Queen's Economics Department Working Paper No. 616

Heteroskedasticity-Robust Tests in Regression Directions

Russell Davidson
Queen's University

James G. MacKinnon
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1985

Heteroskedasticity-Robust Tests in Regression Directions

Russell Davidson

and

James G. MacKinnon

Department of Economics
Queen's University
Kingston, Ontario, Canada
K7L 3N6

Abstract

We develop simple procedures to test for omitted variables and perform other tests in regression directions, which are asymptotically valid in the presence of heteroskedasticity of unknown form. We examine the asymptotic behavior of these tests and use Edgeworth approximations to study their approximate finite-sample performance. We also present results from several Monte Carlo experiments, which suggest that one family of these tests should always be used in preference to the other.

We are grateful to seminar participants at Duke University, the University of Amsterdam, and the Econometric Society Summer Workshop, Paris, for comments. This research was supported, in part, by grants from the Social Sciences and Humanities Research Council of Canada and the Graduate School of Queen's University. This paper appeared in *Annales de l'INSEE*, Numbers 59/60, 1985, pp. 183–218. This version has far fewer typographical errors.

July, 1985

1. Introduction

A great many of the tests which are routinely used by econometricians are what Davidson and MacKinnon (1985) call tests in regression directions. For such a test, both the null hypothesis and the (possibly implicit) alternative against which the test is constructed are, at least locally, regression models of some sort. Any test in regression directions can thus always be computed as a test for one or more omitted variables, and explicit omitted variables tests are one common example of such tests. Other examples include tests for structural change (Chow, 1960), serial correlation (Durbin, 1970; Godfrey, 1978), and exogeneity (Durbin, 1954; Hausman, 1978); many nonnested hypothesis tests (Davidson and MacKinnon, 1981b, 1982); and differencing specification tests (Plosser, Schwert, and White, 1982; Davidson, Godfrey, and MacKinnon, 1985). Note that many of these are specification tests, in the sense of Hausman (1978), while others are classical tests. When the regression errors are homoskedastic, such tests can always be computed as t or F tests for omitted variables, although they will not always be exact in finite samples. When the regression errors display heteroskedasticity of unknown form, however, such tests are no longer valid even asymptotically.

The results of White (1980) make it clear that asymptotically valid tests can indeed be computed in these cases. Building on those results, we show that doing so is not only possible but indeed remarkably easy. We also show that there are numerous asymptotically equivalent test statistics, which may well have different finite sample properties. That is done in Section 2. In Section 3, we then go on to examine the properties of these test statistics under sequences of local data generating processes, or DGPs. In Section 4, we consider extensions to nonlinear models and models estimated by two-stage least squares. In Section 5, we use second-order Edgeworth approximations to analyze the approximate finite-sample performance of two of the test statistics discussed in Section 2. Finally, in Section 6, we present results from several Monte Carlo experiments. These strongly suggest that one family of our tests should generally be used in preference to the other, and also provide evidence on the usefulness of the Edgeworth approximations.

2. Some Easily Computed Tests

We consider the case of testing a linear regression model against an alternative which includes one or more additional regressors. The null and alternative models are

$$H_0: \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \tag{1}$$

and

$$H_1: \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\gamma}\boldsymbol{u}, \tag{2}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are, respectively, $n \times k$ and $n \times r$ matrices of strictly exogenous variables, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are unknown parameter vectors, and $\boldsymbol{u}$ is a vector of error terms with the properties:

$$\mathrm{E}(u_t) = 0,\ \forall\, t, \quad \mathrm{E}(u_t, u_s) = 0,\ \forall\, t \neq s, \quad \mathrm{E}(u_t^2) = \sigma_t^2 < \sigma_{\mathrm{M}}^2. \tag{3}$$

Various technical assumptions are necessary, which are spelled out in White (1984). The important ones are that $\boldsymbol{X}^\top\boldsymbol{X}/n$, $\boldsymbol{Z}^\top\boldsymbol{Z}/n$, $\boldsymbol{X}^\top\boldsymbol{Z}/n$, and $\boldsymbol{Z}^\top\boldsymbol{M_X}\boldsymbol{Z}/n$ should all tend to fixed matrices with ranks of k, r, $\min(r, k)$, and r, respectively, as $n \to \infty$. Here $\boldsymbol{M_X}$ denotes the projection matrix $\mathbf{I} - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. We shall also let $\boldsymbol{\Omega}$ denote the $n \times n$ diagonal matrix with σ_t^2 as a typical diagonal element.

The hypotheses H_0 and H_1 have been stated in a way which makes it appear that we are solely concerned with classical tests of the hypothesis that $\boldsymbol{\gamma} = \mathbf{0}$. That is not in fact the case. Suppose that, following Hausman (1978) and Holly (1982), we are actually interested in the null hypothesis

$$H_0^*: \quad \boldsymbol{X}^\top\boldsymbol{Z}\boldsymbol{\gamma} = \mathbf{0},$$

which implies that estimation of H_0, will yield unbiased estimates of $\boldsymbol{\beta}$. Although the truth of H_0 implies the truth of H_0^*, the converse is not so. It is evident that H_0^* is equivalent to the hypothesis that

$$\mathrm{E}\big(\boldsymbol{X}^\top\boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{M_X}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{M_X}\boldsymbol{y}\big) = \mathbf{0}. \tag{4}$$

Now make the definition

$$\boldsymbol{Z}^* \equiv \boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{M_X}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{X},$$

dropping any redundant columns if $k > r$, and consider the new alternative hypothesis

$$H_1^*: \quad \boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z}^*\boldsymbol{\gamma}^* + \boldsymbol{u}.$$

It is easily seen that

$$\hat{\boldsymbol{\gamma}}^* = \big(\boldsymbol{X}^\top\boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{M_X}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{M_X}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{M_X}\boldsymbol{y},$$

so that a test of $\boldsymbol{y}^* = \mathbf{0}$ is equivalent to a test of (4), and hence to a test of H_0^*. If $k \geq r$, the number of degrees of freedom of the test will be r, and the classical test of H_0 will coincide with the specification test of H_0^*. However, if $k < r$, the number of degrees of freedom of the test will be k, and the two tests will not coincide. We are of course assuming that the matrix $[\boldsymbol{X} \;\; \boldsymbol{Z}^*]$ has full rank; if it does not, some columns of $\boldsymbol{Z}^*$ will have to be dropped, and the degrees of freedom of the test adjusted accordingly. Everything we say below is just as valid for testing H_0 against H_1^* as for testing it against H_1; the reader must simply substitute $\boldsymbol{Z}^*$ for $\boldsymbol{Z}$ and correct the degrees of freedom as necessary.

It is easy to derive the ordinary F statistic for the null hypothesis that $\boldsymbol{y} = \mathbf{0}$. This statistic is

$$\frac{\boldsymbol{y}^\top\boldsymbol{M_X}\boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{M_X}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{M_X}\boldsymbol{y}/r}{\boldsymbol{y}^\top\boldsymbol{M}_{[\boldsymbol{X}\;\boldsymbol{Z}]}\boldsymbol{y}/(n-k-r}, \tag{5}$$

where $\boldsymbol{M}_{[\boldsymbol{X}\;\boldsymbol{Z}]}$ is analogous to $\boldsymbol{M_X}$, except that it projects off $\boldsymbol{X}$ and $\boldsymbol{Z}$ jointly. It is clear by inspection that, under the assumption $\sigma_t^2 = \sigma^2$ for all t, (5) could have been derived as a statistic for testing the hypothesis:

$$\mathrm{E}(\boldsymbol{y}^\top \boldsymbol{M_X Z}) = \mathbf{0}. \tag{6}$$

Indeed, so far as observables are concerned, this hypothesis and the hypothesis $\boldsymbol{\gamma} = \mathbf{0}$ are indistinguishable.

We now consider how to test the hypothesis (6) when the error terms have the properties (3). It is clear that, under H_0, $\boldsymbol{y}^\top \boldsymbol{M_X Z} = \boldsymbol{u}^\top \boldsymbol{M_X Z}$, which has expectation zero and covariance matrix $\boldsymbol{Z}^\top \boldsymbol{M_X \Omega M_X Z}$. If the error terms u_t were normally distributed, it would then follow immediately that the statistic

$$\boldsymbol{y}^\top \boldsymbol{M_X Z}(\boldsymbol{Z}^\top \boldsymbol{M_X \Omega M_X Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{M_X y} \tag{7}$$

is distributed as $\chi^2(r)$ under H_0. That will be true asymptotically even without a normality assumption, provided that a central limit theorem applies to the vector $n^{-1/2}\boldsymbol{u}^\top \boldsymbol{M_X Z}$, which is indeed the case under our assumptions.

The statistic (7) cannot be computed unless $\boldsymbol{\Omega}$ is known. But the results of White (1980), or very slight extensions of them, imply that

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{Z}^\top \boldsymbol{M_X \ddot{\Omega} M_X Z}\Big) = \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{Z}^\top \boldsymbol{M_X \Omega M_X Z}\Big), \tag{8}$$

where $\ddot{\boldsymbol{\Omega}}$ denotes one of a number of possible estimators for $\boldsymbol{\Omega}$. In particular, we consider $\tilde{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\Omega}}$, which have diagonal elements of $\tilde{u}_t^2$ and $\hat{u}_t^2$, respectively, $\tilde{\boldsymbol{u}}$ being the vector of residuals from OLS estimation of the null hypothesis (1), and $\hat{\boldsymbol{u}}$ being the residual vector from OLS estimation of the alternative (2). A proof of (8) can easily be constructed from results in White (1984), and it is therefore omitted. The result (8) implies that the test statistic (7) will still be asymptotically distributed as $\chi^2(r)$ under H_0 if $\boldsymbol{\Omega}$ is replaced by $\tilde{\boldsymbol{\Omega}}$, $\hat{\boldsymbol{\Omega}}$, or any one of a number of similar estimators. Of course, none of these estimators will be consistent for $\boldsymbol{\Omega}$; what matters is that we can estimate $\boldsymbol{Z}^\top \boldsymbol{M_X \Omega M_X Z}/n$ consistently. This is analogous to White's (1980) result that in order to obtain a consistent covariance matrix estimate, what matters is the ability to estimate $\boldsymbol{X}^\top \boldsymbol{\Omega X}/n$ consistently.

Feasible versions of the test statistic (7), which use $\ddot{\boldsymbol{\Omega}}$ instead of $\boldsymbol{\Omega}$, can be computed very easily. Let $\ddot{\sigma}_t$ denote the square root of the t^{th} diagonal element of $\ddot{\boldsymbol{\Omega}}$, and consider the regression

$$y_t/\ddot{\sigma}_t = \ddot{\sigma}_t(\boldsymbol{M_X Z})_t \boldsymbol{\gamma} + \text{errors}. \tag{9}$$

Here $(\boldsymbol{M_X Z})_t$ denotes the t^{th} row of the matrix $\boldsymbol{M_X Z}$, which must be computed first by projecting the columns of $\boldsymbol{Z}$ off $\boldsymbol{X}$. The explained sum of squares from regression (9), taken around zero and not around the mean of the regressand, is easily seen to be the test statistic we want to compute. This regression is similar to, but simpler than, the IV regression proposed by Messer and White (1984) as a procedure for computing heteroskedasticity-consistent standard error estimates. It suffers from two minor disadvantages. First of all, if $\ddot{\sigma}_t$ happens to be exactly zero, as might be the case if an observation has been dummied out, the regressand of (9) cannot be computed. Following a suggestion of Messer and White (1984), one would then have to replace any $\ddot{\sigma}_t$ which were zero by some small number, say .0001. Secondly, few regression packages print the explained sum of squares of a regression about zero, so that calculating the test statistic will usually involve a few auxiliary calculations.

An even simpler artificial regression is available if we restrict our attention to the case of $\boldsymbol{\Omega} = \mathrm{diag}(\tilde{u}_t)$. The regression is

$$\boldsymbol{\iota} = \tilde{\boldsymbol{U}}\boldsymbol{M_X Z\gamma} + \text{errors}, \tag{10}$$

where $\boldsymbol{\iota}$ is an n–vector of ones, and $\tilde{\boldsymbol{U}} = \mathrm{diag}(\tilde{u}_t)$. The explained sum of squares from this regression, which is equal to n minus the sum of squared residuals, is plainly equal to

$$\begin{aligned}&\boldsymbol{\iota}^\top\tilde{\boldsymbol{U}}\boldsymbol{M_X Z}(\boldsymbol{Z}^\top\boldsymbol{M_X}\tilde{\boldsymbol{U}}\tilde{\boldsymbol{U}}\boldsymbol{M_X Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{M_X}\tilde{\boldsymbol{U}}\boldsymbol{\iota}\\ &\quad = \tilde{\boldsymbol{u}}^\top\boldsymbol{M_X Z}(\boldsymbol{Z}^\top\boldsymbol{M_X}\tilde{\boldsymbol{\Omega}}\boldsymbol{M_X Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{M_X}\tilde{\boldsymbol{u}}.\end{aligned} \tag{11}$$

Since $\tilde{\boldsymbol{u}}^\top\boldsymbol{M_X Z} = \boldsymbol{y}^\top\boldsymbol{M_X Z}$, this is the test statistic we are seeking to compute. The artificial regression (10) avoids both of the disadvantages associated with (9), and, as we shall see in Section 6, restricting attention to $\tilde{\boldsymbol{\Omega}}$ is by no means a bad thing to do.

