.072*	3.02	0.082*(0.0044)	0.017*(0.0026)
			HC2	1.044*	3.05	0.072*(0.0042)	0.015 (0.0025)
			HC3	0.989	3.10	0.061*(0.0039)	0.010 (0.0022)
β_0	100	1.036	*OLS*	1.130*	2.97	0.071*(0.0036)	0.018*(0.0019)
		2.99	*HC1*	1.084*	2.98	0.066*(0.0037)	0.016*(0.0021)
			HC2	1.068*	2.99	0.065*(0.0036)	0.014 (0.0021)
			HC3	1.041*	3.01	0.061*(0.0036)	0.012 (0.0019)
β_0	200	1.006	*OLS*	1.087*	3.09	0.074*(0.0030)	0.018*(0.0015)
		3.07	*HC1*	1.029	3.15	0.054 (0.0029)	0.013 (0.0016)
			HC2	1.021	3.16	0.052 (0.0027)	0.013 (0.0016)
			HC3	1.008	3.16	0.050 (0.0028)	0.011 (0.0016)
β_1	50	1.000	*OLS*	1.316*	3.19	0.138*(0.0045)	0.051*(0.0032)
		3.03	*HC1*	1.280*	3.74*	0.117*(0.0054)	0.053*(0.0041)
			HC2	1.210*	3.96*	0.100*(0.0053)	0.043*(0.0039)
			HC3	1.113*	4.27*	0.081*(0.0051)	0.030*(0.0035)
β_1	100	1.013	*OLS*	1.301*	3.08	0.130*(0.0038)	0.047*(0.0024)
		3.01	*HC1*	1.152*	3.33*	0.088*(0.0047)	0.029*(0.0031)
			HC2	1.116*	3.38*	0.075*(0.0045)	0.025*(0.0030)
			HC3	1.068*	3.43*	0.065*(0.0042)	0.022*(0.0029)
β_1	200	1.001	*OLS*	1.271*	3.13	0.126*(0.0028)	0.044*(0.0023)
		3.07	*HC1*	1.077*	3.33*	0.071*(0.0038)	0.019*(0.0024)
			HC2	1.059*	3.34*	0.066*(0.0040)	0.016 (0.0023)
			HC3	1.035	3.36*	0.062*(0.0038)	0.014 (0.0022)
β_2	50	0.976	*OLS*	1.152*	3.09	0.091*(0.0036)	0.034*(0.0028)
		3.03	*HC1*	1.078*	3.23	0.078*(0.0040)	0.023*(0.0027)
			HC2	1.054*	3.23	0.071*(0.0039)	0.022*(0.0028)
			HC3	1.004	3.24	0.058 (0.0038)	0.015 (0.0025)
β_2	100	1.030	*OLS*	1.184*	2.90	0.089*(0.0038)	0.027*(0.0022)
		2.91	*HC1*	1.080*	2.93	0.068*(0.0038)	0.016 (0.0023)
			HC2	1.067*	2.94	0.064*(0.0036)	0.016 (0.0023)
			HC3	1.043*	2.95	0.057 (0.0034)	0.014 (0.0022)
β_2	200	1.007	*OLS*	1.147*	2.96	0.087*(0.0029)	0.023*(0.0018)
		2.94	*HC1*	1.032	3.06	0.063*(0.0032)	0.011 (0.0016)
			HC2	1.026	3.06	0.059*(0.0030)	0.010 (0.0016)
			HC3	1.014	3.06	0.058*(0.0029)	0.010 (0.0016)

[a]See notes to table 1.

Table 3

Case 5: Random coefficient model, weights = (3, 1).[a]

