Queen's Economics Department Working Paper No. 717

Heteroskedasticity-Robust Tests for Structural Change

James G. MacKinnon
Queen's University

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

5-1988

Heteroskedasticity-Robust Tests for Structural Change

James G. MacKinnon

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

Abstract

It is remarkably easy to test for structural change, of the type that the classic F or "Chow" test is designed to detect, in a manner that is robust to heteroskedasticity of possibly unknown form. This paper first discusses how to test for structural change in nonlinear regression models by using a variant of the Gauss-Newton regression. It then shows how to make these tests robust to heteroskedasticity of unknown form and discusses several related procedures for doing so. Finally, it presents the results of a number of Monte Carlo experiments designed to see how well the new tests perform in finite samples.

This research was supported, in part, by the Social Sciences and Humanities Research Council of Canada. I am grateful to Allan Gregory and Simon Power for helpful comments on an earlier draft. This paper was published in the *Empirical Economics*, 14, 1989, 77–92.

May, 1988

1. Introduction

A classic problem in econometrics is testing whether the coefficients of a regression model are the same in two or more separate subsamples. In the case of time-series data, where the subsamples generally correspond to different economic environments, such as different exchange-rate or policy regimes, such tests are generally referred to as tests for structural change. They are equally applicable to cross-section data, where the subsamples might correspond to different groups of observations such as large firms and small firms, rich countries and poor countries, or men and women. Evidently, there could well be more than two such groups of observations.

The classical F test for the equality of two sets of coefficients in linear regression models is commonly referred to by economists as the Chow test, after the early and influential paper by Chow (1960). Another exposition of this procedure is Fisher (1970). The classic approach is to partition the data into two parts, possibly after reordering. The n–vector $\boldsymbol{y}$ of observations on the dependent variable is divided into an n_1–vector $\boldsymbol{y}_1$ and an n_2–vector $\boldsymbol{y}_2$, and the $n \times k$ matrix $\boldsymbol{X}$ of observations on the regressors is divided into an $n_1 \times k$ matrix $\boldsymbol{X}_1$ and an $n_2 \times k$ matrix $\boldsymbol{X}_2$, with $n = n_1 + n_2$. Thus the maintained hypothesis may be written as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 & \mathbf{O} \\ \mathbf{O} & \boldsymbol{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix}, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma_2 \mathbf{I}, \tag{1}$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are each k–vectors of parameters to be estimated. The null hypothesis to be tested is that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}$. Under it, (1) reduces to

$$\boldsymbol{y} \equiv \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix} \equiv \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma_2 \mathbf{I}. \tag{2}$$

In the usual case where both n_1 and n_2 are greater than k, it is easy to construct a test of (2) against (1) by using an ordinary F test. The unrestricted sum of squared residuals from OLS estimation of (1) is

$$\mathrm{USSR} = \mathrm{SSR}_1 + \mathrm{SSR}_2 = \boldsymbol{y}_1{}^\top \boldsymbol{M}_1 \boldsymbol{y}_1 + \boldsymbol{y}_2{}^\top \boldsymbol{M}_2 \boldsymbol{y}_2, \tag{3}$$

where $\boldsymbol{M}_i = \mathbf{I} - \boldsymbol{X}_i(\boldsymbol{X}_i^\top \boldsymbol{X}_i)^{-1}\boldsymbol{X}_i^\top$ for $i = 1, 2$ denotes the $n \times n$ matrix that projects orthogonally off the subspace spanned by the columns of the matrix $\boldsymbol{X}_i$. The vectors $\boldsymbol{M}_1\boldsymbol{y}_1$ and $\boldsymbol{M}_2\boldsymbol{y}_2$ are the residuals from the regressions of $\boldsymbol{y}_1$ on $\boldsymbol{X}_1$ and $\boldsymbol{y}_2$ on $\boldsymbol{X}_2$, respectively. Thus USSR is simply the sum of the two sums of squared residuals.

The restricted sum of squared residuals, from OLS estimation of (2), is

$$\mathrm{RSSR} = \boldsymbol{y}^\top \boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{y}, \tag{4}$$

where $\boldsymbol{M}_{\boldsymbol{X}} = \mathbf{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top$. Thus the ordinary F statistic is

$$\frac{(\boldsymbol{y}^\top \boldsymbol{M_X} \boldsymbol{y} - \boldsymbol{y}_1{}^\top \boldsymbol{M}_1 \boldsymbol{y}_1 - \boldsymbol{y}_2{}^\top \boldsymbol{M}_2 \boldsymbol{y}_2)/k}{(\boldsymbol{y}_1{}^\top \boldsymbol{M}_1 \boldsymbol{y}_1 + \boldsymbol{y}_2{}^\top \boldsymbol{M}_2 \boldsymbol{y}_2)/(n-2k)} = \frac{(\mathrm{RSSR} - \mathrm{SSR}_1 - \mathrm{SSR}_2)/k}{(\mathrm{SSR}_1 + \mathrm{SSR}_2)/(n-2k)}. \tag{5}$$

This test statistic, which is what many applied econometricians refer to as the "Chow statistic", has k and $n-2k$ degrees of freedom, because the unrestricted model has $2k$ parameters while the restricted model has only k. It will be exactly distributed as $F(k, n-2k)$ if the error vector $\boldsymbol{u}$ is normal and independent of the fixed regressors $\boldsymbol{X}$, and k times it will be asymptotically distributed as $\chi^2(k)$ under much weaker conditions.