A great many choices are potentially available for $\ddot{\boldsymbol{\Omega}}$. MacKinnon and White (1985) considered four possibilities, all implicitly based on the unrestricted residuals $\hat{u}_t$, since their paper was concerned with the estimation of heteroskedasticity-consistent covariance matrices. The simplest of these, which we shall call HC_0, uses $\hat{\boldsymbol{\Omega}}$. A slightly more complicated version, which we shall call HC_1, uses $(n/(n-k-r))\hat{\boldsymbol{\Omega}}$, the first factor being a crude correction for degrees of freedom lost in estimating $\hat{\boldsymbol{\Omega}}$. A third version, called HC_2, uses $\mathrm{diag}(\hat{u}_t^2/m_{tt})$, where m_{tt} is the t^{th} diagonal element of the matrix $\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}$; this amounts to using a more sophisticated degrees of freedom correction. The fourth version, called HC_3, utilizes the jackknife covariance matrix estimator. It is somewhat more complicated than the others and will not be described here.

In Monte Carlo work, MacKinnon and White (1985) found that t statistics based on all of the heteroskedasticity-consistent covariance matrices were too prone to reject the null when it was true. In order of diminishing reliability, the ranking was always: HC_3, HC_2, HC_1, HC_0, with HC_3 always performing very much better than HC_0. The superiority of HC_3 is somewhat unfortunate in the present context, because the estimate of $\boldsymbol{\Omega}$ implicit in HC_3 is not (quite) diagonal, so that the artificial regression (9) could not be used to compute the test statistic.

It is easy to construct test statistics based on heteroskedasticity-consistent covariance matrices analogous to HC_0 through HC_3 that use restricted residuals. The simplest, which uses a matrix HCR_0 similar to HC_0, is the test statistic (11). Multiplying (11) by $(n-k)/n$ would be equivalent to using a matrix HCR_1 similar to HC_1. A still more complicated statistic would employ a matrix HCR_2 that uses $\ddot{\sigma}_t^2 = \tilde{u}_t^2/(1-k_{tt})$, where k_{tt} is the t^{th} diagonal element of $\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. In this case, the test statistic could be computed by regression (9) but not by regression (10). One could also construct a statistic based on the jackknife estimator HCR_3, but it would be relatively hard to compute.

In Section 5 below, we study the finite sample properties of tests based on HC_0 and HCR_0 using Edgeworth approximations. Then, in Section 6, we use Monte Carlo methods to study the finite sample performance of all the test statistics mentioned above. We also evaluate the usefulness of the Edgeworth approximations.

3. Asymptotic Behavior of the Tests

In this section, we consider the asymptotic behavior of the test statistics discussed above under sequences of local data generating processes (DGPs). The generic test statistic we are interested in may be written as

$$\boldsymbol{y}^\top \boldsymbol{M_X Z}(\boldsymbol{Z}^\top \boldsymbol{M_X} \ddot{\boldsymbol{\Omega}} \boldsymbol{M_X Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{M_X y}, \tag{12}$$

where $\ddot{\boldsymbol{\Omega}}$ may denote any estimator of $\boldsymbol{\Omega}$ which satisfies (8). The sequence of local DGPs may be written as

$$\boldsymbol{y} = \boldsymbol{X\beta}_0 + n^{-1/2}\boldsymbol{W\delta}_0 + \boldsymbol{u}, \;\; \mathrm{E}(\boldsymbol{u}) = \boldsymbol{0}, \;\; \mathrm{E}(\boldsymbol{uu}^\top) = \boldsymbol{\Omega}_0. \tag{13}$$

The matrix $\boldsymbol{\Omega}_0$ is of course assumed to be diagonal. Strictly speaking, we should write the vectors $\boldsymbol{u}$ and $\boldsymbol{y}$ and the matrices $\boldsymbol{X}$, $\boldsymbol{W}$, and $\boldsymbol{\Omega}_0$ as functions of n, the sample size. In addition to the assumptions made previously, we assume that $\boldsymbol{\delta}_0^\top \boldsymbol{W}^\top \boldsymbol{W\delta}_0/n$, $\boldsymbol{\delta}_0^\top \boldsymbol{W}^\top \boldsymbol{X}/n$, and $\boldsymbol{\delta}_0^\top \boldsymbol{W}^\top \boldsymbol{Z}/n$ all tend to fixed matrices as $n \to \infty$.

The sequence of local DGPs (13) allows for the truth to differ from the null hypothesis in any regression direction, since $\boldsymbol{W\delta}_0$ could be almost anything. As n gets large, the DGP approaches a particular case of the null hypothesis H_0, where $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\boldsymbol{\Omega} = \boldsymbol{\Omega}_0$. It is of course possible for the DGP to lie within the alternative H_1; that will be the case if $\boldsymbol{W\delta}_0$ lies entirely in the space spanned by $\boldsymbol{X}$ and $\boldsymbol{Z}$ jointly.

We now define the matrices $\boldsymbol{P}_0$ and $\ddot{\boldsymbol{P}}$ so that

$$\boldsymbol{P}_0\boldsymbol{P}_0^\top = (\boldsymbol{Z}^\top \boldsymbol{M_X}\boldsymbol{\Omega}_0 \boldsymbol{M_X Z})^{-1} \quad \text{and} \quad \ddot{\boldsymbol{P}}\ddot{\boldsymbol{P}}^\top = (\boldsymbol{Z}^\top \boldsymbol{M_X} \ddot{\boldsymbol{\Omega}} \boldsymbol{M_X Z})^{-1}, \tag{14}$$

noting that $\boldsymbol{P}_0$ and $\ddot{\boldsymbol{P}}$ are $r \times r$ matrices and $O(n^{-1/2})$. This allows us to rewrite the test statistic (12) as

$$\boldsymbol{y}^\top \boldsymbol{M_X Z}\ddot{\boldsymbol{P}}\ddot{\boldsymbol{P}}^\top \boldsymbol{Z}^\top \boldsymbol{M_X y}. \tag{15}$$

From (13) and (14), we see that

$$\boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X y} = n^{-1/2}\boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X W\delta}_0 + \boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X u}, \tag{16}$$

and it is evident that both terms in (16) are $O(1)$. Under the assumptions we have made or alluded to–that is, those of White (1984)–we can clearly apply a central limit theorem to $\boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X u}$. It follows that the vector $\boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X y}$ is asymptotically normal with mean vector

$$n^{-1/2}\boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X W}\boldsymbol{\delta}_0$$

and covariance matrix

$$\mathrm{E}(\boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X u u}^\top \boldsymbol{M_X Z P}_0) = \boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X}\boldsymbol{\Omega}_0 \boldsymbol{M_X Z P}_0 = \mathbf{I}_r. \tag{17}$$

If the DGP were a special case of H_0, it would follow immediately from (8) that $\ddot{\boldsymbol{P}} \to \boldsymbol{P}_0$ as $n \to \infty$. That remains true under (13), because a result equivalent to (8) still holds.[1] Hence the test statistic (15) is asymptotically equivalent to

$$\boldsymbol{y}^\top \boldsymbol{M_X Z P}_0 \boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X y}, \tag{18}$$

which is simply the sum of squares of the elements of the random vector $\boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X y}$. Since those elements are asymptotically normal with variance one, we conclude that the test statistic is asymptotically distributed as noncentral Chi-squared with r degrees of freedom and non-centrality parameter, or NCP,

$$\begin{aligned} &\tfrac{1}{n}\boldsymbol{\delta}_0^\top \boldsymbol{W}^\top \boldsymbol{M_X Z P}_0 \boldsymbol{P}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X W}\boldsymbol{\delta}_0 \\ &\quad = \tfrac{1}{n}\boldsymbol{\delta}_0^\top \boldsymbol{W}^\top \boldsymbol{M_X Z}(\boldsymbol{Z}^\top \boldsymbol{M_X}\boldsymbol{\Omega}_0 \boldsymbol{M_X Z})^{-1}\boldsymbol{Z}^\top \boldsymbol{M_X W}\boldsymbol{\delta}_0. \end{aligned} \tag{19}$$

If in fact the error terms are homoskedastic, so that $\boldsymbol{\Omega}_0 = \sigma_0^2\mathbf{I}$, the NCP (19) will simplify to

$$\tfrac{1}{n}\boldsymbol{\delta}_0^\top \boldsymbol{W}^\top \boldsymbol{M_X Z}(\boldsymbol{Z}^\top \boldsymbol{M_X Z})^{-1}\boldsymbol{Z}^\top \boldsymbol{M_X W}\boldsymbol{\delta}_0/\sigma_0^2, \tag{20}$$

which is identical to the NCP for the Chi-squared version of the ordinary F statistic (5). Thus, when there is in fact no heteroskedasticity, it is asymptotically costless to use a heteroskedasticity-robust test.

When there is heteroskedasticity, (19) makes it clear that the power of the test will depend on the nature of $\boldsymbol{\Omega}_0$. Multiplying all elements of $\boldsymbol{\Omega}_0$ by a factor of λ will of course reduce the NCP by a factor of $1/\lambda$, as in the homoskedastic case. But changes in the pattern of heteroskedasticity, even if they do not affect the average value of $\boldsymbol{\Omega}_0$, may well affect $\boldsymbol{Z}^\top \boldsymbol{M_X}\boldsymbol{\Omega}_0 \boldsymbol{M_X Z}$ and hence affect the power of the test. In the homoskedastic case, if the DGP lies within H_1, so that $\boldsymbol{W}\boldsymbol{\delta}_0$ can be written as $\boldsymbol{M_X Z}\boldsymbol{\gamma}_0$ for some $\boldsymbol{\gamma}_0$, the NCP (20) simplifies to

$$\tfrac{1}{n}\boldsymbol{\gamma}_0^\top \boldsymbol{Z}^\top \boldsymbol{M_X Z}\boldsymbol{\gamma}_0/\sigma_0^2. \tag{21}$$

However, no such simplification occurs for (19).

[1] This statement is quite evidently true from the calculations in Section 5.

4. Some Extensions

The basic ideas of Section 2 can readily be extended to a number of cases more complicated than the one dealt with there. In this section, we first derive heteroskedasticity-robust tests for linear models which have been estimated by two-stage least squares, and then for univariate models which can be estimated by nonlinear least squares. Extensions to other cases, including multivariate models, are of course possible, but they will not be attempted here.

The first case we consider is where H_0 and H_1 are still given by (1) and (2), but some columns of $\boldsymbol{X}$ and/or some columns of $\boldsymbol{Z}$ are asymptotically correlated with $\boldsymbol{u}$. We suppose that there exists a matrix of instruments, $\boldsymbol{W}$, which includes all the columns of $\boldsymbol{X}$ and $\boldsymbol{Z}$ that are valid instruments, plus enough other valid instruments so that 2SLS estimation of both H_0 and H_1 is feasible. For simplicity, we do not consider the possibility that different instruments might be used in the estimation of H_0 and H_1.

The second stage regression, which yields 2SLS estimates of H_1, is

$$\boldsymbol{y} = \boldsymbol{P_W X\beta} + \boldsymbol{P_W Z\gamma} + \text{errors}, \tag{22}$$

where $\boldsymbol{P_W} = \boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top$. As shown by Lovell (1963) and others, the estimate of $\boldsymbol{\gamma}$ from (22) will be identical to the estimate from the regression

$$\boldsymbol{M_{PX} y} = \boldsymbol{M_{PX} Z\gamma} + \text{errors}, \tag{23}$$

where $\boldsymbol{M_{PX} y} = \mathbf{I} - \boldsymbol{P_W X}(\boldsymbol{X}^\top\boldsymbol{P_W X})^{-1}\boldsymbol{X}^\top\boldsymbol{P_W}$. It is thus evident that the hypothesis $\boldsymbol{\gamma} = \mathbf{0}$ is observationally equivalent to the hypothesis

$$\lim_{n\to\infty} \mathrm{E}(n^{-1/2}\boldsymbol{y}^\top\boldsymbol{M_{PX}P_W Z}) = \mathbf{0}. \tag{24}$$

Under the null hypothesis, $\boldsymbol{y}^\top\boldsymbol{M_{PX}P_W Z} = \boldsymbol{u}^\top\boldsymbol{M_{PX}P_W Z}$. The expectation of $n^{-1/2}$ times this quantity clearly tends to zero, and the asymptotic covariance matrix of $n^{-1/2}$ times it is

$$\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{P_W M_{PX}\Omega}_0\boldsymbol{M_{PX}P_W Z}, \tag{25}$$

where $\boldsymbol{\Omega}_0$ is the covariance matrix of $\boldsymbol{u}$. Hence it is obvious that the test statistic

$$\boldsymbol{y}^\top\boldsymbol{M_{PX}P_W Z}(\boldsymbol{Z}^\top\boldsymbol{P_W M_{PX}}\ddot{\boldsymbol{\Omega}}\boldsymbol{M_{PX}P_W Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{P_W M_{PX} y} \tag{26}$$

will be asymptotically $\chi^2(r)$ under H_0, where $\ddot{\boldsymbol{\Omega}}$ is any estimator of $\boldsymbol{\Omega}$ for which a result analogous to (8) holds. The statistic (26) may readily be computed by artificial regressions similar to (9) and (10). In particular, suppose that $\ddot{\boldsymbol{\Omega}} = \tilde{\boldsymbol{\Omega}} = \mathrm{diag}(\tilde{u}_t^2)$, where $\tilde{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}$ (*not* $\boldsymbol{y} - \boldsymbol{P_W X}\tilde{\boldsymbol{\beta}}$), $\tilde{\boldsymbol{\beta}}$ being the 2SLS estimate of $\boldsymbol{\beta}$ from estimation of H_0. Then, letting $\tilde{\boldsymbol{U}}$ denote $\mathrm{diag}(\tilde{u}_t)$, as before, we see that n minus the sum of squared residuals from the artificial regression

$$\boldsymbol{\iota} = \tilde{\boldsymbol{U}}\boldsymbol{M_{PX}P_W Z\gamma} + \text{errors} \tag{27}$$

will yield a version of the test statistic (26) that is analogous to (11).