Coef.	No. of obs.	C.V.	Stat.	S.D.	Kurt.	5%	1%
β_0	50	1.002	*OLS*	1.325*	2.70*	0.142*(0.0050)	0.046*(0.0036)
		3.00	*HC1*	1.236*	2.61*	0.109*(0.0060)	0.024*(0.0034)
			HC2	1.149*	2.69*	0.078*(0.0057)	0.018*(0.0029)
			HC3	1.038	2.83	0.053 (0.0049)	0.012 (0.0024)
β_0	100	1.002	*OLS*	1.307*	2.93	0.133*(0.0042)	0.043*(0.0026)
		3.16	*HC1*	1.143*	2.96	0.087*(0.0048)	0.017 (0.0027)
			HC2	1.096*	2.99	0.070*(0.0046)	0.015 (0.0026)
			HC3	1.037	3.04	0.058 (0.0044)	0.011 (0.0023)
β_0	200	0.963	*OLS*	1.252*	2.92	0.137*(0.0035)	0.045*(0.0022)
		3.04	*HC1*	1.043*	2.99	0.059 (0.0041)	0.019*(0.0027)
			HC2	1.019	3.01	0.054 (0.0040)	0.017*(0.0025)
			HC3	0.991	3.02	0.048 (0.0038)	0.013 (0.0022)
β_1	50	0.997	*OLS*	2.205*	2.57*	0.398*(0.0056)	0.272*(0.0054)
		3.07	*HC1*	1.483*	2.78	0.172*(0.0080)	0.068*(0.0055)
			HC2	1.338*	2.98	0.122*(0.0072)	0.047*(0.0047)
			HC3	1.177*	3.28*	0.082*(0.0061)	0.031*(0.0039)
β_1	100	0.996	*OLS*	2.211*	2.74	0.391*(0.0045)	0.253*(0.0041)
		2.97	*HC1*	1.291*	3.16	0.120*(0.0064)	0.049*(0.0046)
			HC2	1.220*	3.24	0.094*(0.0059)	0.038*(0.0042)
			HC3	1.139*	3.34*	0.076*(0.0056)	0.030*(0.0038)
β_1	200	0.976	*OLS*	2.158*	2.85	0.377*(0.0039)	0.249*(0.0042)
		2.96	*HC1*	1.136*	3.16	0.093*(0.0054)	0.025*(0.0033)
			HC2	1.101*	3.19	0.085*(0.0053)	0.024*(0.0032)
			HC3	1.062*	3.22	0.076*(0.0052)	0.020*(0.0030)
β_2	50	1.002	*OLS*	1.186*	2.80	0.098*(0.0044)	0.023*(0.0030)
		2.98	*HC1*	1.173*	2.72*	0.094*(0.0054)	0.018*(0.0029)
			HC2	1.105*	2.75	0.070*(0.0051)	0.013 (0.0025)
			HC3	1.011	2.84	0.050 (0.0047)	0.009 (0.0021)
β_2	100	1.004	*OLS*	1.165*	2.92	0.091*(0.0037)	0.024*(0.0026)
		3.07	*HC1*	1.108*	2.91	0.077*(0.0043)	0.018*(0.0027)
			HC2	1.070*	2.92	0.065*(0.0041)	0.014 (0.0025)
			HC3	1.021	2.95	0.052 (0.0040)	0.009 (0.0021)
β_2	200	0.969	*OLS*	1.117*	2.94	0.086*(0.0033)	0.024*(0.0022)
		3.01	*HC1*	1.026	3.00	0.061*(0.0040)	0.014 (0.0023)
			HC2	1.007	3.01	0.056 (0.0040)	0.014 (0.0021)
			HC3	0.982	3.02	0.053 (0.0039)	0.013 (0.0020)

[a] See notes to table 1.

Table 4

Case 6: Random coefficient model, weights = (1, 3)

Coef.	No. of obs.	C.V.	Stat.	S.D.	Kurt.	5%	1%
β_0	50	0.986	*OLS*	1.068*	3.13	0.074*(0.0033)	0.016*(0.0016)
		3.08	*HC1*	1.092*	3.23	0.077*(0.0040)	0.021*(0.0024)
			HC2	1.062*	3.27	0.074*(0.0040)	0.018*(0.0022)
			HC3	1.005	3.34*	0.057 (0.0039)	0.011 (0.0019)
β_0	100	1.002	*OLS*	1.067*	3.03	0.062*(0.0029)	0.016*(0.0015)
		3.01	*HC1*	1.060*	3.08	0.064*(0.0035)	0.016*(0.0019)
			HC2	1.043*	3.09	0.059*(0.0034)	0.016*(0.0019)
			HC3	1.015	3.09	0.052 (0.0034)	0.012 (0.0019)
β_0	200	1.007	*OLS*	1.063*	3.01	0.064*(0.0023)	0.015*(0.0016)
		2.98	*HC1*	1.033	3.05	0.060*(0.0027)	0.015*(0.0019)
			HC2	1.024	3.05	0.054 (0.0027)	0.015 (0.0019)
			HC3	1.010	3.06	0.050 (0.0027)	0.014 (0.0018)
β_1	50	1.013	*OLS*	1.269*	3.17	0.114*(0.0039)	0.042*(0.0026)
		3.10	*HC1*	1.319*	3.78*	0.129*(0.0058)	0.052*(0.0041)
			HC2	1.242*	3.96*	0.110*(0.0056)	0.040*(0.0038)
			HC3	1.137*	4.21*	0.084*(0.0053)	0.030*(0.0035)
β_1	100	1.007	*OLS*	1.249*	2.93	0.114*(0.0030)	0.037*(0.0024)
		2.84	*HC1*	1.180*	3.46*	0.096*(0.0045)	0.026*(0.0030)
			HC2	1.140*	3.56*	0.087*(0.0045)	0.023*(0.0029)
			HC3	1.089*	3.67*	0.063*(0.0042)	0.019*(0.0027)
β_1	200	1.034	*OLS*	1.266*	2.75	0.114*(0.0026)	0.034*(0.0025)
		2.74	*HC1*	1.122*	2.99	0.072*(0.0043)	0.020*(0.0022)
			HC2	1.100*	3.01	0.067*(0.0041)	0.017*(0.0021)
			HC3	1.073*	3.03	0.062*(0.0042)	0.015 (0.0020)
β_2	50	0.984	*OLS*	1.143*	3.04	0.094*(0.0033)	0.025*(0.0023)
		2.95	*HC1*	1.088*	3.25	0.073*(0.0043)	0.023*(0.0027)
			HC2	1.063*	3.28	0.069*(0.0043)	0.021*(0.0026)
			HC3	1.013	3.33*	0.059 (0.0040)	0.014 (0.0022)
β_2	100	0.989	*OLS*	1.132*	3.04	0.085*(0.0031)	0.024*(0.0019)
		3.06	*HC1*	1.044*	3.13	0.058 (0.0036)	0.017*(0.0021)
			HC2	1.031	3.13	0.054 (0.0037)	0.016*(0.0020)
			HC3	1.007	3.13	0.050 (0.0034)	0.015*(0.0019)
β_2	200	1.001	*OLS*	1.137*	3.01	0.086*(0.0028)	0.024*(0.0016)
		2.96	*HC1*	1.027	3.04	0.054 (0.0030)	0.012 (0.0018)
			HC2	1.020	3.04	0.051 (0.0029)	0.012 (0.0018)
			HC3	1.008	3.04	0.050 (0.0029)	0.011 (0.0017)