Tests based on the Chow statistic (5) have one obvious and very serious limitation. Like all conventional F tests, they are (in general) valid only under the rather strong assumption that $\mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma^2 \mathbf{I}$. This assumption may be particularly implausible when one is testing the equality of two sets of regression parameters, since if the parameter vector $\boldsymbol{\beta}$ differs between two regimes the variance σ^2 may well be different as well. A number of papers have addressed this issue, including Toyoda (1974), Jayatissa (1977), Schmidt and Sickles (1977), Watt (1979), Honda (1982), Phillips and McCabe (1983), Ohtani and Toyoda (1985), Toyoda and Ohtani (1986), and Weerahandi (1987). However, none of these papers proposes the very simple approach of using a test which is robust to heteroskedasticity of unknown form. The work of Eicker (1963) and White (1980) has made such tests available, and Davidson and MacKinnon (1985) have provided simple ways to calculate them using artificial regressions. In this paper, I show how the results of the latter paper may be used to calculate several heteroskedasticity-robust variants of the Chow test.

The plan of the paper is as follows. In Section 2, I discuss how to test for structural change in nonlinear regression models by using a variant of the Gauss-Newton regression. In Section 3, I then discuss ways to make the tests discussed in Section 2 robust to heteroskedasticity of unknown form. Finally, in Section 4, I present the results of some Monte Carlo experiments designed to see how well the new tests perform in finite samples.

2. Testing for Structural Change in Nonlinear Regression Models

Nonlinear regression models may seem unnecessarily complicated, but studying them makes it easier to see how to make Chow-type tests robust to heteroskedasticity. Suppose that the null hypothesis is

$$\mathrm{H}_0: \quad y_t = x_t(\boldsymbol{\beta}) + u_t, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma^2 \mathbf{I}, \tag{6}$$

where the regression functions $x_t(\boldsymbol{\beta})$, which may depend on exogenous and/or lagged dependent variables and on a k–vector of parameters $\boldsymbol{\beta}$, are assumed to be twice continuously differentiable. The matrix $\boldsymbol{X}(\boldsymbol{\beta})$, with typical element

$$X_{ti}(\boldsymbol{\beta}) = \frac{\partial x_t(\boldsymbol{\beta})}{\partial \beta_i}, \tag{7}$$

will play a major role in the analysis. In the case of the linear regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$, $\boldsymbol{X}(\boldsymbol{\beta})$ is simply the matrix $\boldsymbol{X}$. It is assumed that

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\boldsymbol{X}^{\top}(\boldsymbol{\beta})\boldsymbol{X}(\boldsymbol{\beta})\Big) \tag{8}$$

exists and is a positive-definite matrix.

For simplicity, it will be assumed that the sample is to be divided into only two groups of observations; extensions to the many-group case are obvious. We first define a vector $\boldsymbol{\delta} \equiv [\delta_1 \ldots \delta_n]^{\top}$, letting $\delta_t = 0$ if observation t belongs to group 1 and $\delta_t = 1$ if observation t belongs to group 2. Note that it would be possible to let δ_t, take on values between zero and one for some observations, which might be useful if it were thought that the transition between regimes was gradual rather than abrupt. If the null hypothesis is (6), the alternative hypothesis may be written as

$$\mathrm{H}_1: \;\; y_t = x_t\big(\boldsymbol{\beta}_1(1-\delta_t) + \boldsymbol{\beta}_2\delta_t\big) + u_t, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^{\top}) = \sigma^2\mathbf{I}. \tag{9}$$

Thus the regression function is $x_t(\boldsymbol{\beta}_1)$ if $\delta_t = 0$ and $x_t(\boldsymbol{\beta}_2)$ if $\delta_t = 1$.

The alternative hypothesis H_1 can be rewritten as

$$y_t = x_t\big(\boldsymbol{\beta}_1 + (\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)\delta_t\big) + u_t = x_t(\boldsymbol{\beta}_1 + \boldsymbol{\gamma}\delta_t) + u_t, \tag{10}$$

where $\boldsymbol{\gamma} \equiv \boldsymbol{\beta}_2 - \boldsymbol{\beta}_1$. This makes it clear that H_0 is equivalent to the null hypothesis is that $\boldsymbol{\gamma} = \mathbf{0}$. Since the latter is simply a set of zero restrictions on the parameters of a nonlinear regression function, we can use a Gauss-Newton regression to test it; see Engle (1982b) or Davidson and MacKinnon (1984). The Gauss-Newton regression, or GNR, for testing H_0 against H_1 is easily seen to be

$$y_t - x_t(\tilde{\boldsymbol{\beta}}) = \boldsymbol{X}_t(\tilde{\boldsymbol{\beta}})\boldsymbol{b} + \delta_t\boldsymbol{X}_t(\tilde{\boldsymbol{\beta}})\boldsymbol{c} + \text{residuals}, \tag{11}$$

where $\tilde{\boldsymbol{\beta}}$ denotes the vector of nonlinear least squares (NLS) estimates of $\boldsymbol{\beta}$ for the whole sample.

The GNR (11) may be written more compactly as

$$\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{X}}\boldsymbol{b} + \boldsymbol{\delta} * \tilde{\boldsymbol{X}}\boldsymbol{c} + \text{residuals}, \tag{12}$$

where $\tilde{\boldsymbol{u}}$ is an n–vector with typical element $y_t - x_t(\tilde{\boldsymbol{\beta}})$, and $\tilde{\boldsymbol{X}}$ is an $n \times k$ matrix with typical row $\boldsymbol{X}_t(\tilde{\boldsymbol{\beta}})$. Here "$*$" denotes the direct product of two matrices, a typical element of $\boldsymbol{\delta} * \boldsymbol{X}$ being $\delta_t X_{ti}(\tilde{\boldsymbol{\beta}})$, so that $\boldsymbol{\delta} * \tilde{\boldsymbol{X}}$, equals $\tilde{\boldsymbol{X}}_t$ when $\delta_t = 1$ and $\mathbf{0}$ when $\delta_t = 0$. Thus we can perform the test by estimating the model using the entire sample and regressing the residuals on the matrix of derivatives $\tilde{\boldsymbol{X}}$ and on the matrix $\boldsymbol{\delta} * \tilde{\boldsymbol{X}}$, which is $\tilde{\boldsymbol{X}}$ with the rows that correspond to group 1 observations set to zero. There is no need to reorder the data. Several asymptotically valid test statistics can then be computed, including the ordinary F statistic for the null hypothesis that

$\boldsymbol{c} = \mathbf{0}$. In the usual case where k is less than $\min(n_1, n_2)$, it will have k degrees of freedom in the numerator and $n - 2k$ degrees of freedom in the denominator.