We now consider the case of nonlinear regression models, writing the general model as

$$H_1: \quad \boldsymbol{y} = \boldsymbol{f}(\boldsymbol{\beta}, \boldsymbol{\gamma}) + \boldsymbol{u}, \tag{28}$$

where $\boldsymbol{f}(\cdot)$ represents a vector of twice continuously differentiable functions $f_t(\cdot)$, which depend implicitly on exogenous and/or predetermined variables, and explicitly on parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. The vector of error terms is assumed to have the same properties as the errors adhering to (1) and (2). The null hypothesis to be tested is $\boldsymbol{\gamma} = \mathbf{0}$, and in view of the nonlinearity of the model, we shall restrict our attention to tests which do not require estimation under the alternative.

Under homoskedasticity, a particularly easy test statistic to compute is the Lagrange Multiplier statistic, which is equal to the explained sum of squares (or n times the uncentered R^2) from the artificial regression

$$(1/\tilde{\sigma})(\boldsymbol{y} - \tilde{\boldsymbol{f}}) = \tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}\boldsymbol{b} + \tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}\boldsymbol{c} + \text{errors}, \tag{29}$$

where $\tilde{\boldsymbol{f}} \equiv \boldsymbol{f}(\tilde{\boldsymbol{\beta}}, \mathbf{0})$, $\tilde{\sigma}^2 = (1/n)(\boldsymbol{y} - \tilde{\boldsymbol{f}})^\top(\boldsymbol{y} - \tilde{\boldsymbol{f}})$, and $\tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}$ are matrices of derivatives of $\boldsymbol{f}(\cdot)$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively, evaluated at the restricted estimates $(\tilde{\boldsymbol{\beta}}, \mathbf{0})$. The test statistic may be written explicitly as

$$(1/\tilde{\sigma}^2)(\boldsymbol{y} - \tilde{\boldsymbol{f}})^\top \tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} \tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}} (\tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}^\top \tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} \tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}})^{-1} \tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}^\top \tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} (\boldsymbol{y} - \tilde{\boldsymbol{f}}), \tag{30}$$

where $\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} \equiv \mathbf{I} - \tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}(\tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}^\top \tilde{\boldsymbol{F}}_{\boldsymbol{\beta}})^{-1}\tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}^\top$. It will be asymptotically distributed as $\chi^2(r)$ under H_0. For more details, see Engle (1982) or Davidson and MacKinnon (1984).

Under H_0 and sequences of local DGPs, the test statistic (30) tends to

$$(1/\sigma^2)\boldsymbol{u}^\top \boldsymbol{M}_{\boldsymbol{\beta}} \boldsymbol{F}_{\boldsymbol{\gamma}} (\boldsymbol{F}_{\boldsymbol{\gamma}}^\top \boldsymbol{M}_{\boldsymbol{\beta}} \boldsymbol{F}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{F}_{\boldsymbol{\gamma}}^\top \boldsymbol{M}_{\boldsymbol{\beta}} \boldsymbol{u}$$

as $n \to \infty$. Hence it is clear that (30) is really testing the hypothesis that

$$\lim_{n \to \infty} \mathrm{E}(n^{-1/2}\boldsymbol{u}^\top \boldsymbol{M}_{\boldsymbol{\beta}} \boldsymbol{F}_{\boldsymbol{\gamma}}) = \mathbf{0}. \tag{31}$$

Under heteroskedasticity, the variance-covariance matrix of $\boldsymbol{u}^\top \boldsymbol{M}_{\boldsymbol{\beta}} \boldsymbol{F}_{\boldsymbol{\gamma}}$ is

$$\boldsymbol{F}_{\boldsymbol{\gamma}}^\top \boldsymbol{M}_{\boldsymbol{\beta}} \boldsymbol{\Omega} \boldsymbol{M}_{\boldsymbol{\beta}} \boldsymbol{F}_{\boldsymbol{\gamma}},$$

and so it is clear that an asymptotically valid test statistic is

$$(\boldsymbol{y} - \tilde{\boldsymbol{f}})^\top \tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} \tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}} (\tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}^\top \tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} \tilde{\boldsymbol{\Omega}} \tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} \tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}})^{-1} \tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}}^\top \tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} (\boldsymbol{y} - \tilde{\boldsymbol{f}}), \tag{32}$$

where $\tilde{\boldsymbol{\Omega}} = \mathrm{diag}(\tilde{u}_t^2)$ as usual.

The test statistic (32) is simply n minus the sum of squared residuals from the artificial regression.

$$\boldsymbol{\iota} = \tilde{\boldsymbol{U}} \tilde{\boldsymbol{M}}_{\boldsymbol{\beta}} \tilde{\boldsymbol{F}}_{\boldsymbol{\gamma}} \boldsymbol{c} + \text{errors}, \tag{33}$$

where $\tilde{\boldsymbol{u}} = \mathrm{diag}(\tilde{u}_t)$. Thus in order to compute the test statistic, it is simply necessary to do the following:

1. Obtain nonlinear least squares estimates $\tilde{\boldsymbol{\beta}}$ under H_0, and retain $\tilde{\boldsymbol{u}} = \boldsymbol{y} - \tilde{\boldsymbol{f}}$.
2. Calculate the matrices of derivatives $\boldsymbol{F}_{\boldsymbol{\beta}}$ and $\boldsymbol{F}_{\gamma}$ and evaluate them at $(\tilde{\boldsymbol{\beta}}, \mathbf{0})$ to obtain $\tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{F}}_{\gamma}$.
3. Regress the columns of $\tilde{\boldsymbol{F}}_{\gamma}$ on $\tilde{\boldsymbol{F}}_{\boldsymbol{\beta}}$ to obtain residuals $\tilde{\boldsymbol{M}}_{\boldsymbol{\beta}}\tilde{\boldsymbol{F}}_{\gamma}$.
4. Run the artificial regression (33), and compute n minus the sum of squared residuals, which will be asymptotically distributed as $\chi^2(r)$ under H_0.

5. Finite Sample Corrections to the Behavior of the Tests

In this section, we make use of Edgeworth expansions of the distributions of asymptotic t statistics based on the covariance matrices HC_0 and HCR_0 in order to obtain corrections to their asymptotic distributions. It is possible to perform those expansions not only for the case in which the null hypothesis is satisfied, but also for a sequence of local DGPs such as (13). We begin with a general result, similar to one found in Rothenberg (1984), who treats a wider variety of problems than we do in this paper, but in less detail.

THEOREM: For each positive integer n, consider the random variable

$$x_n \equiv \frac{n^{-1/2}\boldsymbol{z}_n^\top \boldsymbol{y}_n}{(n^{-1}\boldsymbol{y}_n^\top \boldsymbol{A}_n \boldsymbol{y}_n)^{1/2}} \tag{34}$$

computed from a random n–vector $\boldsymbol{y}_n$ that satisfies the equation

$$\boldsymbol{y}_n = n^{-1/2}\boldsymbol{w}_n\delta + \boldsymbol{u}_n, \tag{35}$$

in which the random n–vector $\boldsymbol{u}_n$ is distributed as $\mathrm{N}(\mathbf{0}, \mathbf{I}_n)$, where $\mathbf{I}_n$ is an $n \times n$ identity matrix, the non-random n–vector $\boldsymbol{w}_n$ satisfies $\boldsymbol{w}_n^\top \boldsymbol{w}_n = n$, and δ is a scalar independent of n. Further, let the non-random n–vector $\boldsymbol{z}_n$ satisfy the conditions $\boldsymbol{z}_n^\top \boldsymbol{z}_n = n$ and $\boldsymbol{w}_n^\top \boldsymbol{z}_n = \cos\theta$. Let the non-random $n \times n$ symmetric matrix $\boldsymbol{A}_n$ be such that the five quantities

$$\begin{aligned} a_n &= n^{-1}\boldsymbol{z}_n^\top \boldsymbol{A}_n \boldsymbol{z}_n \\ b_n &= n^{-1}\mathrm{Tr}(\boldsymbol{A}_n^2) \\ c_n &= n^{-1}\boldsymbol{w}_n^\top \boldsymbol{A}_n \boldsymbol{z}_n \\ d_n &= n^{-1}\boldsymbol{w}_n^\top \boldsymbol{A}_n \boldsymbol{w}_n \\ e_n &= \mathrm{Tr}\boldsymbol{A}_n - n \end{aligned} \tag{36}$$

are all $O(1)$ as $n \to \infty$.

Then the characteristic function of x_n given by $\psi(t) \equiv \mathrm{E}(e^{itx_n})$, equals

$$e^{it\delta\cos\theta}e^{-t^2/2}\Big(1 + \frac{1}{n}\sum_{r=1}^{4}(-it)^r a_{nr}\Big) + o(n^{-1}) \tag{37}$$

where the four functions a_{nr} are defined by

$$\begin{aligned}
a_{n1} &= c_n\delta + \Big(\frac{1}{2}e_n - \frac{3}{4}b_n\Big)\,\delta\cos\theta + \frac{1}{2}d_n\delta^3\cos\theta;\\
a_{n2} &= b_n - a_n - \frac{1}{2}e_n - \frac{1}{2}d_n\delta^2 + \frac{1}{4}b_n\delta^2\cos^2\theta - c_n\delta^2\cos\theta;\\
a_{n3} &= c_n\delta + \frac{1}{2}(a_n - b_n)\delta\cos\theta;\\
a_{n4} &= \frac{1}{4}b_n - \frac{1}{2}a_n.
\end{aligned} \tag{38}$$

REMARK: The a_{ni} can readily be interpreted in terms of the moments or cumulants of the distribution of x_n. In fact,

$$\mathrm{E}(x_n) = \delta\cos\theta - n^{-1}a_{n1} + o(n^{-1})$$

and the variance of x_n is

$$1 + 2n^{-1}a_{n2} + o(n^{-1}).$$

Similarly, the skewness of x_n is

$$\mathrm{E}\big(x_n - \mathrm{E}(x_n)\big)^3\Big/\big(\mathrm{Var}(x_n)\big)^{3/2} = -6n^{-1}a_{n3} + o(n^{-1}),$$

and the kurtosis of x_n is

$$\mathrm{E}\big(x_n - \mathrm{E}(x_n)\big)^4\Big/\big(\mathrm{Var}(x_n)\big)^2 - 3 = 24n^{-1}a_{n4} + o(n^{-1}).$$

Note also that if $\delta = 0$, so that $\boldsymbol{y}_n$ is just white noise, the mean and skewness of x_n are zero to order n^{-1}, while to this order the variance and kurtosis differ from their limits as $n \to \infty$, which are 1 and 0, respectively, by

$$n^{-1}(2b_n - 2a_n - e_n) \quad \text{and} \quad 6n^{-1}(b_n - 2a_n). \tag{39}$$

The proof of this theorem is in the Appendix.