[a]See notes to table 1.

3. When there is no heteroskedasticity, all the *HC* estimators are less reliable than *OLS*, but *HC3* does not seem to be much less reliable.

5. An alternative approach

What we have done so far is to modify the heteroskedasticity-consistent covariance estimator so as to obtain test statistics whose finite sample distributions are closer to their asymptotic ones. This is not the only approach to making more accurate inferences in finite samples. An alternative approach, which is theoretically appealing but technically demanding, would be to use the original test statistic based on *HC* in conjunction with size-corrected critical values. The latter may be obtained by the use of Edgeworth expansions, in this case second-order asymptotic approximations to the distribution of the test statistic.

In a recent paper, Rothenberg (1984) has applied this technique to exactly the problem that interests us in this paper. His fundamental result is that

$$t'_\alpha = t_\alpha\left[1 + \left(c_1\left(1 + t_\alpha^2\right) + c_2\left(1 - t_\alpha^2\right) + c_3\right)/2n\right], \tag{18}$$

where t_α is a size-α critical value for the normal distribution and t'_α is an adjusted size-α critical value. The parameters c_1, c_2 and c_3 are constants which depend in a complicated way on the regressors, the pattern of heteroskedasticity, and the coefficient (or linear combination of coefficients) for which the test is to be conducted. In practice, the parameters c_1 through c_3 will have to be estimated using the least squares residuals, since the pattern of heteroskedasticity is unknown.

We conducted a number of experiments to see how this approach of using *HC* with adjusted critical values compares with the much simpler approach of using *HC3* with the usual asymptotic critical values. We looked only at cases 2 and 4, the ones which were not reported in tables 1 to 4. Case 2 was chosen because the heteroskedasticity was relatively mild in that case, and case 4 was chosen because it was representative of all the random coefficient cases. Results for both these cases for samples of size 50, 100, 200 and 400 are shown in table 5, which tabulates rejection frequencies for tests which are nominally at the 5% and 1% levels. '*Edge-E*' shows the rejection frequencies when c_1, c_2 and c_3 are estimated from the data, as they would have to be in practice, while '*Edge-T*' shows the rejection frequencies when the true values of those parameters are used. All results are based on 10,000 replications, so experimental error should be very small.[3]

[3] We used 10,000 replications here instead of 2000 because early results showed that *HC3* and *Edge-E* performed similarly for samples of medium size, and we wanted to minimize experimental error. Results are based on ten sets of 1000 replications.