Unlike the ordinary "Chow statistic" given in (5), this procedure is applicable even if $\min(n_1, n_2) < k$. Suppose, without loss of generality, that $n_2 < k$ and $n_1 > k$. Then the matrix $\boldsymbol{\delta} * \tilde{\boldsymbol{X}}$, which has k columns, will have $n_2 < k$ rows that are not just rows of zeros, and hence will have rank at most n_2. When equation (12) is estimated, at most n_2 elements of $\boldsymbol{c}$ will be identifiable, and the residuals corresponding to all observations which belong to group 2 will be zero. Thus the degrees of freedom for the numerator of the F statistic, which is equal to the rank of $[\tilde{\boldsymbol{X}} \quad \boldsymbol{\delta} * \tilde{\boldsymbol{X}}]$ minus the rank of $\tilde{\boldsymbol{X}}$, must be at most n_2. The degrees of freedom for the denominator will normally be $n_1 - k$. Note that when $x_t(\boldsymbol{\beta}) = \boldsymbol{X\beta}$ and $\min(n_1, n_2) > k$, the F statistic based on the GNR (12) is numerically identical to the "Chow statistic" (5). This follows from the fact that the sum of squared residuals from (12) will then be equal to $\mathrm{SSR}_1 + \mathrm{SSR}_2$, the sum of the SSRs from estimating the regression separately over the two groups of observations.

It may be of interest to test whether a subset of the parameters of a model, rather than all of the parameters, is the same over two (or more) subsamples. It is easy to modify the tests already discussed to deal with this case. The null and alternative hypotheses can now be written as

$$\mathrm{H}_0: \; y_t = x_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) + u_t, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma^2 \mathbf{I}, \tag{13}$$

and

$$\mathrm{H}_1: \; y_t = x_t\big(\boldsymbol{\alpha}, (1 - \delta_t)\boldsymbol{\beta}_1 + \delta_t \boldsymbol{\beta}_2\big) + u_t, \quad \mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma^2 \mathbf{I}, \tag{14}$$

where $\boldsymbol{\alpha}$ is an l–vector of parameters that are assumed to be the same over the two subsamples, and $\boldsymbol{\beta}$ is an m–vector of parameters the constancy of which is to be tested. The Gauss-Newton regression is easily seen to be

$$\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{X}}_{\boldsymbol{\alpha}} \boldsymbol{a} + \tilde{\boldsymbol{X}}_{\boldsymbol{\beta}} \boldsymbol{b} + \boldsymbol{\delta} * \tilde{\boldsymbol{X}}_{\boldsymbol{\beta}} \boldsymbol{c} + \text{residuals}, \tag{15}$$

where $\tilde{\boldsymbol{X}}_{\boldsymbol{\alpha}}$ is an $n \times l$ matrix with typical element $\partial x_t(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial \alpha_i$, and $\tilde{\boldsymbol{X}}_{\boldsymbol{\beta}}$ is an $n \times m$ matrix with typical element $\partial x_t(\boldsymbol{\alpha}, \boldsymbol{\beta})/\partial \beta_j$, both evaluated at the estimates $(\tilde{\boldsymbol{\alpha}}, \tilde{\boldsymbol{\beta}})$ from (13). One would then use the F statistic for $\boldsymbol{c} = \mathbf{0}$, which if $m < \min(n_1, n_2)$ will have m and $n - l - 2m$ degrees of freedom.

There are several asymptotically equivalent test statistics which may be calculated from the artificial regression (12). They all have the same numerator, which is the explained sum of squares from that regression. The denominator may be anything that consistently estimates σ^2, and if the statistic is to be compared to the $F(k, 2n-k)$ rather than the $\chi^2(k)$ distribution, it must first be multiplied by $(n - 2k)/k$. If we let $\tilde{\boldsymbol{Z}}$ denote $\boldsymbol{\delta} * \tilde{\boldsymbol{X}}$, then the numerator of all the test statistics is

$$\tilde{\boldsymbol{u}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{Z}} (\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{Z}})^{-1} \tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{u}}, \tag{16}$$

where $\boldsymbol{M}_{\tilde{\boldsymbol{X}}} \equiv \mathbf{I} - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}^\top$. What may be the best of the many valid test statistics is the ordinary F statistic for $\boldsymbol{c} = \boldsymbol{0}$ in (12), which is

$$\frac{\tilde{\boldsymbol{u}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}}(\tilde{\boldsymbol{Z}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}})^{-1}\tilde{\boldsymbol{Z}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{u}}/k}{\tilde{\boldsymbol{u}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}};\,\tilde{\boldsymbol{Z}}}\tilde{\boldsymbol{u}}/(n-2k)}, \tag{17}$$

where $\boldsymbol{M}_{\tilde{\boldsymbol{X}};\,\tilde{\boldsymbol{Z}}}$ is the matrix that projects orthogonally off the subspace spanned by $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{Z}}$ jointly. Expression (17) is just $(n-2k)/k$ times the explained sum of squares from the GNR (12) divided by the sum of squared residuals that artificial regression.