The result of the above theorem is readily applied to situations in which the vector $\boldsymbol{y}$ (subscripts n will henceforth be dropped unless they are essential for comprehension) is generated by a sequence of local DGPs similar to (13), with the disturbances $\boldsymbol{u}$ subject to a normality requirement. This sequence is

$$\boldsymbol{y} = \boldsymbol{X\beta}_0 + n^{-1/2}\boldsymbol{w}\delta + \boldsymbol{\Omega}_0^{1/2}\boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_n). \tag{40}$$

The random variable x is then defined by

$$x = \frac{n^{-1/2}\boldsymbol{z}^\top\boldsymbol{M_X y}}{(n^{-1}\boldsymbol{y}^\top\boldsymbol{M_X A M_X y})^{1/2}}. \tag{41}$$

We also define $\boldsymbol{\eta} \equiv \boldsymbol{\Omega}_0^{-1/2}(\boldsymbol{M_X y} + \boldsymbol{P_X}\boldsymbol{\Omega}_0^{1/2}\boldsymbol{\varepsilon})$, where $\boldsymbol{P_X} = \mathbf{I} - \boldsymbol{M_X}$. Thus $\boldsymbol{\eta}$ obeys the equation

$$\boldsymbol{\eta} = n^{-1/2}(\boldsymbol{\Omega}_0^{-1/2}\boldsymbol{w})\delta + \boldsymbol{\varepsilon}, \tag{42}$$

and we renormalize $\boldsymbol{w}$ so that

$$\boldsymbol{w}^\top \boldsymbol{M_X}\boldsymbol{\Omega}_0^{-1}\boldsymbol{M_X w} = n. \tag{43}$$

Then the random variable x of (41) can be expressed in terms of $\boldsymbol{\eta}$ as follows:

$$x = \frac{n^{-1/2}\boldsymbol{z}^\top \boldsymbol{M_X}\boldsymbol{\Omega}_0^{1/2}\boldsymbol{\eta}}{(n^{-1}\boldsymbol{\eta}^\top\boldsymbol{\Omega}_0^{1/2}\boldsymbol{M_X A M_X}\boldsymbol{\Omega}_0^{1/2}\boldsymbol{\eta})^{1/2}}. \tag{44}$$

The required normalizations are

$$n^{-1}\boldsymbol{z}^\top\boldsymbol{M_X}\boldsymbol{\Omega}_0\boldsymbol{M_X z} = 1 \quad \text{and} \quad n^{-1}\mathrm{Tr}(\boldsymbol{\Omega}_0\boldsymbol{M_X A M_X}) = 1 + O(n^{-1}). \tag{45}$$

We may now express the asymptotic t statistics based on HC_0 and HCR_0 in the form (44). From (12), it is easy to see that these statistics have the general form

$$\frac{\boldsymbol{y M_X z}}{(\boldsymbol{z}^\top\boldsymbol{M_X}\ddot{\boldsymbol{\Omega}}\boldsymbol{M_X z})^{1/2}}$$

We replace $\ddot{\boldsymbol{\Omega}}$ by $\hat{\boldsymbol{\Omega}}$ for HC_0 and by $\tilde{\boldsymbol{\Omega}}$ for HCR_0, where

$$\hat{\boldsymbol{\Omega}} = \mathrm{diag}\big((\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{y})_t^2\big) \quad \text{and} \quad \tilde{\boldsymbol{\Omega}} = \mathrm{diag}\big((\boldsymbol{M_X y})_t^2\big).$$

Let $\boldsymbol{D} \equiv \mathrm{diag}\big((\boldsymbol{M_X z})_t^2\big)$. Then, for HC_0,

$$\boldsymbol{z}^\top\boldsymbol{M_X}\hat{\boldsymbol{\Omega}}\boldsymbol{M_X z} = \boldsymbol{y}^\top\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{y} = \boldsymbol{\eta}^\top\boldsymbol{\Omega}_0^{1/2}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{\Omega}_0^{1/2}\boldsymbol{\eta}. \tag{46}$$

Similarly, for HCR_0,

$$\boldsymbol{z}^\top\boldsymbol{M_X}\tilde{\boldsymbol{\Omega}}\boldsymbol{M_X z} = \boldsymbol{y}^\top\boldsymbol{M_X D M_X y} = \boldsymbol{\eta}^\top\boldsymbol{\Omega}_0^{1/2}\boldsymbol{M_X D M_X}\boldsymbol{\Omega}_0^{1/2}\boldsymbol{\eta}. \tag{47}$$

The expressions in (46) and (47) are of the correct form to be denominators in (44), since

$$\mathrm{Tr}(\boldsymbol{\Omega}_0\boldsymbol{M_X D M_X z}) = \mathrm{Tr}(\boldsymbol{\Omega}_0\boldsymbol{D}) + O(1) = \boldsymbol{z}^\top\boldsymbol{M_X}\boldsymbol{\Omega}_0\boldsymbol{M_X z} + O(1) = n + O(1)$$

by (45). A similar argument applies to $\mathrm{Tr}(\boldsymbol{\Omega}_0\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]})$.

Expression (37) for the characteristic function can now be worked out for the t statistics based on HC_0 and HCR_0. It is enough to give expressions for the quantities a, b, c, d, and e and for the angle θ. From the definitions, we have for HC_0:

$$\begin{aligned} a &= n^{-1}\boldsymbol{z}^\top \boldsymbol{M_X}\boldsymbol{\Omega}_0\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{\Omega}_0\boldsymbol{M_X}\boldsymbol{z};\\ b &= n^{-1}\mathrm{Tr}\big(\boldsymbol{\Omega}_0\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{\Omega}_0\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\big);\\ c &= n^{-1}\boldsymbol{w}^\top\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{\Omega}_0\boldsymbol{M_X}\boldsymbol{z};\\ d &= n^{-1}\boldsymbol{w}^\top\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{w};\\ e &= \mathrm{Tr}\big(\boldsymbol{\Omega}_0\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\big) - n;\\ \cos\theta &= n^{-1}\boldsymbol{w}^\top\boldsymbol{M_X}\boldsymbol{z}. \end{aligned} \tag{48}$$

For HCR_0, it suffices to replace $\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}$ everywhere by $\boldsymbol{M_X}$.

For dealing with questions of the size or power of test statistics that use HC_0 and HCR_0, we need to pass from the characteristic function (37) to an expression for the density of the statistics. It is straightforward to show that to (37) corresponds the density

$$\phi(x-\delta\cos\theta)\Big(1+\frac{1}{n}\sum_{r=1}^{4}a_rH_r(x-\delta\cos\theta)\Big)+o(n^{-1}), \tag{49}$$

where $\phi(\cdot)$ is the standard normal density and $H_r(\cdot)$ is the r^{th} Hermite polynomial, defined by the relation

$$\frac{d^r\phi(x)}{dx^r} = \phi(x)H_r(x). \tag{50}$$

In particular, we have

$$\begin{aligned} H_0(x) &= 1;\\ H_1(x) &= -x;\\ H_2(x) &= x^2-1;\\ H_3(x) &= 3x-x^3;\\ H_4(x) &= x^4-6x^2+3. \end{aligned} \tag{51}$$

For a random variable x with density (49), the probability in the tail where $x > y$ is then

$$\Phi\big(-(y-\delta\cos\theta)\big) - \frac{1}{n}\phi(y-\delta\cos\theta)\sum_{r=1}^{4}a_rH_{r-1}(x-\delta\cos\theta), \tag{52}$$

with $\Phi(\cdot)$ the distribution function of the standard normal. Now let x^*_α be the critical value for a one-sided test of size α, where $0 < \alpha < 1$, based on a statistic with tail probability (52) for $\delta = 0$, and let x_α be the same thing for a test based on a statistic that follows the $\mathrm{N}(0,1)$ distribution Then, from (38) and (51),

$$\begin{aligned} \alpha = \Phi(-x^*_\alpha) - \frac{1}{n}\phi(x^*_\alpha)\bigg(&x^*_\alpha\Big(a-b+\frac{1}{2}e\Big)\\ &+(3x^*_\alpha-(x^*_\alpha)^3)\Big(\frac{1}{4}b-\frac{1}{2}a\Big)\bigg)+o(n^{-1}). \end{aligned} \tag{53}$$

This equation may either be solved directly (numerically) for x^*_α, or else a Taylor approximation can be made around x_α, with the result

$$x^*_\alpha = x_\alpha - \frac{1}{n}\Big(x_\alpha\Big(a - b + \frac{1}{2}e\Big) + (3x_\alpha - x^3_\alpha)\Big(\frac{1}{4}b - \frac{1}{2}a\Big)\Big) + o(n^{-1}). \tag{54}$$

Since the density (49) is, for $\delta = 0$ and $a_1 = a_3 = 0$, symmetric about the origin, equations (53) and (54) are equally applicable for two-sided tests of size 2α.

For either (53) or (54), the $o(n^{-1})$ symbol is *not* uniform in α, so that for small α the approximations cannot be expected to be very reliable. In particular, the density (49) need not even be non-negative definite. If in (34) one puts $\boldsymbol{A}_n = n/(n-1)\boldsymbol{M}_{\boldsymbol{Z}}$, x will for $\delta = 0$ have the Student's t distribution with $n-1$ degrees of freedom. Since $\mathrm{Tr}(\boldsymbol{A}_n) = n$, we get $e = 0$. Also, $a = 0$, and $b = n/(n-1) = 1 + O(n^{-1})$. Thus the approximation (54) becomes

$$x^*_\alpha = x_\alpha + (4n)^{-1}(x_\alpha + x^3_\alpha),$$

which is, to order n^{-1}, the Cornish-Fisher approximation for critical values of the Student's t distribution. To provide a standard of comparison, see Table 1, which compares true values of x^*_α for $\alpha = .025$ and $\alpha = .005$ with the Cornish-Fisher approximation and with the solution to equation (53).

Nevertheless, (53) and (54) do demonstrate why a t statistic based on HCR_0 can often be expected to behave better under the null hypothesis than one based on HC_0. For the simple case in which there are no regressors $\boldsymbol{X}$ in the null hypothesis, so that $\boldsymbol{M}_{\boldsymbol{X}} = \mathbf{I}$, we find from (48) that, for HCR_0, $a = b = n^{-1}\mathrm{Tr}(\boldsymbol{\Omega}_0\boldsymbol{D})^2$, and $e = 0$ (with $\delta = 0$). Since $a_1 = a_2 = a_3 = 0$ in this case, the only discrepancy from the $\mathrm{N}(0,1)$ distribution comes therefore from a_4, which is equal to $-a/4$. To this order, then, HCR_0 has mean zero, variance one, no skewness, and a touch of platykurtosis (i.e., thinner tails than the standard normal). The distribution of HC_0 is not so simple, and it can thus be expected to differ more from its asymptotic distribution.

Unfortunately, since in practice $\boldsymbol{\Omega}_0$ is not known, the quantities a, b, and e appearing in (53) and (54) must be estimated. We now discuss some convenient and consistent estimators. Let

$$\hat\sigma^2_t = (\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{y})^2_t \quad \text{and} \quad \tilde\sigma^2_t = (\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y})^2_t.$$

In order to achieve the normalization (45), we now set

$$\hat\omega_t = \hat\sigma^2_t \Big/ \Big(\frac{1}{n}\sum_{t=1}^n \hat\sigma^2_t(\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z})^2_t\Big), \text{ and}$$

$$\tilde\omega_t = \tilde\sigma^2_t \Big/ \Big(\frac{1}{n}\sum_{t=1}^n \tilde\sigma^2_t(\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z})^2_t\Big).$$

Then $\ddot{\boldsymbol{\Omega}}$ can be either $\mathrm{diag}(\hat{\omega}_t)$ or $\mathrm{diag}(\tilde{\omega}_t)$. For an estimate of e, we use

$$\hat{e} = \mathrm{Tr}\big(\ddot{\boldsymbol{\Omega}}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\big) - n \quad \text{for HC}_0, \quad \text{or}$$

$$\tilde{e} = \mathrm{Tr}\big(\ddot{\boldsymbol{\Omega}}\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{D}\boldsymbol{M}_{\boldsymbol{X}}\big) - n \quad \text{for HCR}_0.$$

Since $\boldsymbol{\Omega}_0$ appears quadratically in the expressions for a and b, we need slightly more complicated estimators. One can see from (48) that $b = n^{-1}\mathrm{Tr}(\boldsymbol{\Omega}_0^2\boldsymbol{D}^2) + O(n^{-1})$, and so a consistent estimator is

$$\hat{b} = (3n)^{-1}\mathrm{Tr}(\ddot{\boldsymbol{\Omega}}^2\boldsymbol{D}^2),$$

since $\mathrm{E}(\hat{\sigma}_t^4) = 3\sigma_t^4 + o(1)$.

In the expression for a, because neither $\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{D}\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}$ nor $\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{D}\boldsymbol{M}_{\boldsymbol{X}}$ is diagonal, cross terms with products $\omega_t\omega_{t'}$, for $t \neq t'$, appear as well as terms with ω_t^2. Thus we need the estimator

$$\hat{a} = \frac{1}{n}\boldsymbol{z}^\top\boldsymbol{M}_{\boldsymbol{X}}\ddot{\boldsymbol{\Omega}}\boldsymbol{M}\boldsymbol{D}\boldsymbol{M}\ddot{\boldsymbol{\Omega}}\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z} - \frac{2}{3}\frac{1}{n}\boldsymbol{z}^\top\boldsymbol{M}_{\boldsymbol{X}}\ddot{\boldsymbol{\Omega}}\,\mathrm{diag}(\boldsymbol{M}\boldsymbol{D}\boldsymbol{M})\ddot{\boldsymbol{\Omega}}\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z},$$

where $\boldsymbol{M} = \boldsymbol{M}_{\boldsymbol{X}}$ or $\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}$, as required, and $\mathrm{diag}(\boldsymbol{M}\boldsymbol{D}\boldsymbol{M})$ denotes a diagonal matrix with the same diagonal elements as $\boldsymbol{M}\boldsymbol{D}\boldsymbol{M}$.