Table 5

Performance of Edgeworth critical values.[a]

Coef.	No. of obs.	Test	Rejection frequencies: Case 2 5%	Rejection frequencies: Case 2 1%	Rejection frequencies: Case 4 5%	Rejection frequencies: Case 4 1%
β_0	50	*HC*	0.079*(0.0018)	0.023*(0.0012)	0.101*(0.0021)	0.034*(0.0015)
		HC3	0.052 (0.0016)	0.014*(0.0010)	0.060*(0.0020)	0.017*(0.0012)
		Edge-E	0.059*(0.0017)	0.019*(0.0011)	0.069*(0.0021)	0.022*(0.0013)
		Edge-T	0.050 (0.0015)	0.010 (0.0008)	0.040*(0.0015)	0.005*(0.0007)
β_0	100	*HC*	0.065*(0.0014)	0.015*(0.0009)	0.081*(0.0018)	0.022*(0.0011)
		HC3	0.052 (0.0014)	0.011 (0.0007)	0.059*(0.0017)	0.014*(0.0010)
		Edge-E	0.054*(0.0014)	0.012*(0.0008)	0.060*(0.0018)	0.015*(0.0010)
		Edge-T	0.051 (0.0014)	0.008 (0.0007)	0.047 (0.0015)	0.008*(0.0007)
β_0	200	*HC*	0.057*(0.0012)	0.012*(0.0007)	0.064*(0.0015)	0.017*(0.0009)
		HC3	0.051 (0.0012)	0.011 (0.0007)	0.053 (0.0015)	0.013*(0.0009)
		Edge-E	0.052 (0.0012)	0.012*(0.0007)	0.051 (0.0015)	0.014*(0.0009)
		Edge-T	0.050 (0.0012)	0.010 (0.0006)	0.049 (0.0014)	0.009 (0.0007)
β_0	400	*HC*	0.053*(0.0010)	0.011 (0.0006)	0.058*(0.0013)	0.012 (0.0017)
		HC3	0.048 (0.0009)	0.010 (0.0005)	0.054*(0.0013)	0.011 (0.0007)
		Edge-E	0.048 (0.0009)	0.010 (0.0005)	0.054*(0.0013)	0.011 (0.0007)
		Edge-T	0.048 (0.0009)	0.009 (0.0005)	0.052 (0.0012)	0.010 (0.0006)
β_1	50	*HC*	0.122*(0.0024)	0.051*(0.0017)	0.175*(0.0032)	0.078*(0.0025)
		HC3	0.080*(0.0022)	0.027*(0.0014)	0.092*(0.0027)	0.038*(0.0019)
		Edge-E	0.116*(0.0027)	0.089*(0.0026)	0.123*(0.0032)	0.090*(0.0028)
		Edge-T	0.043*(0.0017)	0.005*(0.0006)	0.031*(0.0017)	0.003*(0.0005)
β_1	100	*HC*	0.087*(0.0020)	0.029*(0.0013)	0.116*(0.0026)	0.048*(0.0019)
		HC3	0.062*(0.0018)	0.018*(0.0011)	0.082*(0.0024)	0.030*(0.0016)
		Edge-E	0.068*(0.0020)	0.029*(0.0015)	0.076*(0.0024)	0.033*(0.0017)
		Edge-T	0.044*(0.0016)	0.009 (0.0008)	0.045*(0.0018)	0.006*(0.0007)
β_1	200	*HC*	0.070*(0.0016)	0.021*(0.0010)	0.085*(0.0022)	0.029*(0.0014)
		HC3	0.056*(0.0015)	0.014*(0.0009)	0.067*(0.0020)	0.021*(0.0013)
		Edge-E	0.055*(0.0016)	0.017*(0.0011)	0.061*(0.0019)	0.019*(0.0012)
		Edge-T	0.047 (0.0013)	0.008*(0.0006)	0.047 (0.0017)	0.008*(0.0008)
β_1	400	*HC*	0.060*(0.0013)	0.015*(0.0009)	0.070*(0.0018)	0.019*(0.0011)
		HC3	0.053 (0.0013)	0.014*(0.0008)	0.060*(0.0017)	0.014*(0.0009)
		Edge-E	0.052 (0.0013)	0.014*(0.0009)	0.055*(0.0017)	0.015*(0.0009)
		Edge-T	0.049 (0.0013)	0.010 (0.0007)	0.049 (0.0016)	0.010 (0.0008)
β_2	50	*HC*	0.080*(0.0018)	0.026*(0.0013)	0.091*(0.0019)	0.030*(0.0013)
		HC3	0.055*(0.0017)	0.015*(0.0010)	0.058*(0.0018)	0.016*(0.0011)
		Edge-E	0.058*(0.0017)	0.020*(0.0012)	0.067*(0.0019)	0.021*(0.0012)
		Edge-T	0.051 (0.0016)	0.011 (0.0009)	0.048 (0.0016)	0.009 (0.0008)
β_2	100	*HC*	0.064*(0.0015)	0.016*(0.0009)	0.073*(0.0017)	0.020*(0.0010)
		HC3	0.053 (0.0014)	0.012 (0.0008)	0.056*(0.0016)	0.014*(0.0009)
		Edge-E	0.053 (0.0014)	0.013*(0.0008)	0.057*(0.0016)	0.015*(0.0010)
		Edge-T	0.051 (0.0014)	0.009 (0.0007)	0.051 (0.0015)	0.009 (0.0007)
β_2	200	*HC*	0.060*(0.0014)	0.013*(0.0006)	0.061*(0.0014)	0.013*(0.0007)
		HC3	0.053 (0.0013)	0.010 (0.0006)	0.053*(0.0013)	0.011 (0.0007)
		Edge-E	0.053 (0.0013)	0.011 (0.0006)	0.053 (0.0013)	0.012*(0.0007)
		Edge-T	0.050 (0.0013)	0.009 (0.0006)	0.050 (0.0013)	0.010 (0.0007)
β_2	400	*HC*	0.053*(0.0010)	0.012 (0.0006)	0.055*(0.0011)	0.012 (0.0007)
		HC3	0.049 (0.0010)	0.011 (0.0006)	0.050 (0.0011)	0.011 (0.0007)
		Edge-E	0.050 (0.0010)	0.011 (0.0006)	0.050 (0.0011)	0.011 (0.0006)
		Edge-T	0.049 (0.0010)	0.010 (0.0006)	0.048 (0.0011)	0.010 (0.0006)