Rewriting expression (16) so that all factors are $O_p(1)$, we obtain

$$(n^{-1/2}\tilde{\boldsymbol{u}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}})(n^{-1}\tilde{\boldsymbol{Z}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}})^{-1}(n^{-1/2}\tilde{\boldsymbol{Z}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{u}}). \tag{18}$$

This expression is a quadratic form in the vector

$$n^{-1/2}\tilde{\boldsymbol{Z}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{u}}. \tag{19}$$

Standard asymptotic theory tells us that this vector is asymptotically normally distributed with mean vector zero and covariance matrix

$$\sigma^2 \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}}\Big). \tag{20}$$

The middle matrix in (18), times anything that consistently estimates σ^2, provides a consistent estimate of the inverse of (20). Thus (18), divided by anything that consistently estimates σ^2, must be asymptotically distributed as $\chi^2(k)$.

The key point which emerges from the above discussion is that every test statistic based on the GNR (12) is actually testing whether the k–vector (19) has mean zero asymptotically. Under relatively weak assumptions, this vector will be asymptotically normal, since it is essentially a weighted sum of n independent random variables, namely, the elements of the vector $\boldsymbol{u}$. Under the much stronger assumption of homoskedasticity, its asymptotic covariance matrix will be given by (20), which allows us to use tests based on the GNR. Without this assumption, we will still be able to compute test statistics as quadratic forms in the vector $n^{-1/2}\tilde{\boldsymbol{Z}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{u}}$ and expect them to be asymptotically distributed as $\chi^2(k)$, provided that we can somehow obtain an estimate of the asymptotic covariance matrix of $n^{-1/2}\tilde{\boldsymbol{Z}}^\top\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{u}}$ that is consistent in the presence of heteroskedasticity. How this may be done is discussed in the next section.

3. Heteroskedasticity-Robust Tests

We are now ready to drop the often implausible assumption that $\mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \sigma^2\mathbf{I}$. Instead, we shall assume initially that

$$\mathrm{E}(\boldsymbol{u}\boldsymbol{u}^\top) = \boldsymbol{\Omega}, \quad \Omega_{tt} = \sigma_2^2, \quad \Omega_{ts} = 0 \text{ for } t \neq s, \quad 0 < \sigma_t < \sigma_{\max}. \tag{21}$$

Thus $\boldsymbol{\Omega}$, the covariance matrix of the error terms $\boldsymbol{u}$, is assumed to be an $n \times n$ diagonal matrix with σ_t^2 as its t^{th} diagonal element. Except that σ_t is assumed to be bounded from above by some possibly very large number $\sigma_{\max}$, we are not assuming that anything is known about the σ_t. Since there is nothing that prevents σ_t from depending on variables which affect $x_t(\boldsymbol{\beta})$, these assumptions admit virtually any interesting pattern of heteroskedasticity, including autoregressive conditional heteroskedasticity, or ARCH; see Engle (1982a). They do, however, rule out serial correlation or any other sort of dependence across observations.

Under the assumptions (21), it is easy to see that the asymptotic covariance matrix of the vector (19) is

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}}\boldsymbol{\Omega}\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}}\Big). \tag{22}$$

It is, in general, not possible to estimate $\boldsymbol{\Omega}$ an $n \times n$ matrix which in this case has n non-zero elements, consistently. However, by a slight modification of the arguments used by White (1980), one can show that the matrix

$$\frac{1}{n}\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}}\acute{\boldsymbol{\Omega}}\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}} \tag{23}$$

consistently estimates (22), where $\acute{\boldsymbol{\Omega}}$ is an $n \times n$ diagonal matrix with $\acute{\sigma}_t^2$ as the t^{th} diagonal element, and the diagonal elements $\acute{\sigma}_t^2$ have the property that

$$\acute{\sigma}_t^2 \to \sigma_t^2 + v_t \ \text{ as } n \to \infty. \tag{24}$$

Here v_t is a random variable which asymptotically has mean zero and finite variance and is independent of $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{Z}}$. There are many choices for $\acute{\sigma}_t^2$, of which the simplest is just $\tilde{u}_t^2$, the square of the t^{th} residual from the initial NLS estimation of H_0.

Combining (19) and (23), we obtain the family of test statistics

$$\begin{aligned}&(n^{-1/2}\tilde{\boldsymbol{u}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}})(n^{-1}\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}}\acute{\boldsymbol{\Omega}}\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}})^{-1}(n^{-1/2}\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{u}})\\ &= \tilde{\boldsymbol{u}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}}(\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}}\acute{\boldsymbol{\Omega}}\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}})^{-1}\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{u}}.\end{aligned} \tag{25}$$

Since $n^{-1/2}\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{u}}$ is asymptotically normal with covariance matrix (22), and the matrix (23) consistently estimates (22), it is clear that (25) will be asymptotically distributed as $\chi^2(k)$ under H_0. As shown by Davidson and MacKinnon (1985), variants of (25) can be computed by means of two different artificial regressions. The most generally applicable of these is

$$\tilde{u}_t/\acute{\sigma}_t = \acute{\sigma}_t(\boldsymbol{M}_{\tilde{\boldsymbol{X}}}\tilde{\boldsymbol{Z}})_t\boldsymbol{c} + \text{residual}. \tag{26}$$

The explained sum of squares from regression (26) is the test statistic (25). The inner product of the regressor matrix with itself is $\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \acute{\boldsymbol{\Omega}} \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{Z}}$, and its inner product with the regressand is $\tilde{\boldsymbol{u}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{Z}}$. The latter expression does not involve $\acute{\sigma}_t$, because the $\acute{\sigma}_t$ which multiplies each of the regressors cancels with the $1/\acute{\sigma}_t$ which multiplies the regressand. For regression (26) to be computable, $\acute{\sigma}_t$ must never be exactly equal to zero, since, if it were, the regressand would be undefined; this problem can be avoided in practice by setting $\acute{\sigma}_t$ to a very small number whenever it should really be zero.