We shall conclude this section with a discussion of the power of t statistics based on HC_0 and HCR_0 in the event that the sequence of local DGPs satisfies the alternative hypothesis implied by the use of the vector $\boldsymbol{z}$. Thus the DGP is assumed to be

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}_0 + n^{-1/2}\boldsymbol{z}\gamma_0 + \boldsymbol{\Omega}_0^{1/2}\boldsymbol{\varepsilon}. \tag{55}$$

This means that, for (43) to be satisfied, we replace the vector $\boldsymbol{w}$ of (40) by $\zeta^{-1/2}\boldsymbol{z}$, where ζ is defined as $n^{-1}\boldsymbol{z}^\top\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{\Omega}_0^{-1}\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z}$, and we replace the δ of (40) by $\zeta^{-1/2}\gamma_0$.

From the definition of $\cos\theta$, we find that $\cos\theta = \zeta^{-1/2}n^{-1}\boldsymbol{z}^\top\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z}$, and so the noncentrality parameter of the asymptotic distribution of the test statistics, which from (37) is $\delta\cos\theta$, becomes $\gamma_0 n^{-1}\boldsymbol{z}^\top\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z}$. In fact, it is quite compatible with all earlier normalizations to impose in addition that $\boldsymbol{z}^\top\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{z} = n$, and this fixes also the normalization of γ_0, so that the noncentrality parameter is just γ_0. For a test in chi-squared form, the noncentrality parameter (19) would then be γ_0^2.

Now we define the power function of a one-sided test based on a statistic x as $P(\alpha, \gamma_0) = \Pr(x > x_\alpha^*)$, where the probability is calculated under the DGP (55), and x_α^* is the level α critical value. From (52), it follows that

$$P(\alpha, \gamma_0) = \Phi\big(-(x_\alpha^* - \gamma_0)\big) - \frac{1}{n}\phi(x_\alpha^* - \gamma_0)\sum_{r=1}^{4} a_r(\gamma_0)H_{r-1}(x_\alpha^* - \gamma_0) + o(n^{-1}),$$

where the coefficients a_r have been written as explicit functions of γ_0. But a Taylor expansion about x_α, the $\mathrm{N}(0,1)$ critical value, gives with the aid of (54) that

$$\Phi\big(-(x_\alpha^* - \gamma_0)\big) = \Phi\big(-(x_\alpha - \gamma_0)\big) + \frac{1}{n}\phi(x_\alpha - \gamma_0)\sum_{r=1}^{4} a_r(\gamma_0)H_{r-1}(x_\alpha) + o(n^{-1}),$$

so that

$$P(\alpha,\gamma_0) = \Phi\bigl(-(x_\alpha-\gamma_0)\bigr) - \frac{1}{n}\phi(x_\alpha-\gamma_0)\sum_{r=1}^{4} a_r(\gamma_0)H_{r-1}(x_\alpha-\gamma_0) \\ + \frac{1}{n}\phi(x_\alpha-\gamma_0)\sum_{r=1}^{4} a_r(0)H_{r-1}(x_\alpha) + o(n^{-1}). \tag{56}$$

Notice that $\Phi\bigl(-(x_\alpha-\gamma_0)\bigr)$ is the asymptotic power function, that is, the power function of a statistic that is $\mathrm{N}(0,1)$ under the null and $\mathrm{N}(\gamma_0,1)$ under the alternative (55). The summations in (56) can be evaluated by use of (38) and (51), and in the notation of (38) the result is

$$\frac{1}{2}\delta^2 x_\alpha(d-2\cos\theta+c\cos^2\theta) + \delta x_\alpha^2\bigl(c-(a-b/4)\cos\theta\bigr).$$

From (48), we find for the present circumstances in which the vector $\boldsymbol{w}$ of (48) is replaced by $\zeta^{-1/2}\boldsymbol{z}$, δ is replaced by $\zeta^{1/2}\gamma_0$, and $\cos\theta$ is replaced by $\zeta^{-1/2}$, that

$$\frac{1}{2}\delta^2 x_\alpha(d-2c\cos\theta+a\cos^2\theta) = \frac{1}{2n}\gamma_0^2 x_\alpha \boldsymbol{z}^\top(\mathbf{I}-\boldsymbol{\Omega}_0\boldsymbol{M_X})^\top\boldsymbol{MDM}(\mathbf{I}-\boldsymbol{\Omega}_0\boldsymbol{M_X})\boldsymbol{z}, \tag{57}$$

and

$$\delta x_\alpha^2\bigl(c-(a-b/4)\cos\theta\bigr)) \\ = \gamma_0 x_\alpha^2\Bigl(\frac{1}{n}\boldsymbol{z}^\top(\mathbf{I}-\boldsymbol{M_X\Omega}_0)\boldsymbol{MDM\Omega}_0\boldsymbol{M_X z} + \frac{1}{4n}\mathrm{Tr}(\boldsymbol{\Omega}^2\boldsymbol{D}^2)\Bigr), \tag{58}$$

where, as before, $\boldsymbol{M}$ denotes either $\boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}$, for HC_0, or $\boldsymbol{M_X}$, for HCR_0. A little algebra shows that (57) yields the same expression for both possibilities for $\boldsymbol{M}$, which is most easily written as

$$\frac{1}{2n}\gamma_0^2 x_\alpha \boldsymbol{z}^\top\boldsymbol{M_X}(\mathbf{I}-\boldsymbol{\Omega}_0\boldsymbol{M_X})\boldsymbol{D}(\mathbf{I}-\boldsymbol{M_X\Omega}_0)\boldsymbol{M_X z}.$$

On the other hand, (58) yields for HC_0

$$\gamma_0 x_\alpha^2\frac{1}{n}\Bigl(-\boldsymbol{z}^\top\boldsymbol{M_X}(\mathbf{I}-\boldsymbol{\Omega}_0\boldsymbol{M_X})\boldsymbol{D}(\mathbf{I}-\boldsymbol{M_X\Omega}_0)\boldsymbol{M_X z} + \frac{1}{4}\mathrm{Tr}(\boldsymbol{\Omega}^2\boldsymbol{D}^2)\Bigr)$$

but for HCR_0

$$\gamma_0 x_\alpha^2\frac{1}{n}\Bigl(\boldsymbol{z}^\top\boldsymbol{M_X}(\mathbf{I}-\boldsymbol{\Omega}_0\boldsymbol{M_X})\boldsymbol{DM_X\Omega}_0\boldsymbol{M_X z} + \frac{1}{4}\mathrm{Tr}(\boldsymbol{\Omega}^2\boldsymbol{D}^2)\Bigr).$$

The difference in the power functions is thus, to order n^{-1}, simply

$$n^{-2}\phi(x_\alpha-\gamma_0)\,\boldsymbol{z}^\top\boldsymbol{M_X}(\mathbf{I}-\boldsymbol{\Omega}_0\boldsymbol{M_X})\boldsymbol{DM_X z}. \tag{59}$$

This quantity cannot be signed, so that there is no unambiguous ranking of the two t statistics, based on HC_0 and HCR_0, with regard to power. In the absence of heteroskedasticity, with $\boldsymbol{\Omega}_0 = \mathbf{I}$ (since our normalizations require this rather than that

$\boldsymbol{\Omega}_0 = \sigma_0^2\mathbf{I}$), expression (59) vanishes, and for both tests the power function is to order n^{-1} given by

$$P(\alpha, \gamma_0) = \Phi\big(-(x_\alpha - \gamma_0)\big) - \frac{1}{4n^2}\,\phi(x_\alpha - \gamma_0)\gamma_0 x_\alpha^2 \operatorname{Tr}(\boldsymbol{D}^2).$$

This can be compared with the power function of the standard t statistic, which is given by

$$\frac{n^{-1/2}\boldsymbol{z}^\top \boldsymbol{M_X y}}{\big((n-k-1)^{-1}\boldsymbol{y}^\top \boldsymbol{M}_{[\boldsymbol{X}\ \boldsymbol{Z}]}\boldsymbol{y}\big)^{1/2}}, \tag{60}$$

where k is the rank of $\boldsymbol{X}$; recall that $\boldsymbol{zM_X z} = n$. Calculations based on expression (60) lead to the following power function to order n^{-1}:

$$P(\alpha, \gamma_0) = \Phi\big(-(x_\alpha - \gamma_0)\big) - \frac{1}{4n}\phi(x_\alpha - \gamma_0)\gamma_0 x_\alpha^2.$$

The loss of power resulting from the use of HC_0 and HCR_0 rather than the usual OLS standard error is thus measured by $n^{-1}\operatorname{Tr}(\boldsymbol{D}^2) - 1$. this quantity is, not surprisingly, always non-negative, since

$$\operatorname{Tr}(\boldsymbol{D}^2) = \sum_{t=1}^{n}(\boldsymbol{M_X z})_t^4 \geq \tfrac{1}{n}\Big(\sum_{t=1}^{n}(\boldsymbol{M_X z})_t^2\Big)^2 = n, \tag{61}$$

by the Cauchy-Schwartz inequality and the normalization. Equality in (61) occurs only if all the $(\boldsymbol{M_X z})_t$ are equal.

6. Monte Carlo Experiments

In this section, we present the results of several Monte Carlo experiments designed to study the performance of the test statistics HC_0 through HC_3 and HCR_0 through HCR_3 under the null hypothesis. We also investigate the usefulness of the Edgeworth expansions derived in the previous section. Because those expansions are for the case of testing for a single omitted variable, we only report results for that case, and we express all the test statistics as quasi t statistics. We also performed several unreported experiments in which we looked at tests for two or more omitted variables, and we obtained results qualitatively similar to those reported below.

In all of our experiments, we utilized the following model:

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \gamma_1 Z_{1t} + u_t, \tag{62}$$

where n = 50, 100, 200, or 400. In all but one case, the regressors X_1 and X_2 were the ninety day Treasury Bill rate for Canada and the change in the log of real GNP for Canada, seasonally adjusted at annual rates, while Z_1 was the unemployment rate for Canada, seasonally adjusted. These series were for the period 1970:3 to 1984:4, a total of fifty observations. When more than fifty observations were used, the original fifty were replicated the required number of times. We chose the regressors in this way because we wanted them to be representative of real data, and so that the matrix $\boldsymbol{X}^\top\boldsymbol{X}/n$ would not change as the sample size n was changed.

We report the results of four sets of experiments. In the first set, Case 1, the errors were normal, and there was in fact no heteroskedasticity. In the second set, Case 2, the variance of u_t was proportional to Z_{1t}^2. In the third set, Case 3, the variance of u_t changed abruptly, as if due to some sort of structural change. The errors u_t were specified to be $\mathrm{N}(0, \sigma^2)$ for $t = l, \ldots, 25$, $t = 51, \ldots, 75$, and so on, and $\mathrm{N}(0, 16\sigma^2)$ for $t = 26, \ldots, 50$, $t = 76, \ldots, 100$, and so on. This pattern was chosen so that the relationship between the regressors and the variance of the error term would not change with the sample size. The final set of experiments, Case 4, was similar to Case 3 except that the regressors were different. In this case, X_1 was the unemployment rate, X_2 was the price of the U.S. dollar in terms of Canadian dollars, and Z_t was the change in the log of the GNE price deflator for Canada, seasonally adjusted at annual rates.

In Tables 2 through 5, we report the results of sixteen experiments (four cases for each of $n = 50$, $n = 100$, $n = 200$, and $n = 400$). Each experiment involved 2,000 replications. In addition to the quasi t statistics based on HC_0 through HC_3 and HCR_0 through HCR_3, we calculated the ordinary t statistic for the (true) hypothesis that $\gamma_1 = 0$, and also a control variate which utilizes the true covariance matrix and is thus exactly $\mathrm{N}(0, 1)$. To save space, not all the results are reported for n greater than 50. We omit results for HC_1 and HC_2 because they always fall between those for HC_0 and HC_3, and we omit results for HCR_2 and HCR_3 because those statistics always perform very similarly to HCR_1, and they do not seem to be worth the additional complexity of calculation. Differences between HCR_1, HCR_2, and HCR_3 are of course most marked for $n = 50$, so any differences worth noticing do show up in the tables.

We report the standard deviation and the kurtosis of both the control variate and the actual statistics. These should be roughly one and three respectively, and when they differ significantly at the one per cent level from those values, we note the fact with an asterisk. The information about the control variate can be useful in interpreting the results. For example, in Case 3 some of the results for $n = 400$ actually seem to be worse than those for $n = 200$. But this is clearly explained by the fact that the control variate has a standard deviation of 1.031 in the former case and only 0.992 in the latter.

Our estimates of rejection frequencies at nominal significance levels explicitly incorporate the information in the control variate, using a technique proposed by Davidson and MacKinnon (1981a). If the control variate has exceeded its critical value by 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 standard error of the resulting estimate depends on the amount of correlation between the control variate and the statistic in question, and thus tends to fall as n increases. For example, in

Table 2 the standard errors of the rejection frequencies at the 5% nominal level for HC_0 are 0.38 for $n = 50$ and 0.27 for $n = 400$. If the usual estimator of these frequencies had been used (i.e. one that simply divides the observed number of rejections by the number of replications), and if the estimates had been the same, these standard errors would have been 0.68 and 0.53, respectively. There are thus substantial gains to be had by using the information in the control variate in this way.