[a] Estimated rejection frequencies are simple averages of 10 sets of control variate estimates, each based on 1000 replications.

An asterisk indicates that a rejection frequency is significantly different from 0.05 or 0.01 at the 1% level.

The results for *Edge-T* show that Rothenberg's Edgeworth expansions are generally quite good, and become very good indeed as the sample size gets past 100. If anything, the corrected critical values tend to be too conservative. Unfortunately, these good results usually do not carry over to *Edge-E*, for which the correct critical values are almost always not conservative enough. When the sample size is 50, *HC3* always yields more accurate inferences than *Edge-E*, and that is usually the case for $n = 100$ as well. For $n = 200$ and $n = 400$, however, *HC3* no longer outperforms *Edge-E* overall, although both perform very well. As one might expect from the nature of Edgeworth approximations, *Edge-E* typically performs less well at the 1% level than at the 5% level. Except for a very few cases at the 1% level with $n = 50$, *Edge-E* does always outperform *HC*.

These results suggest that Edgeworth expansions for t-statistics based on *HC* are valuable, but may be more useful as a theoretical tool than as a practical method to obtain corrected critical values. This may however be too pessimistic. In principal, Rothenberg's technique could be applied to *HC1*, *HC2* or *HC3* instead of to *HC*, and it is quite possible that this would produce improved results. The approach could also be modified by the use of alternative asymptotic expansions, by improved methods for estimating the parameters c_1, c_2 and c_3, or by more sophisticated methods for choosing a critical value, not necessarily equal to t'_α, but making use of the information that t'_α conveys. Thus future research may well make Edgeworth expansions look more attractive than they do at present.

6. Tests for heteroskedasticity

Using the heteroskedasticity-consistent covariance matrix estimator as a starting point, White (1980) proposed a test for heteroskedasticity of unknown form. In the case of our model (14), the White test may be carried out by regressing the squared OLS residuals $\hat{u}_t^2$ on a constant, X_1, X_2, X_1^2, X_2^2 and X_1X_2. The test statistic is n times the R^2 from this regression, and it is asymptotically distributed as chi-squared with (in this case) 5 degrees of freedom. In the tables, this test will be referred to as H.

In view of the success of *HC2* and *HC3*, it is natural to wonder whether modified versions of the White test might perform better than the original. In the case of *HC3* it is not obvious how one should modify the test. However, in the case of *HC2* it is straightforward to modify it by using $\tilde{\sigma}_t^2$ instead of $\hat{u}_t^2$ as the regressand. Unfortunately, this modified version of H turned out to have poorer small-sample properties under the null than the original, and we therefore dropped it from our experiments.

Lagrange Multiplier tests for heteroskedasticity have recently become very popular. In the case of the random coefficient model described in section 3, a particularly simple form of the *LM* test may be computed by regressing $\hat{u}_t^2$ on a constant, X_1^2 and X_2^2. The test statistic is then n times the R^2 from this regression, and it is asymptotically distributed as chi-squared with (in this case) 2 degrees of freedom. For details, see Koenker (1981) and Breusch and Pagan (1979). A similar test may be constructed to test against a structural change in variance. In this case $\hat{u}_t^2$ is regressed on a constant and on a dummy variable equal to zero half the time and one the other half; the test has one degree of freedom. These tests will be referred to as *LM1* and *LM2*, respectively.