If $\tilde{u}_t$ is used for $\acute{\sigma}_t$, and it is probably the most natural choice, an even simpler artificial regression is available. It is

$$\boldsymbol{\iota} = \tilde{\boldsymbol{U}} \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{Z}} \boldsymbol{c} + \text{errors}, \tag{27}$$

where $\boldsymbol{\iota}$ is an n–vector of ones, and $\tilde{\boldsymbol{U}}$ is an $n \times n$ diagonal matrix with $\tilde{u}_t$ as the t^{th} diagonal element. The explained sum of squares from regression (27) is

$$\boldsymbol{\iota}^\top \tilde{\boldsymbol{U}} \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{Z}} (\tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{U}} \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{Z}})^{-1} \tilde{\boldsymbol{Z}}^\top \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{U}} \boldsymbol{\iota}. \tag{28}$$

The vector $\boldsymbol{\iota}^\top \tilde{\boldsymbol{U}}$ is simply $\tilde{\boldsymbol{u}}^\top$, and the matrix $\tilde{\boldsymbol{U}}^\top \tilde{\boldsymbol{U}}$ is simply $\acute{\boldsymbol{\Omega}}$ with $\tilde{u}_t^2$ being used for $\acute{\sigma}_t^2$, so that (28) is just a special case of (25).

The artificial regression (27) is very easy to compute. The regressand is a vector of ones. Each of the regressors is the vector of residuals from a regression of $\tilde{\boldsymbol{Z}}$ on $\tilde{\boldsymbol{X}}$, each element of which has been multiplied by the appropriate element of $\tilde{\boldsymbol{u}}$ (to see this, observe that $\tilde{\boldsymbol{U}} \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{Z}} = \tilde{\boldsymbol{u}} * \boldsymbol{M}_{\tilde{\boldsymbol{X}}} \tilde{\boldsymbol{Z}}$). Thus one simply has to perform $k+1$ linear regressions. Since k of them involve the same set of regressors, namely, the matrix $\tilde{\boldsymbol{X}}$, the computational burden (given appropriate software) is only moderately greater than that of performing two linear regressions.

There are other choices for $\acute{\sigma}_t^2$ besides $\tilde{u}_t^2$. One that was proposed in the context of heteroskedasticity-consistent covariance matrix estimators (HCCMEs) for linear regression models by MacKinnon and White (1985) is

$$\ddot{\sigma}_t^2 = \tilde{u}_t^2 / (\boldsymbol{M}_{\tilde{\boldsymbol{X}}})_{tt}, \tag{29}$$

where $(\boldsymbol{M}_{\tilde{\boldsymbol{X}}})_{tt}$ denotes the t^{th} diagonal element of the matrix $\boldsymbol{M}_{\tilde{\boldsymbol{X}}}$. The reason for using (29) is that, in the case of a linear regression model with homoskedastic residuals, it provides an unbiased estimate of $\sigma_t^2 = \sigma^2$, correcting the tendency of squared residuals to be too small.

In the context of testing for structural change, assumptions (21) may seem more unrestrictive than is needed. What has traditionally worried econometricians about the ordinary F test is not the possibility that there may be heteroskedasticity of unknown form, but the possibility that the variance of the error terms may simply be different in the two sub-samples. It is easy to derive a version which allows only for this possibility. First, estimate the model over each of the two groups of observations, obtaining sums of squared residuals SSR_1, and SSR_2, respectively. Then make the definitions:

$$\hat{\sigma}_1 = \left(\frac{\mathrm{SSR}_1}{n_1 - k}\right)^{1/2} \quad \text{and} \quad \hat{\sigma}_2 = \left(\frac{\mathrm{SSR}_2}{n_2 - k}\right)^{1/2}, \tag{30}$$

and let $\acute{\sigma}_t = \hat{\sigma}_1$ for all observations where $\delta_t = 0$ and $\acute{\sigma}_t = \hat{\sigma}_2$ for all observations where $\delta_t = 1$. Now run regression (26) using the $\acute{\sigma}_t$ so defined. The explained sum of squares from this regression will have the form of (25), and it will clearly provide an asymptotically valid test statistic if in fact group 1 observations have variance σ_1^2 and group 2 observations have variance σ_2^2. Of course, if one is willing to make the assumption that the variance is constant over each of the sub-samples, various other procedures are available; see Jayatissa (1977), Weerahandi (1987), Watt (1979), Honda (1982), and Ohtani and Toyoda (1985), among others.

4. Finite-Sample Properties of the Tests

The tests suggested in the previous section are valid only asymptotically. If they are to be useful in practice, their known asymptotic distributions must provide reasonably good approximations to their unknown finite-sample distributions. In this section, I report the results of several Monte Carlo experiments designed to investigate whether this is so. For obvious reasons, attention is restricted to the case of linear regression models. Experiments were run for samples of sizes 50, 200, and 800, with n_1 equal to θn, θ being either 0.5 or 0.2, and with σ_1 variously equal to σ_2, four times σ_2, or one quarter of σ_2. In all experiments, there were four regressors including a constant term. The $\boldsymbol{X}$ matrix was initially chosen for a sample of size 50 and replicated as many times as necessary as the sample size was increased, so as to ensure that the matrix $n^{-1}\boldsymbol{X}^\top\boldsymbol{X}$ did not change. The regressors were a constant, the 90-day treasury bill rate for Canada, the quarterly percentage rate of change in real Canadian GNP, seasonally adjusted at annual rates, and the exchange rate between the Canadian and U.S. dollars, in Canadian dollars per U.S. dollar, all for the period 1971:3 to 1983:4.