Tables 2 through 5 are reasonably self-explanatory. As in MacKinnon and White (1985), HC_3 always outperforms HC_2, which always outperforms HC_1, which always outperforms HC_0. The latter is often very unreliable, typically performing worse than the ordinary t statistic (OLS) for $n = 50$, except in the extreme case of Table 5. On the other hand, the tests based on restricted residuals perform strikingly well. Even HCR_0 usually yields more reliable inferences than HC_3, especially when the sample size is small and heteroskedasticity is severe; see in particular Table 5. Differences among the HCR statistics are relatively slight, with HCR_1 usually doing just about as well as any of them.

One major difference between the tests based on restricted and unrestricted residuals is that the former invariably display platykurtosis, while the latter invariably display leptokurtosis, at least for the smaller sample sizes. This is what the Edgeworth expansions predict. As a result, all the tests are more reliable at the 10% level than at the 5% or 1% levels. Tests based on restricted residuals tend to reject too infrequently at low significance levels, while tests based on unrestricted residuals tend to reject too frequently at those levels.

In Tables 6 through 9, we evaluate the performance of the Edgeworth expansions derived in Section 5. We see how well the standard deviation predicted by expression (39) predicts the standard deviation actually observed, and how well the 10%, 5% and 1% critical values obtained by solving equation (53) compare with the estimated 95% confidence intervals for the corresponding estimated critical values. These confidence intervals are based on non-parametric inference, and they are therefore not symmetric around the estimate; for details on their calculation, see Mood and Graybill (1963, pp. 406–409).

We also estimate rejection frequencies using critical values based on the Edgeworth approximations, using the information in the control variate as before. This is done twice, once for the critical values calculated knowing $\boldsymbol{\Omega}$ (referred to as "True Coefficients" in the table), and once for critical values calculated at each replication using $\tilde{\boldsymbol{\Omega}}$ or $\hat{\boldsymbol{\Omega}}$ (referred to as "Estimated Coefficients"). The critical values based on true coefficients perform very well indeed for HCR_0, and reasonably well for HC_0, when n is 100 or more. The critical values based on estimated coefficients perform much less well. One would generally obtain more reliable inferences simply by using HCR_1 than by using HCR_0 with corrected critical values, and one would almost always obtain more reliable inferences by using HC_3 than by using HC_0 with corrected critical values, especially when the sample size is small. Thus it would appear that the Edgeworth expansion approach may be more useful in improving our theoretical understanding of the properties of these tests (see Section 5) than in actually helping us to make accurate inferences in practice.

7. Conclusion

In this paper, we have proposed computationally simple procedures for performing heteroskedasticity-robust tests in regression directions. Although tests based on unrestricted residuals are often quite unreliable in small samples, and should therefore be used with caution, tests based on restricted residuals seem to perform remarkably well. If they have a fault, it is a tendency not to reject frequently enough when the nominal size of the test is small. We would argue that such tests should be used routinely when the sample size is not too small and there is any doubt about the validity of the usual homoskedasticity assumption.

References

Chow, G. C. (1960). "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28, 591–605.

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

Davidson, R. and MacKinnon, J. G. (1981b). "Several tests for model specification in the presence of alternative hypotheses," *Econometrica*, 49, 781–793.

Davidson, R. and MacKinnon, J. G. (1982). "Some non-nested hypothesis tests and the relations among them," *Review of Economic Studies*, 49, 551-565.

Davidson, R. and MacKinnon, J. G. (1984). "Model specification tests based on artificial linear regressions," *International Economic Review*, 25, 485-502.

Davidson, R. and MacKinnon, J. G. (1985). "The interpretation of test statistics," *Canadian Journal of Economics*, 18, 38–57.

Davidson, R., Godfrey, L. G., and MacKinnon, J. G. (1985). "A simplified version of the differencing test," *International Economic Review*, 26, 639–647.

Durbin, J. (1954). "Errors in variables," *Review of the International Statistical Institute*, 22, 23–32.

Durbin, J. (1970). "Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables," *Econometrica*, 38, 410–421.

Engle, R. F. (1982). "A general approach to Lagrange multiplier model diagnostics," *Journal of Econometrics*, 20, 83–104.

Godfrey, L. G. (1978). "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables," *Econometrica*, 46, 1293–1301.

Hausman, J. A. (1978). "Specification tests in econometrics," *Econometrica*, 46, 1251–1272.

Holly, A. (1982). "A remark on Hausman's specification test," *Econometrica*, 50, 749–759.

Lovell, M. C. (1963). "Seasonal adjustment of economic time series," *Journal of the American Statistical Association*, 58, 993–1010.

MacKinnon, J. G. and White, H. (1985). "Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties," *Journal of Econometrics*, 29, 305–325.

Messer, K. and White, H. (1984). "A note on computing the heteroskedasticity consistent covariance matrix using instrumental variables techniques," *Bulletin of the Oxford Institute of Statistics*, 46, 181–184.

Mood, A. M. and Graybill, F. A. (1963). *Introduction to the Theory of Statistics*, Second Edition, New York, McGraw-Hill.

Plosser, C. E., Schwert, G. W., and White, H. (1982). "Differencing as a test of specification," *International Economic Review*, 23, 535–552.

Rothenberg, T. J. (1984). "Approximate power functions for some robust tests of regression coefficients," U. C. Berkeley Discussion Paper 84-1.

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

White, H. (1984). *Asymptotic Theory for Econometricians*, New York, Academic Press.

Appendix

Subscripts n will be dropped throughout. Let the random variable x_0 be defined as $n^{-1/2}\boldsymbol{z}^\top\boldsymbol{u}$. Then $x_0 \sim \mathrm{N}(0,1)$. From (35), we obtain:

$$n^{-1/2}\boldsymbol{z}^\top\boldsymbol{y} = n^{-1}\boldsymbol{z}^\top\boldsymbol{w}\delta + x_0 = \delta\cos\theta + x_0. \tag{63}$$

Using the definitions in (36),

$$\begin{aligned} n^{-1}\boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{y} &= n^{-1}\boldsymbol{u}^\top\boldsymbol{A}\boldsymbol{u} + 2n^{-3/2}\delta\boldsymbol{w}^\top\boldsymbol{A}\boldsymbol{u} + n^{-2}\delta^2\boldsymbol{w}^\top\boldsymbol{A}\boldsymbol{w} \\ &= n^{-1}(\boldsymbol{u}^\top\boldsymbol{M_z}\boldsymbol{A}\boldsymbol{M_z}\boldsymbol{u} + 2\boldsymbol{u}^\top\boldsymbol{M_z}\boldsymbol{A}\boldsymbol{P_z}\boldsymbol{u} + \boldsymbol{u}^\top\boldsymbol{P_z}\boldsymbol{A}\boldsymbol{P_z}\boldsymbol{u}) \\ &\quad + 2n^{-3/2}\delta\boldsymbol{w}^\top\boldsymbol{A}\boldsymbol{u} + n^{-1}\delta^2 d, \end{aligned} \tag{64}$$

where $\boldsymbol{P_z} = \mathbf{I} - \boldsymbol{M_z} = n^{-1}\boldsymbol{z}^\top\boldsymbol{z}$.

Now $\boldsymbol{P_u} = n^{-1/2}x_0\boldsymbol{z}$ and $\boldsymbol{M_z}\boldsymbol{u}$ is a vector of jointly normal random variables independent of x_0. Further

$$n^{-1}\mathrm{E}(\boldsymbol{u}^\top\boldsymbol{M_z}\boldsymbol{A}\boldsymbol{M_z}\boldsymbol{u}) = n^{-1}\mathrm{Tr}(\boldsymbol{M_z}\boldsymbol{A}) = 1 + n^{-1}(e-a).$$

Define a random variable z, independent of x_0, by

$$n^{-1}\boldsymbol{u}^\top\boldsymbol{M_z}\boldsymbol{A}\boldsymbol{M_z}\boldsymbol{u} = 1 + n^{-1/2}z + n^{-1}(e-a).$$

It follows that $\mathrm{E}(z) = 0$, and one calculates further that

$$\mathrm{Var}(z) = \frac{1}{n}\,\mathrm{Var}(\boldsymbol{u}^\top\boldsymbol{M_z}\boldsymbol{A}\boldsymbol{M_z}\boldsymbol{u}) = \frac{2}{n}\mathrm{Tr}(\boldsymbol{M_z}\boldsymbol{A}\boldsymbol{M_z}\boldsymbol{A}) = 2b + o(1).$$

Thus $z = O_2(1)$, where the order symbol O_2 refers to the mean-square norm $\|\cdot\|_2$. For a random variable v, $\|v\|_2 = \mathrm{E}(v^2)$.

Next, we define two more normal, zero-mean random variables, both of order unity in mean square and independent of x_0, by the equations:

$$\begin{aligned} w_1 &= n^{-1/2}\boldsymbol{w}^\top\boldsymbol{A}\boldsymbol{M_z}\boldsymbol{u}, \text{ and} \\ w_2 &= n^{-1/2}\boldsymbol{z}^\top\boldsymbol{A}\boldsymbol{M_z}\boldsymbol{u}. \end{aligned}$$

Equation (64) can now be rewritten as

$$n^{-1}\boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{y} = 1 + n^{-1/2}z + n^{-1}(e - a + 2x_0w_2 + ax_0^2 + 2\delta w_1 + 2\delta dx_0 + \delta^2 d).$$

The random variable in parentheses is $O_2(1)$, and thus it is bounded in probability. Consequently, $(n^{-1}\boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{y})^{-1/2}$ can be expanded by the binomial theorem:

$$\begin{aligned} (n^{-1}\boldsymbol{y}^\top\boldsymbol{A}\boldsymbol{y})^{-1/2} &= 1 - \frac{1}{2}n^{-1/2}z \\ &\quad - \frac{1}{2n}\Big(e - a + \delta^2 d + 2\delta w_1 + 2x_0(\delta d + w_2) + ax_0^2 - \frac{3}{4}z^2\Big) + o_p(1). \end{aligned}$$

The calculation of the characteristic function $\psi(t)$ of the theorem can best be undertaken in two stages, the first involving the computation of the conditional expectation $\mathrm{E}(e^{itx} \mid x_0)$. By a Taylor expansion of the exponential function, we get

$$\mathrm{E}(e^{itx} \mid x_0) = e^{it(x_0+\delta\cos\theta)}\Big(1 - \frac{1}{4n}bt^2(x_0+\delta\cos\theta)^2 \\ - \frac{1}{2n}it(x_0+\delta\cos\theta)\big(e - a + \delta^2 d - \frac{3}{2}b + 2\delta d x_0 + a x_0^2\big)\Big) + o_p(1).$$

However,

$$\mathrm{E}(e^{itx_0}) = e^{-t^2/2},$$

since $x_0 \sim \mathrm{N}(0,1)$, and

$$\mathrm{E}(x^r e^{itx_0}) = (-1)^r H_r(t) e^{-t^2/2} \text{ for } r = 1, 2, \ldots,$$

as can readily be seen from the definition (50) of the Hermite polynomials. Thus it is a matter of simple and tedious algebra to verify the conclusion of the theorem, with the proviso that the random variable denoted $o_p(n^{-1})$ possesses an expectation. But this follows from standard results on quadratic forms of normal variables.