Over the years, numerous *ad hoc* tests for heteroskedasticity have been proposed. Among the most popular of these is the F test suggested by Goldfeld and Quandt (1965). The data are ordered by time or by one of the regressors, separate regressions are performed on the first and last thirds of the data (leaving out a third in the middle), and the ratio of the sums of squared residuals is then formed. Under the null, this ratio is distributed as F with both numerator and denominator degrees of freedom equal to $n/3-k$. This test has the advantage of being exact, but may have little power if the actual heteroskedasticity is not closely related to time or to one of the regressors. We calculated three tests of this type. In all cases the partial regressions used 17, 34, or 68 observations (so that 16 were omitted in the middle of each 50). *F1* is the test based on ordering the data in the same way that they are ordered for the structural change in variance (i.e., by time, given the odd way that time works in our experiments). *F2* is the test based on ordering the data according to X_1, and *F3* is the test based on ordering according to X_2.

Before we can examine the power of any of these tests, we must determine how well the asymptotic tests (H, *LM1* and *LM2*) perform under the null. Unfortunately, there are no obvious control variates comparable to the ones used in our previous experiments. Thus in order to obtain reasonably accurate estimates, we utilized 8000 replications. The results of these experiments are shown in table 6. The left-hand columns show the estimated rejection probabilities at nominal levels of 5% and 1%, together with estimated standard errors. An asterisk indicates that the estimate differs from the nominal level by more than 2.576 estimated standard errors. It is noteworthy that *LM1* always rejects the null significantly less often than it should, while H also tends to reject the null too infrequently. The right-hand columns of table 6 show estimated critical values, followed by 95% confidence intervals based on the usual non-parametric approximations. These estimated critical values will be used in comparing the power of different tests, and the fact that they are only estimates should be borne in mind.

The powers of various tests for heteroskedasticity are compared in tables 7, 8 and 9, which deal with cases 2, 4, and 6, respectively. For the most part these tables are self-explanatory, so we will mention only a few points of interest. The H test performs least well relative to some of the other tests when the heteroskedasticity takes the form of a structural change in variance. *LM2* and *F1*, which are specifically designed to test against this form of heteroskedastic-

Table 6

Tests for heteroskedasticity: Performance under the null.[a]

No. of obs.	Test	Estimated rejection probabilities 5%	Estimated rejection probabilities 1%	Estimated critical values 5%	Estimated critical values 1%
50	H	0.042*(0.0022)	0.012 (0.0012)	10.65(10.48–10.94)	15.53(14.86–16.19)
	$LM1$	0.030*(0.0019)	0.009 (0.0010)	4.99(4.86– 5.17)	8.76(8.06– 9.45)
	$LM2$	0.047 (0.0024)	0.006*(0.0008)	3.74(3.61– 3.90)	5.91(5.62– 6.12)
100	H	0.045 (0.0023)	0.014*(0.0013)	10.86(10.51–11.09)	16.08(15.50–16.91)
	$LM1$	0.038*(0.0021)	0.012 (0.0012)	5.46(5.21– 5.61)	9.79(9.20–10.95)
	$LM2$	0.052 (0.0025)	0.008 (0.0010)	3.90(3.74– 4.05)	6.25(6.05– 6.63)
200	H	0.046 (0.0023)	0.011 (0.0012)	10.88(10.64–11.13)	15.47(14.96–16.04)
	$LM1$	0.042*(0.0023)	0.012 (0.0013)	5.66(5.45– 5.86)	9.79(9.22–10.40)
	$LM2$	0.051 (0.0025)	0.010 (0.0011)	3.86(3.67– 4.03)	6.75(6.35– 7.10)

[a] Number of replications = 8000.

An asterisk indicates that a quantity is significantly different at the 1% level from what it should be if the statistic had its asymptotic distribution. The statistics H, $LM1$ and $LM2$ should be asymptotically distributed as chi-square with 5, 2 and 1 degrees of freedom, respectively.

Table 7

Tests for heteroskedasticity: Case 2: Structural change in variance, $\alpha = 2$.[a]

No. of obs.	Test	Estimated power using asymptotic critical values 5%	Estimated power using asymptotic critical values 1%	Estimated power using estimated critical values 5%	Estimated power using estimated critical values 1%
50	H	0.161(0.0082)	0.063(0.0054)	0.176(0.0085)	0.053(0.0050)
	$LM1$	0.173(0.0084)	0.065(0.0055)	0.225(0.0093)	0.077(0.0059)
	$LM2$	0.790(0.0091)	0.415(0.0110)	0.802(0.0089)	0.513(0.0112)
	$F1$	0.720(0.0100)	0.464(0.0112)		
	$F2$	0.205(0.0090)	0.077(0.0060)		
	$F3$	0.184(0.0087)	0.070(0.0057)		
100	H	0.281(0.0101)	0.141(0.0078)	0.291(0.0102)	0.113(0.0071)
	$LM1$	0.332(0.0105)	0.163(0.0082)	0.366(0.0108)	0.144(0.0079)
	$LM2$	0.993(0.0019)	0.947(0.0050)	0.993(0.0019)	0.958(0.0045)
	$F1$	0.975(0.0035)	0.900(0.0067)		
	$F2$	0.303(0.0103)	0.140(0.0078)		
	$F3$	0.333(0.0105)	0.163(0.0083)		
200	H	0.531(0.0112)	0.322(0.0104)	0.544(0.0111)	0.306(0.0103)
	$LM1$	0.594(0.0110)	0.387(0.0109)	0.621(0.0108)	0.353(0.0107)
	$LM2$	1.000	1.000	1.000	1.000
	$F1$	1.000	0.998(0.0010)		
	$F2$	0.490(0.0112)	0.293(0.0102)		
	$F3$	0.564(0.0111)	0.358(0.0107)		