Choosing the $\boldsymbol{X}$ matrix in this way makes it easy to see how the sample size affects the results. However, it may make the performance of the heteroskedasticity-robust (HR) tests appear to be unrealistically good in moderately large samples. As Chesher and Jewitt (1987) have shown, the values of the few smallest diagonal elements of $\boldsymbol{M_X}$ can have a very big impact on the finite-sample performance of HCCMEs. Replicating the $\boldsymbol{X}$ matrix as the sample size is increased ensures that all diagonal elements of $\boldsymbol{M_X}$ approach one at a rate proportional to $1/n$, so that, once n becomes large, the HR tests are bound to perform reasonably well. With real data sets, one would certainly expect the smallest elements of $\boldsymbol{M_X}$ to approach one as n tends to infinity, but possibly at a rate much slower than $1/n$, thus implying that the HR tests might perform less well for larger samples than these experiments suggest. In the experiments, the smallest diagonal elements of $\boldsymbol{M_X}$ were 0.7965 for $n = 50$, 0.9491 for $n = 200$, and 0.9873 for $n = 800$.

The four test statistics that were computed in the course of the experiments were the following:

1. The ordinary F statistic, expression (5), which is valid only under homoskedasticity. It will be denoted F.
2. The heteroskedasticity-robust test statistic (28), based on the artificial regression (27). It will be denoted HR_1.
3. A heteroskedasticity-robust test statistic like (25), in which $\ddot{\sigma}_t^2$ defined by (29) is used in place of $\tilde{u}_t^2$. This statistic, which will be denoted HR_2, is somewhat harder to compute than HR_1.
4. A test statistic with the form of (25), but where $\acute{\sigma}_t$ is either $\hat{\sigma}_1$ or $\hat{\sigma}_2$, where the latter were defined in (30). This statistic, which will be denoted 2V (for two variances) will be asymptotically valid under much less general assumptions than HR_1 and HR_2.

The results of the Monte Carlo experiments are presented in Tables 1 and 2. Table 1 contains results for 18 experiments where the null hypothesis that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$ was correct. The percentage of the time that each test rejected the null hypothesis at the nominal 1%, 5%, and 10% levels is shown in the table. These numbers should thus be very close to 1.0, 5.0, and 10.0 if the tests are behaving in finite samples as asymptotic theory says they should.

In the first group of experiments, the variance in the two subsamples was equal. The ordinary F test is thus completely valid, and, as we would expect, the rejection frequencies for the F test were indeed very close to what they should be. All the other tests performed reasonably well when $\sigma_1 = \sigma_2$. However, HR_1 and HR_2 tended to under-reject, especially for $\theta = 0.2$, when n_1 was one-quarter the size of n_2, while 2V tended to over-reject somewhat. The performance of all tests improved sharply with the sample size, and one could feel confident about using any of them for $n \geq 200$.

In the second group of experiments, σ_2 was four times as large as σ_1. The F test was therefore no longer valid, but it continued to perform quite well for $\theta = 0.5$. However, it rejected the null far too infrequently for $\theta = 0.2$. The two HR tests performed reasonably well for $\theta = 0.5$, but they also grossly under-rejected for $\theta = 0.2$. Even for $n = 800$, they tended to reject too infrequently in the latter case. The 2V test over-rejected quite severely for $n = 50$ and moderately for $n = 200$, but it performed very well for $n = 800$.

The third group of experiments was similar to the second, except that σ_1 was now four times as large as σ_2. This changed many results dramatically. The F test continued to perform surprisingly well for $\theta = 0.5$, but it rejected the null far too often for $\theta = 0.2$. The two HR tests generally performed well, although they over-rejected somewhat when $n = 50$. The 2V test continued to over-reject quite severely when $n = 50$ and moderately when $n = 200$.

From Table 1, two conclusions emerge. First, the two HR tests generally perform quite well, but they usually tend to under-reject. There is thus no reason to prefer

HR_2 to the simpler HR_1; the former simply under-rejects more severely in most cases. Nevertheless, there are evidently some cases where HR_1 can seriously over-reject, at least for small samples, so that routine use of this test as if it were an exact test is not justified. Secondly, the 2V test performs very well in medium and large samples, but it tends to over-reject in smaller ones. Its good performance in reasonably large samples makes sense, because it would be an exact test if $\hat{\sigma}_1$ and $\hat{\sigma}_2$ were replaced by σ_1 and σ_2. Provided that both n_1 and n_2 are reasonably large, $\hat{\sigma}_1$ and $\hat{\sigma}_2$ will provide good estimates of σ_1 and σ_2, and hence it is not surprising that the test performs well. Of course, in these circumstances, the Wald test examined by Watt (1979), Honda (1982), and Ohtani and Toyoda (1985), which also uses the estimates $\hat{\sigma}_1$ and $\hat{\sigma}_2$, might well perform even better.

Table 2 presents results for 18 experiments where the null hypothesis was false. The parameters were chosen so that for the case where $\sigma_1 = \sigma_2$ and $\theta = 0.5$, the F test would reject the null roughly half the time. The difference between $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ was made proportional to $n^{-1/2}$ so that there would be no tendency for the rejection frequencies to increase with the sample size. What should happen under this scheme as $n \to \infty$ is that all tests which are asymptotically equivalent will tend to the same random variable, and thus reject the null the same fraction of the time. The results in Table 2 largely speak for themselves. Once again, the 2V test performs well. It performs quite similarly to HR_1 and HR_2 in most cases for $n = 800$, but it generally rejects the null more frequently for smaller sample sizes.

The limited Monte Carlo experiments reported on here certainly do not provide a definitive study of heteroskedasticity-robust tests for structural change. For example, no attempt was made to study the effect of combining the ordinary F test with the 2V test by first doing a pretest of the hypothesis that $\sigma_1 = \sigma_2$; see Phillips and McCabe (1983) or Toyoda and Ohtani (1986). Such a strategy seems appealing, and it would presumably produce results somewhere between those for F and 2V, depending on the significance level of the pretest. There was also no attempt to quantify the size-power tradeoffs of the various tests, although how useful such an exercise is when size is not known in practice is unclear.