Table 1. Performance of Two Approximations

	D.F.	Cornish-Fisher Values	Solution of (53)	True Values
$\alpha = .025$	40	2.019	2.020	2.021
	30	2.039	2.041	2.042
	10	2.198	2.206	2.228
$\alpha = .005$	40	2.699	2.694	2.704
	30	2.740	2.731	2.750
	10	3.068	2.972	3.169

Table 2. Case 1: No Heteroskedasticity

N. Obs.	C.V.	Stat.	S.D.	Kurt.	10%		5%		1%	
50	1.004	OLS	1.022	3.16	10.91	(0.38)	5.36	(0.32)	1.01	(0.14)
	3.01	HC_0	1.214*	3.64*	16.71*	(0.55)	10.31*	(0.38)	3.67*	(0.35)
		HC_1	1.165*	3.64*	15.20*	(0.52)	9.22*	(0.47)	3.17*	(0.33)
		HC_2	1.113*	3.70*	13.43*	(0.53)	7.60*	(0.45)	2.64*	(0.29)
		HC_3	1.024	3.79*	10.45	(0.52)	5.71	(0.41)	1.43	(0.23)
		HCR_0	1.023	2.45*	9.96	(0.52)	4.65	(0.41)	0.38*	(0.13)
		HCR_1	0.992	2.45*	9.21	(0.51)	4.01	(0.40)	0.29*	(0.12)
		HCR_2	0.988	2.45*	9.11	(0.51)	3.91*	(0.39)	0.24*	(0.11)
		HCR_3	0.989	2.58*	9.29	(0.51)	4.01	(0.40)	0.53*	(0.15)
100	0.993	OLS	1.004	3.10	10.51	(0.31)	5.72*	(0.25)	0.95	(0.16)
	3.04	HC_0	1.101*	3.34*	13.87*	(0.46)	7.48*	(0.39)	2.21*	(0.26)
		HC_3	1.011	3.39*	10.80	(0.46)	5.68	(0.36)	1.43	(0.22)
		HCR_0	1.007	2.76	10.55	(0.48)	4.98	(0.37)	0.73	(0.17)
		HCR_1	0.992	2.76	10.04	(0.46)	4.57	(0.35)	0.68	(0.16)
200	1.011	OLS	1.014	2.88	10.11	(0.24)	4.99	(0.22)	0.89	(0.12)
	2.89	HC_0	1.067*	3.05	11.73*	(0.41)	5.93*	(0.33)	1.66*	(0.23)
		HC_3	1.021	3.07	9.79	(0.39)	4.97	(0.31)	1.36	(0.19)
		HCR_0	1.024	2.79	10.48	(0.40)	4.72	(0.31)	0.93	(0.17)
		HCR_1	1.016	2.79	10.19	(0.40)	4.48	(0.31)	0.88	(0.16)
400	1.018	OLS	1.021	3.13	10.11	(0.19)	4.85	(0.21)	1.30	(0.12)
	3.11	HC_0	1.049*	3.28	10.97*	(0.31)	6.04*	(0.27)	1.47*	(0.17)
		HC_3	1.027	3.29*	9.97	(0.31)	5.03	(0.27)	1.27	(0.16)
		HCR_0	1.025	3.08	10.22	(0.31)	4.88	(0.25)	0.97	(0.16)
		HCR_1	1.021	3.08	10.12	(0.31)	4.91	(0.26)	0.83	(0.14)

Numbers under "C.V." are the standard deviation and kurtosis of the control variate.

Numbers under "S.D." and "Kurt." are the standard deviation and kurtosis of the test statistics.

Numbers under "10%", "5%", and "1%" are the estimated rejection percentages at those nominal levels. The standard errors of these estimates, which incorporate the information in the control variate, are in parentheses.

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

Number of replications = 2,000.

Table 3. Case 2: Variance Proportional to Square of Omitted Variable

N. Obs.	C.V.	Stat.	S.D.	Kurt.	10%	5%	1%
50	0.985	OLS	1.326*	3.09	22.19* (0.46)	14.06* (0.38)	5.44* (0.34)
	2.99	HC_0	1.365*	3.87*	21.10* (0.65)	14.09* (0.60)	6.25* (0.46)
		HC_1	1.309*	3.87*	19.07* (0.65)	13.18* (0.58)	5.53* (0.44)
		HC_2	1.233*	4.02*	16.60* (0.64)	10.77* (0.57)	4.46* (0.42)
		HC_3	1.119*	4.19*	13.17* (0.62)	8.30* (0.54)	3.34* (0.37)
		HCR_0	1.039	2.42*	11.26 (0.61)	4.36 (0.43)	0.72 (0.19)
		HCR_1	1.008	2.42*	9.55 (0.59)	3.54* (0.40)	0.46* (0.15)
		HCR_2	1.002	2.43*	9.02 (0.58)	3.54* (0.40)	0.46* (0.15)
		HCR_3	1.003	2.58*	9.23 (0.59)	3.78* (0.41)	0.77 (0.19)
100	1.030	OLS	1.377*	3.15	22.03* (0.36)	14.65* (0.35)	5.40* (0.26)
	3.18	HC_0	1.228*	3.65*	15.53* (0.54)	9.77* (0.49)	4.27* (0.37)
		HC_3	1.104*	3.75*	11.77* (0.55)	6.99* (0.44)	2.86* (0.34)
		HCR_0	1.058*	2.65*	11.38 (0.55)	5.62 (0.43)	1.02 (0.22)
		HCR_1	1.042*	2.65*	10.91 (0.54)	5.31 (0.42)	0.70 (0.18)
200	0.996	OLS	1.324*	3.05	21.42* (0.31)	13.99* (0.29)	5.42* (0.23)
	3.03	HC_0	1.108*	3.29*	13.71* (0.51)	8.73* (0.45)	2.32* (0.27)
		HC_3	1.049*	3.31*	11.97* (0.49)	7.24* (0.42)	1.77* (0.23)
		HCR_0	1.031	2.77	11.69* (0.49)	6.00 (0.43)	0.76 (0.14)
		HCR_1	1.023	2.77	11.18 (0.49)	5.72 (0.42)	0.66* (0.13)
400	1.023	OLS	1.349*	3.03	21.83* (0.28)	13.74* (0.24)	5.40* (0.20)
	3.03	HC_0	1.073*	3.25	10.95 (0.37)	6.83* (0.34)	1.61* (0.20)
		HC_3	1.044*	3.25	10.23 (0.36)	5.84* (0.31)	1.49* (0.17)
		HCR_0	1.032	2.89	10.25 (0.38)	4.96 (0.30)	0.83 (0.14)
		HCR_1	1.028	2.89	10.07 (0.38)	4.99 (0.31)	0.76 (0.14)

Numbers under "C.V." are the standard deviation and kurtosis of the control variate.

Numbers under "S.D." and "Kurt." are the standard deviation and kurtosis of the test statistics.

Numbers under "10%", "5%", and "1%" are the estimated rejection percentages at those nominal levels. The standard errors of these estimates, which incorporate the information in the control variate, are in parentheses.

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

Number of replications = 2,000.

Table 4. Case 3: Structural Change in Variance

N. Obs.	C.V.	Stat.	S.D.	Kurt.	10%	5%	1%
50	0.996	OLS	1.220*	3.20	17.16* (0.47)	10.83* (0.44)	3.46* (0.24)
	3.09	HC_0	1.285*	3.63*	19.22* (0.63)	13.34* (0.57)	4.70* (0.37)
		HC_1	1.232*	3.63*	17.53* (0.62)	12.30* (0.55)	3.89* (0.32)
		HC_2	1.171*	3.77*	15.57* (0.60)	10.45* (0.54)	2.93* (0.31)
		HC_3	1.073*	3.96*	13.09* (0.60)	7.54* (0.51)	1.76* (0.26)
		HCR_0	1.052*	2.49*	12.77* (0.62)	4.94 (0.44)	0.68 (0.18)
		HCR_1	1.020	2.49*	10.90 (0.60)	4.23 (0.40)	0.43* (0.15)
		HCR_2	1.014	2.49*	10.32 (0.58)	3.95* (0.39)	0.39* (0.14)
		HCR_3	1.017	2.65*	10.36 (0.59)	4.60 (0.42)	0.68 (0.18)
100	1.030	OLS	1.232*	3.17	17.58* (0.37)	10.14* (0.35)	3.24* (0.22)
	3.18	HC_0	1.156*	3.36*	13.66* (0.48)	8.09* (0.44)	2.76* (0.30)
		HC_3	1.050*	3.43*	10.43 (0.47)	5.64 (0.41)	1.86* (0.24)
		HCR_0	1.040	2.72*	10.34 (0.51)	4.98 (0.40)	0.81 (0.19)
		HCR_1	1.025	2.72*	9.81 (0.51)	4.74 (0.38)	0.77 (0.19)
200	0.996	OLS	1.187*	3.03	16.70* (0.30)	10.54* (0.26)	3.19* (0.21)
	3.03	HC_0	1.057*	3.24	11.96* (0.45)	6.31* (0.35)	1.57* (0.21)
		HC_3	1.006	3.26	10.20 (0.44)	5.44 (0.31)	1.31 (0.21)
		HCR_0	1.000	2.85	10.14 (0.43)	4.84 (0.31)	1.05 (0.18)
		HCR_1	0.992	2.85	9.65 (0.42)	4.78 (0.31)	1.00 (0.17)
400	1.023	OLS	1.227*	3.05	16.73* (0.32)	9.76* (0.29)	2.84* (0.18)
	3.03	HC_0	1.073*	3.22	11.40* (0.38)	6.99* (0.32)	1.51* (0.19)
		HC_3	1.047*	3.22	10.33 (0.39)	5.90* (0.31)	1.35 (0.18)
		HCR_0	1.041	2.89	10.28 (0.39)	5.80 (0.31)	0.97 (0.17)
		HCR_1	1.037	2.89	9.88 (0.39)	5.59 (0.32)	0.97 (0.17)

Numbers under “C.V.” are the standard deviation and kurtosis of the control variate.

Numbers under “S.D.” and “Kurt.” are the standard deviation and kurtosis of the test statistics.

Numbers under “10%”, “5%”, and “1%” are the estimated rejection percentages at those nominal levels. The standard errors of these estimates, which incorporate the information in the control variate, are in parentheses.

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

Number of replications = 2,000.

Table 5. Case 4: Structural Change in Variance; Different Regressors

N. Obs.	C.V.	Stat.	S.D.	Kurt.	10%	5%	1%
50	0.984	OLS	1.879*	2.52*	41.94* (0.52)	32.73* (0.56)	17.91* (0.54)
	2.80	HC_0	1.481*	3.43*	25.10* (0.82)	17.03* (0.73)	7.34* (0.56)
		HC_1	1.420*	3.43*	22.84* (0.80)	16.00* (0.72)	6.16* (0.52)
		HC_2	1.332*	3.63*	19.70* (0.78)	13.19* (0.69)	4.70* (0.47)
		HC_3	1.208*	3.86*	15.56* (0.73)	9.50* (0.63)	3.15* (0.39)
		HCR_0	1.074*	2.09*	9.82 (0.65)	3.79* (0.43)	0.60 (0.17
		HCR_1	1.042*	2.09*	8.42* (0.61)	3.08* (0.39)	0.35* (0.13)
		HCR_2	1.048*	2.08*	8.78 (0.62)	3.08* (0.39)	0.40* (0.14)
		HCR_3	1.059*	2.18*	9.22 (0.64)	3.85* (0.43)	0.55* (0.17)
100	0.968	OLS	1.834*	2.93	39.88* (0.50)	30.96* (0.52)	17.21* (0.41)
	3.09	HC_0	1.211 *	3.55*	17.02* (0.61)	10.81* (0.54)	3.73* (0.40)
		HC_3	1.087*	3.64*	12.91* (0.58)	7.55* (0.51)	2.18* (0.31)
		HCR_0	1.013	2.46*	10.72 (0.56)	4.16 (0.43)	0.40* (0.14)
		HCR_1	0.997	2.46*	10.05 (0.56)	3.50* (0.40)	0.35* (0.13)
200	0.990	OLS	1.857*	2.93	38.22* (0.41)	30.35* (0.43)	17.52* (0.40)
	3.16	HC_0	1.108*	3.00	14.19* (0.54)	8.59* (0.45)	2.48* (0.29)
		HC_3	1.049*	3.01	12.28* (0.52)	7.41 * (0.44)	1.71* (0.26)
		HCR_0	1.014	2.56*	10.75 (0.50)	5.70 (0.40)	0.24* (0.11)
		HCR_1	1.006	2.56*	10.49 (0.49)	5.02 (0.39)	0.24* (0.11)
400	1.002	OLS	1.886*	2.91	37.73* (0.30)	29.86* (0.32)	16.55* (0.30)
	2.90	HC_0	1.077*	3.23	11.84* (0.45)	6.95* (0.39)	1.79* (0.20)
		HC_3	1.047*	3.23	10.88 (0.43)	6.44* (0.40)	1.58* (0.20)
		HCR_0	1.025	2.85	10.65 (0.44)	6.04* (0.39)	0.95 (0.19)
		HCR_1	1.022	2.85	10.60 (0.44)	5.73 (0.38)	0.85 (0.17)

Numbers under "C.V." are the standard deviation and kurtosis of the control variate.

Numbers under "S.D." and "Kurt." are the standard deviation and kurtosis of the test statistics.

Numbers under "10%", "5%", and "1%" are the estimated rejection percentages at those nominal levels. The standard errors of these estimates, which incorporate the information in the control variate, are in parentheses.

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

Number of replications = 2,000.