[a] Number of replications = 2000. Quantities in brackets are estimated standard errors.

Table 8

Tests for heteroskedasticity: Case 4: Random coefficient model, weights = (1, 1).[a]

No. of obs.	Test	Estimated power using asymptotic critical values		Estimated power using estimated critical values	
		5%	1%	5%	1%
50	*H*	0.281(0.0100)	0.152(0.0080)	0.295(0.0102)	0.145(0.0079)
	LM1	0.267(0.0099)	0.170(0.0084)	0.314(0.0104)	0.185(0.0087)
	LM2	0.056(0.0051)	0.009(0.0021)	0.060(0.0053)	0.015(0.0027)
	F1	0.100(0.0067)	0.024(0.0034)		
	F2	0.090(0.0064)	0.027(0.0036)		
	F3	0.081(0.0061)	0.017(0.0028)		
100	*H*	0.560(0.0111)	0.419(0.0110)	0.570(0.0111)	0.385(0.0109)
	LM1	0.567(0.0111)	0.446(0.0111)	0.591(0.0110)	0.424(0.0111)
	LM2	0.097(0.0066)	0.015(0.0027)	0.096(0.0066)	0.022(0.0032)
	F1	0.202(0.0090)	0.083(0.0062)		
	F2	0.202(0.0090)	0.076(0.0059)		
	F3	0.104(0.0068)	0.029(0.0037)		
200	*H*	0.870(0.0075)	0.760(0.0096)	0.877(0.0073)	0.748(0.0097)
	LM1	0.853(0.0079)	0.767(0.0095)	0.862(0.0077)	0.749(0.0097)
	LM2	0.180(0.0086)	0.039(0.0043)	0.178(0.0086)	0.035(0.0041)
	F1	0.369(0.0108)	0.183(0.0086)		
	F2	0.384(0.0109)	0.202(0.0090)		
	F3	0.141(0.0078)	0.048(0.0048)		

[a]See notes to table 7.

ity, both outperform the *H* test substantially. Even *LM1* and the other *F* tests do as well as or better than *H* in this case. The facts that *H* has any power at all here, and likewise that the *OLS* covariance matrix is inconsistent, are attributable principally to the larger variance in X_1 in the second half of the sample.

When the heteroskedasticity arises from a random coefficient model, *H* performs very well. Curiously, *LM1*, which is specifically designed to test against this alternative, does not perform much better than *H*, on average; it outperforms it in most cases, but not in all. When the weights for the random coefficient model are (1, 3) or (3, 1), so that most of the heteroskedasticity is associated with only one of the regressors, the corresponding *F* test performs very well, and somewhat better than *H*.

The results, then, are somewhat mixed. No one test has greatest power against all alternatives. Perhaps the most interesting result is that, in many cases, the power of all the tests is fairly low, even though, as we saw earlier, there is enough heteroskedasticity in the errors to cause serious errors of inference when using *OLS* *t*-statistics. This suggests that a strategy of first testing for heteroskedasticity, and then using either *OLS* or *HC3* depending on the outcome of the test, may *not* be a good one.

Table 9

Tests for heteroskedasticity: Case 6: Random coefficient model, weights = (1, 3).[a]