The most substantial omission is that the undoubtedly very complex relationships between test performance, the number of regressors, and the structure of the $\boldsymbol{X}$ matrix were not studied at all. To do so would be a major undertaking, because it seems unlikely that Monte Carlo evidence alone, without a strong theoretical framework based on work like that of Chesher and Jewitt (1987), would allow one to say anything interesting about those relationships. Nevertheless, a few fairly strong results do seem to emerge from the Monte Carlo experiments. These are:

1. There seems to be no reason to use HR_2 instead of the simpler HR_1.
2. Since HR_1 never seriously over-rejects at the 1% level, one should probably view an HR_1 statistic which is significant at the 1% level as providing quite strong evidence against the null hypothesis.

3. The 2V test performs very well in medium and large samples, although it over-rejects somewhat in small samples. It generally has more power than the HR tests, but they of course require much weaker distributional assumptions.

5. Conclusion

This paper has shown that it is remarkably easy to test for structural change in a fashion which is robust to heteroskedasticity of unknown form. The tests can also be modified so that they are robust only to a more structured form of heteroskedasticity in which the variance differs between the two subsamples, although since numerous other solutions to this simpler problem are available, this modification may be of limited interest. The new tests are asymptotically valid for both linear and nonlinear regression models. Monte Carlo evidence for the linear case suggests that, although the finite-sample performance of even the best tests is sometimes poor, the ordinary F test can be so misleading that it clearly makes no sense to ignore the possibility of heteroskedasticity when testing for structural change. At the very least, one should double-check the results of the F test by using one of the tests discussed in this paper.

References

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

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

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

Davidson, R., and J. G. MacKinnon (1985). "Heteroskedasticity-robust tests in regression directions," *Annales de l'INSEE*, 59/60, 183–218.

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

Engle, R. F. (l982a). "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1007.

Engle, R. F. (l982b). "A general approach to Lagrange multiplier model diagnostics," *Journal of Econometrics*, 20, 83–104.

Fisher, F. M. (1970). "Tests of equality between sets of coefficients in two linear regressions: An expository note," *Econometrica*, 38, 361–366.

Honda, Y. (1982). “On tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal,” *The Manchester School*, 49, 116–125.

Jayatissa, W. A. (1977). “Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal,” *Econometrica*, 45, 1291–1292.

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

Ohtani, K., and T. Toyoda (1985). “Small sample properties of tests of equality between sets of coefficients in two linear regressions under heteroskedasticity,” *International Economic Review*, 26, 37–44.

Phillips, G. D. A., and B. P. McCabe (1983). “The independence of tests for structural change in regression models,” *Economics Letters*, 12, 283–287.

Schmidt, P., and R. Sickles (1977). “Some further evidence on the use of the Chow test under heteroskedastieity,” *Econometrica*, 45, 1293–1298.

Toyoda, T. (1974). “Use of the Chow test under heteroskedasticity,” *Econometrica*, 42, 601–608.

Toyoda, T., and K. Ohtani (1986). “Testing equality between sets of coefficients after a preliminary test for equality of disturbance variances in two linear regressions,” *Journal of Econometrics*, 31, 67–80.

Watt, P. A. (1979). “Tests of equality between sets of coefficients in two linear regressions when disturbance variances are unequal: Some small sample properties,” *The Manchester School*, 47, 391–396.

Weerahandi, S. (1987). “Testing regression equality with unequal variances,” *Econometrica*, 55, 1211–1215.

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

Table 1. Rejection Frequencies when the Null Hypothesis is True

n	σ_1/σ_2	θ	Test	1% R.F.	5% R.F.	10% R.F.	θ	Test	1% R.F.	5% R.F.	10% R.F.
50	1/1	0.5	F	1.10	5.15	10.30	0.2	F	0.70	4.65	9.85
			HR_1	0.45	5.00	10.30		HR_1	0.00^{**}	0.70^{**}	5.10^{**}
			HR_2	0.25^{**}	3.00^{**}	8.05^{*}		HR_2	0.00^{**}	0.25^{**}	2.40^{**}
			2V	2.60^{**}	7.30^{**}	13.05^{**}		2V	6.10^{**}	12.90^{**}	19.00^{**}
200	1/1	0.5	F	1.05	4.85	10.25	0.2	F	1.35	5.90	9.95
			HR_1	0.55	4.50	9.65		HR_1	0.55	3.75	9.70
			HR_2	0.55	4.25	9.10		HR_2	0.55	3.55^{*}	8.90
			2V	1.30	5.30	11.00		2V	2.25^{**}	7.55^{**}	11.80^{*}
800	1/1	0.5	F	1.25	5.30	10.40	0.2	F	1.25	5.30	10.30
			HR_1	1.15	5.45	10.05		HR_1	0.95	5.00	9.85
			HR_2	1.10	5.45	9.90		HR_2	0.95	4.85	9.60
			2V	1.35	5.40	10.50		2V	1.20	5.60	10.70
50	1/4	0.5	F	2.65^{**}	7.10^{**}	11.80^{*}	0.2	F	0.00^{**}	0.10^{**}	0.15^{**}
			HR_1	0.60	4.45	11.70		HR_1	0.00^{**}	0.25^{**}	0.80^{**}
			HR_2	0.15^{**}	2.70^{**}	7.80^{**}		HR_2	0.00^{**}	0.10^{**}	0.55^{**}
			2V	2.50^{**}	7.50^{**}	13.80^{**}		2V	3.70^{**}	9.60^{**}	14.75^{**}
200	1/4	0.5	F	2.35^{**}	8.35^{**}	12.60^{**}	0.2	F	0.00^{**}	0.15^{**}	0.30^{**}
			HR_1	1.20	5.45	10.80		HR_1	0.25^{**}	1.90^{**}	5.65^{**}
			HR_2	0.95	5.05	10.40		HR_2	0.20^{**}	1.65^{**}	5.30^{**}
			2V	1.60^{*}	6.60^{*}	11.95^{*}		2V	1.65^{*}	6.05	11.55
800	1/4	0.5	F	1.80^{**}	5.35	10.10	0.2	F	0.00^{**}	0.05^{**}	0.15^{**}
			HR_1	1.00	4.85	10.10		HR_1	0.75	4.05	7.55^{**}
			HR_2	1.00	4.70	10.00		HR_2	0.75	3.95	7.35^{**}
			2V	1.20	5.25	9.90		2V	1.45	5.75	9.60
50	4/1	0.5	F	2.45^{**}	8.70^{**}	13.90^{**}	0.2	F	47.45^{**}	63.70^{**}	70.50^{**}
			HR_1	0.60	4.95	11.15		HR_1	0.90	7.40^{**}	16.05^{**}
			HR_2	0.20^{**}	3.05^{**}	7.75^{**}		HR_2	0.40^{*}	4.90	11.60
			2V	2.45^{**}	7.45^{**}	12.40^{**}		2V	9.15^{**}	16.15^{**}	22.20^{**}
200	4/1	0.5	F	2.70^{**}	8.05^{**}	12.95^{**}	0.2	F	38.70^{**}	56.70^{**}	64.90^{**}
			HR_1	1.10	4.90	10.30		HR_1	1.25	5.30	10.40
			HR_2	0.95	4.60	9.75		HR_2	0.90	4.50	9.70
			2V	1.50	5.90	11.35		2V	2.30^{**}	6.65^{**}	11.95^{*}
800	4/1	0.5	F	2.40^{**}	7.60	12.60^{**}	0.2	F	39.05^{**}	55.60^{**}	65.15^{**}
			HR_1	0.65	4.35	10.25		HR_1	1.00	5.75	10.60
			HR_2	0.65	4.35	10.05		HR_2	0.95	5.45	10.50
			2V	0.80	4.65	9.90		2V	0.95	5.55	10.85