Table 6. Performance of Edgeworth Approximations: Case 1

Sample Size:	50	100	200	400
Unrestricted residuals ($\mathbf{HC_0}$)				
Standard Deviation: Predicted	1.153	1.090	1.049	1.025
Estimated (S.E.)	1.214 (0.019)	1.101 (0.017)	1.067 (0.017)	1.049 (0.017)
10% Critical Value: Predicted	1.908	1.797	1.725	1.686
95% Conf. Int.	[1.911, 2.061]	[1.769, 1.892]	[1.661, 1.799]	[1.629, 1.794]
5% Critical Value: Predicted	2.283	2.158	2.068	2.016
95% Conf. Int.	[2.275, 2.527]	[2.101, 2.342]	[1.978, 2.191]	[1.983, 2.170]
1% Critical Value: Predicted	2.998	2.867	2.749	2.670
95% Conf. Int.	[2.967, 3.737]	[2.776, 3.248]	[2.586, 3.080]	[2.669, 3.171]
Rej. Frequencies: True coefficients				
10% Level (S.E.)	11.15 (0.51)	10.70 (0.45)	9.64 (0.38)	9.97 (0.31)
5% Level (S.E.)	5.96 (0.38)	5.05 (0.33)	4.77 (0.31)	4.92 (0.27)
1% Level (S.E.)	1.31 (0.23)	1.17 (0.19)	1.04 (0.13)	1.14 (0.15)
Rej. Frequencies: Est. coefficients				
10% Level (S.E.)	14.38* (0.55)	11.77* (0.49)	9.94 (0.39)	9.98 (0.33)
5% Level (S.E.)	8.98* (0.49)	6.63* (0.39)	5.31 (0.32)	5.11 (0.27)
1% Level (S.E.)	3.81* (0.36)	2.21* (0.26)	1.56* (0.22)	1.37 (0.15)
Restricted residuals ($\mathbf{HCR_0}$)				
Standard Deviation: Predicted	1.018	1.013	1.008	1.004
Estimated (S.E.)	1.023 (0.016)	1.007 (0.016)	1.024 (0.016)	1.025 (0.016)
10% Critical Value: Predicted	1.684	1.672	1.660	1.653
95% Conf. Int.	[1.600, 1.724]	[1.611, 1.731]	[1.610, 1.725]	[1.600, 1.752]
5% Critical Value: Predicted	1.945	1.960	1.962	1.962
95% Conf. Int.	[1.871, 2.004]	[1.880, 2.019]	[1.888, 2.041]	[1.936, 2.098]
1% Critical Value: Predicted	2.369	2.476	2.527	2.552
95% Conf. Int.	[2.214, 2.543]	[2.338, 2.701]	[2.448, 2.788]	[2.558, 2.865]
Rej. Frequencies: True coefficients				
10% Level (S.E.)	9.56 (0.52)	9.99 (0.46)	10.10 (0.40)	10.02 (0.31)
5% Level (S.E.)	4.82 (0.42)	4.90 (0.37)	4.72 (0.31)	4.88 (0.25)
1% Level (S.E.)	0.79 (0.19)	0.87 (0.19)	1.14 (0.17)	1.26 (0.19)
Rej. Frequencies: Est. coefficients				
10% Level (S.E.)	9.56 (0.52)	9.99 (0.47)	10.12 (0.41)	9.88 (0.31)
5% Level (S.E.)	5.30 (0.43)	5.67 (0.39)	4.96 (0.31)	5.28 (0.26)
1% Level (S.E.)	1.20 (0.23)	1.62* (0.23)	1.40 (0.20)	1.42 (0.18)

Table 7. Performance of Edgeworth Approximations: Case 2

Sample Size:	50	100	200	400
Unrestricted residuals (HC_0)				
Standard Deviation: Predicted	1.218	1.146	1.084	1.045
Estimated (S.E.)	1.365 (0.022)	1.228 (0.019)	1.108 (0.018)	1.073 (0.017)
10% Critical Value: Predicted	2.009	1.895	1.785	1.718
95% Conf. Int.	[2.117, 2.289]	[1.900, 2.086]	[1.771, 1.953]	[1.664, 1.849]
5% Critical Value: Predicted	2.375	2.269	2.144	2.059
95% Conf. Int.	[2.524, 2.732]	[2.358, 2.664]	[2.131, 2.306]	[2.020, 2.206]
1% Critical Value: Predicted	3.066	2.986	2.849	2.735
95% Conf. Int.	[3.232, 3.789]	[3.171, 3.712]	[2.710, 3.144]	[2.670, 3.182]
Rej. Frequencies: True coefficients				
10% Level (S.E.)	13.37* (0.59)	10.90 (0.52)	10.99 (0.48)	9.91 (0.36)
5% Level (S.E.)	8.08* (0.52)	6.28* (0.42)	5.88 (0.40)	4.96 (0.30)
1% Level (S.E.)	3.50* (0.38)	1.82* (0.27)	1.12 (0.18)	0.91 (0.15)
Rej. Frequencies: Est. coefficients				
10% Level (S.E.)	17.06* (0.65)	13.26* (0.56)	11.99* (0.50)	9.91 (0.36)
5% Level (S.E.)	12.08* (0.58)	8.65* (0.49)	7.62* (0.43)	5.73 (0.31)
1% Level (S.E.)	6.64* (0.50)	3.91* (0.38)	2.32* (0.27)	1.54* (0.18)
Restricted residuals (HCR_0)				
Standard Deviation: Predicted	1.034	1.031	1.019	1.011
Estimated (S.E.)	1.039 (0.016)	1.058 (0.017)	1.031 (0.016)	1.032 (0.016)
10% Critical Value: Predicted	1.708	1.701	1.680	1.665
95% Conf. Int.	[1.633, 1.717]	[1.681, 1.807]	[1.657, 1.808]	[1.623, 1.776]
5% Critical Value: Predicted	1.939	1.976	1.977	1.971
95% Conf. Int.	[1.830, 1.979]	[1.957, 2.117]	[1.955, 2.088]	[1.915, 2.083]
1% Critical Value: Predicted	2.276	2.446	2.520	2.551
95% Conf. Int.	[2.284, 2.601]	[2.423, 2.688]	[2.371, 2.600]	[2.489, 2.830]
Rej. Frequencies: True coefficients				
10% Level (S.E.)	9.01 (0.58)	10.39 (0.54)	10.69 (0.48)	9.82 (0.37)
5% Level (S.E.)	4.63 (0.45)	5.49 (0.43)	5.72 (0.42)	4.99 (0.31)
1% Level (S.E.)	1.57 (0.27)	1.34 (0.25)	1.06 (0.16)	0.91 (0.15)
Rej. Frequencies: Est. coefficients				
10% Level (S.E.)	10.02 (0.60)	10.63 (0.54)	10.69 (0.48)	9.73 (0.37)
5% Level (S.E.)	4.88 (0.45)	6.27* (0.44)	6.25* (0.42)	5.18 (0.31)
1% Level (S.E.)	1.95* (0.30)	2.19* (0.31)	2.04* (0.23)	1.63* (0.20)

Table 8. Performance of Edgeworth Approximations: Case 3

Sample Size:	50	100	200	400
Unrestricted residuals ($\mathbf{HC_0}$)				
Standard Deviation: Predicted	1.190	1.119	1.066	1.035
Estimated (S.E.)	1.285 (0.020)	1.156 (0.018)	1.057 (0.017)	1.073 (0.017)
10% Critical Value: Predicted	1.969	1.848	1.755	1.702
95% Conf. Int.	[2.065, 2.221]	[1.823, 1.995]	[1.670, 1.823]	[1.698, 1.856]
5% Critical Value: Predicted	2.344	2.219	2.107	2.037
95% Conf. Int.	[2.427, 2.619]	[2.176, 2.409]	[2.027, 2.228]	[2.047, 2.189]
1% Critical Value: Predicted	3.052	2.937	2.804	2.704
95% Conf. Int.	[3.181, 4.223]	[2.954, 3.496]	[2.600, 3.160]	[2.631, 2.989]
Rej. Frequencies: True coefficients				
10% Level (S.E.)	13.20* (0.56)	9.99 (0.47)	9.93 (0.42)	9.89 (0.40)
5% Level (S.E.)	7.50* (0.49)	5.33 (0.39)	4.92 (0.28)	5.40 (0.31)
1% Level (S.E.)	1.61 (0.25)	1.45 (0.21)	1.12 (0.18)	1.13 (0.16)
Rej. Frequencies: Est. coefficients				
10% Level (S.E.)	16.67* (0.63)	11.73* (0.50)	10.43 (0.45)	10.28 (0.39)
5% Level (S.E.)	12.25* (0.58)	6.56* (0.42)	5.75 (0.34)	5.97* (0.32)
1% Level (S.E.)	5.88* (0.47)	2.95* (0.32)	1.50 (0.23)	1.43 (0.19)
Restricted residuals ($\mathbf{HCR_0}$)				
Standard Deviation: Predicted	1.037	1.027	1.016	1.009
Estimated (S.E.)	1.052 (0.017)	1.040 (0.016)	1.000 (0.016)	1.041 (0.016)
10% Critical Value: Predicted	1.710	1.694	1.674	1.661
95% Conf. Int.	[1.672, 1.753]	[1.635, 1.758]	[1.595, 1.696]	[1.649, 1.815]
5% Critical Value: Predicted	1.965	1.981	1.976	1.970
95% Conf. Int.	[1.906, 2.020]	[1.914, 2.097]	[1.886, 2.066]	[1.980, 2.112]
1% Critical Value: Predicted	2.367	2.491	2.540	2.560
95% Conf. Int.	[2.358, 2.626]	[2.397, 2.775]	[2.335, 2.723]	[2.497, 2.712]
Rej. Frequencies: True coefficients				
10% Level (S.E.)	10.03 (0.58)	9.29 (0.51)	9.13 (0.41)	9.91 (0.39)
5% Level (S.E.)	4.86 (0.44)	4.89 (0.39)	4.78 (0.31)	5.59 (0.32)
1% Level (S.E.)	1.37 (0.25)	0.96 (0.21)	1.10 (0.18)	0.97 (0.17)
Rej. Frequencies: Est. coefficients				
10% Level (S.E.)	11.49 (0.61)	9.57 (0.51)	9.14 (0.42)	9.86 (0.38)
5% Level (S.E.)	5.88 (0.47)	5.28 (0.41)	4.95 (0.30)	5.89* (0.31)
1% Level (S.E.)	1.69* (0.26)	1.69* (0.25)	1.36 (0.21)	1.33 (0.18)

Table 9. Performance of Edgeworth Approximations: Case 4

Sample Size:	50	100	200	400
Unrestricted residuals (HC_0)				
Standard Deviation: Predicted	1.230	1.167	1.099	1.054
Estimated (S.E.)	1.481 (0.023)	1.211 (0.019)	1.108 (0.017)	1.077 (0.017)
10% Critical Value: Predicted	2.002	1.924	1.811	1.734
95% Conf. Int.	[2.281, 2.460]	[1.900, 2.091]	[1.760, 1.927]	[1.679, 1.833]
5% Critical Value: Predicted	2.336	2.286	2.167	2.075
95% Conf. Int.	[2.687, 2.964]	[2.371, 2.526]	[2.162, 2.360]	[2.067, 2.262]
1% Critical Value: Predicted	2.967	2.977	2.861	2.748
95% Conf. Int.	[2.142, 4.417]	[2.996, 3.701]	[2.760, 3.058]	[2.692, 3.123]
Rej. Frequencies: True coefficients				
10% Level (S.E.)	16.58* (0.73)	11.30 (0.55)	10.67 (0.49)	10.39 (0.44)
5% Level (S.E.)	10.62* (0.65)	6.25* (0.47)	6.54* (0.41)	5.87 (0.37)
1% Level (S.E.)	4.01* (0.43)	1.57 (0.27)	1.10 (0.22)	1.24 (0.19)
Rej. Frequencies: Est. coefficients				
10% Level (S.E.)	19.85* (0.77)	14.51* (0.59)	11.52* (0.52)	10.54 (0.45)
5% Level (S.E.)	14.97* (0.72)	8.94* (0.52)	7.62* (0.44)	6.44* (0.40)
1% Level (S.E.)	7.70* (0.57)	4.65* (0.41)	2.75* (0.30)	1.79* (0.19)
Restricted residuals (HCR_0)				
Standard Deviation: Predicted	1.035	1.033	1.021	1.011
Estimated (S.E.)	1.074 (0.017)	1.013 (0.016)	1.014 (0.016)	1.025 (0.016)
10% Critical Value: Predicted	1.712	1.706	1.684	1.667
95% Conf. Int.	[1.592, 1.680]	[1.608, 1.708]	[1.603, 1.752]	[1.606, 1.735]
5% Critical Value: Predicted	1.914	1.960	1.968	1.966
95% Conf. Int.	[1.780, 1.940]	[1.840, 1.974]	[1.906, 2.053]	[1.946, 2.112]
1% Critical Value: Predicted	2.179	2.362	2.467	2.522
95% Conf. Int.	[2.207, 2.579]	[2.171, 2.509]	[2.318, 2.505]	[2.438, 2.722]
Rej. Frequencies: True coefficients				
10% Level (S.E.)	7.98* (0.60)	8.85 (0.55)	10.21 (0.50)	10.29 (0.44)
5% Level (S.E.)	4.38 (0.46)	4.16 (0.43)	5.55 (0.39)	5.73 (0.38)
1% Level (S.E.)	1.64 (0.28)	0.86 (0.21)	0.62 (0.17)	1.00 (0.19)
Rej. Frequencies: Est. coefficients				
10% Level (S.E.)	8.29* (0.61)	9.30 (0.56)	10.21 (0.50)	10.24 (0.44)
5% Level (S.E.)	4.75 (0.47)	4.76 (0.45)	6.36* (0.42)	6.05* (0.39)
1% Level (S.E.)	1.40 (0.26)	1.28 (0.25)	1.63 (0.25)	1.74* (0.20)