No. of obs.	Test	Estimated power using asymptotic critical values 5%	Estimated power using asymptotic critical values 1%	Estimated power using estimated critical values 5%	Estimated power using estimated critical values 1%
50	*H*	0.284(0.0101)	0.122(0.0073)	0.310(0.0103)	0.108(0.0069)
	LM1	0.370(0.0108)	0.169(0.0084)	0.451(0.0111)	0.188(0.0087)
	LM2	0.131(0.0075)	0.023(0.0042)	0.138(0.0077)	0.036(0.0042)
	F1	0.257(0.0098)	0.095(0.0065)		
	F2	0.118(0.0072)	0.031(0.0039)		
	F3	0.397(0.0109)	0.189(0.0087)		
100	*H*	0.589(0.0110)	0.357(0.0107)	0.600(0.0110)	0.309(0.0103)
	LM1	0.699(0.0103)	0.475(0.0112)	0.726(0.0100)	0.437(0.0111)
	LM2	0.260(0.0098)	0.081(0.0061)	0.255(0.0097)	0.096(0.0066)
	F1	0.483(0.0112)	0.263(0.0098)		
	F2	0.185(0.0087)	0.067(0.0056)		
	F3	0.713(0.0101)	0.482(0.0112)		
200	*H*	0.915(0.0063)	0.782(0.0092)	0.918(0.0061)	0.761(0.0095)
	LM1	0.946(0.0051)	0.865(0.0077)	0.954(0.0047)	0.848(0.0080)
	LM2	0.519(0.0112)	0.256(0.0098)	0.516(0.0112)	0.248(0.0096)
	F1	0.796(0.0090)	0.597(0.0110)		
	F2	0.318(0.0104)	0.147(0.0079)		
	F3	0.963(0.0042)	0.880(0.0073)		

[a] See notes to table 7.

We investigated the effects of using such a strategy, based on the H test with asymptotic size of 20%, 10% and 5%, for all the cases we studied. One might expect the properties of the resulting pretest t-statistic to be a convex combination of the properties of the *HC3* and *OLS* t-statistics, with weights given by the power of the test. In fact, the pretest t-statistics did not perform as badly as that; they were closer to the *HC3* t-statistics than the power of the test would suggest. This presumably indicates that the H test tends to have power when the heteroskedasticity in the sample is particularly damaging. Nevertheless, whenever there actually was heteroskedasticity, we found that t-statistics based on pretesting were consistently and often substantially less well-behaved than those based on *HC3*. This was most apparent when the size of the test was low and the sample size small, so that the power of the H test was low. Since the cost of using *HC3* instead of *OLS* when heteroskedasticity is absent is apparently not very great (see table 1), it would seem wise to employ t-statistics based on *HC3* even when there is little evidence of heteroskedasticity.

7. Conclusions

We have examined the performance of three modified versions of White's (1980) heteroskedasticity-consistent covariance matrix estimator. All of them can be thought of as in some way derived from the jackknife, and the one which is explicitly the jackknife covariance estimator, *HC3*, always performs better than the other two, which in turn always outperform the original. We have also studied an alternative approach to obtaining reliable inferences in small samples when there is heteroskedasticity of unknown form, namely the Edgeworth approximations of Rothenberg (1984). This approach is a good deal more difficult to implement than using *HC3*, and appears to perform less well than the latter when the sample size is small.

In addition, we have studied the properties of several alternative tests for heteroskedasticity, and found that they often lack power to detect damaging levels of it. This fact, together with our other results, suggests that it may be wise to use *HC3* in preference to the usual *OLS* covariance estimator, even when there is little evidence of heteroskedasticity. This of course is subject to the proviso that the sample size should not be extremely small, nor the design of the $X'X$ matrix extremely unbalanced, so that *HC3* might perform significantly less well than it did in our experiments.

References

Breusch, T.S. and A.R. Pagan, 1979, A simple test for heteroskedasticity and random coefficient variation, Econometrica 47, 1287–1294.

Cragg, J.G., 1983, More efficient estimation in the presence of heteroskedasticity of unknown form, Econometrica 51, 751–763.

Davidson, R. and J.G. MacKinnon, 1981, Efficient estimation of tail-area probabilities in sampling experiments, Economics Letters 8, 73–77.

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 estimators for families of linear regressions, Annals of Mathematical Statistics 34, 447–456.

Goldfeld, S.M. and R.E. Quandt, 1965, Some tests for homoskedasticity, Journal of the American Statistical Association 60, 539–547.

Hinkley, D.V., 1977, Jackknifing in unbalanced situations, Technometrics 19, 285–292.

Horn, S.D., R.A. Horn and D.B. Duncan, 1975, Estimating heteroskedastic variances in linear models, Journal of the American Statistical Association 70, 380–385.

Koenker, R., 1981, A note on studentizing a test for heteroskedasticity, Journal of Econometrics 17, 107–112.

Messer, K. and H. White, 1984, A note on computing the heteroskedasticity consistent covariance matrix using instrumental variable techniques, Oxford Bulletin of Economics and Statistics 46, 181–184.

Nicholls, D.F. and A.R. Pagan, 1983, Heteroskedasticity in models with lagged dependent variables, Econometrica 51, 1233–1242.

Rothenberg, T.J., 1984, Approximate power functions for some robust tests of regression coefficients, Research paper in economics no. 84-1 (University of California, Berkeley, CA).

White, H., 1980, A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity, Econometrica 48, 817–838.