All results are based on 2000 replications.

* and ** indicate that the rejection frequency in question differs significantly at the 0.01 and 0.001 level, respectively, from what it should be if the test statistic were distributed as $\chi^2(4)$ or $F(4, n-8)$.

Table 2. Rejection Frequencies when the Null Hypothesis is False

n	σ_1/σ_2	θ	Test	1% R.F.	5% R.F.	10% R.F.	θ	Test	1% R.F.	5% R.F.	10% R.F.
50	1/1	0.5	F	22.40	45.60	59.05	0.2	F	16.70	38.65	52.35
			HR_1	10.65	39.00	56.70		HR_1	0.25	8.00	23.15
			HR_2	5.65	29.40	47.90		HR_2	0.15	3.60	13.75
			2V	30.45	52.80	63.70		2V	31.45	47.85	58.00
200	1/1	0.5	F	26.00	50.20	62.20	0.2	F	20.55	41.50	54.60
			HR_1	23.65	48.50	61.05		HR_1	9.80	31.85	47.60
			HR_2	22.45	46.55	59.70		HR_2	8.90	30.25	45.55
			2V	28.05	51.30	63.10		2V	23.65	43.90	56.30
800	1/1	0.5	F	26.80	51.25	64.60	0.2	F	21.60	42.55	55.05
			HR_1	26.40	51.20	64.15		HR_1	18.05	40.20	53.50
			HR_2	25.70	51.05	63.95		HR_2	17.40	39.75	53.20
			2V	27.25	51.35	64.80		2V	22.00	42.10	56.00
50	1/4	0.5	F	23.50	43.90	56.55	0.2	F	0.95	5.45	12.10
			HR_1	11.90	39.40	57.85		HR_1	0.00	3.20	19.15
			HR_2	6.40	29.85	48.95		HR_2	0.00	1.10	9.45
			2V	36.55	57.35	69.50		2V	52.05	72.45	80.55
200	1/4	0.5	F	25.70	46.65	58.05	0.2	F	0.85	5.95	12.90
			HR_1	24.95	51.05	64.95		HR_1	15.50	45.60	63.15
			HR_2	23.70	49.20	63.65		HR_2	14.60	42.80	61.10
			2V	31.90	56.15	67.40		2V	47.75	71.50	80.10
800	1/4	0.5	F	24.80	44.90	57.45	0.2	F	0.95	6.50	14.20
			HR_1	28.80	52.25	65.80		HR_1	37.10	65.30	76.95
			HR_2	28.50	51.90	65.50		HR_2	36.85	64.70	76.65
			2V	30.95	53.40	66.30		2V	47.50	71.10	80.40
50	4/1	0.5	F	24.35	47.55	61.60	0.2	F	78.30	88.70	92.45
			HR_1	19.95	55.50	73.35		HR_1	2.35	19.30	36.90
			HR_2	11.95	44.10	65.10		HR_2	1.05	11.60	27.50
			2V	47.95	69.85	80.25		2V	26.25	40.05	48.95
200	4/1	0.5	F	26.20	51.70	63.25	0.2	F	76.15	86.30	89.80
			HR_1	39.50	66.05	77.55		HR_1	9.25	29.20	42.60
			HR_2	37.50	64.35	76.25		HR_2	8.15	27.50	41.25
			2V	46.55	69.75	79.00		2V	17.10	33.25	44.60
800	4/1	0.5	F	28.80	51.70	64.00	0.2	F	75.75	85.75	90.15
			HR_1	44.55	69.00	79.75		HR_1	13.45	30.55	43.60
			HR_2	44.20	68.70	79.50		HR_2	13.15	30.30	43.05
			2V	45.55	69.70	79.95		2V	14.85	32.50	43.25

All results are based on 2000 replications.