Some Specification Tests for the Linear Regression Model

Article *in* Sociological Methods & Research · November 1992

DOI: 10.1177/0049124192021002003

2 authors:

J Scott Long
Indiana University Bloomington

Pravin Trivedi
Indiana University Bloomington

SOME SPECIFICATION TESTS FOR THE LINEAR REGRESSION MODEL

J. Scott Long
Department of Sociology
Indiana University
Bloomington, IN 47401, U.S.A.

Pravin K. Trivedi
Department of Economics
Indiana University
Bloomington, IN 47401, U.S.A.

July 3, 1991
Revised: December 3, 1991
Revised: January 13, 1992

Acknowledgements: The authors would like to thank Ken Bollen, Ronald Schoenberg and Karl Schuessler for their comments on this paper.

Note: This paper in slightly modified form was published as: Long, J.S. and Pravin Trivedi, 1992, Some Specification Tests for the Linear Regression Model. *Sociological Methods and Research* 21:161-204. Reprinted in Bollen and Long, 1993.

SOME SPECIFICATION TESTS FOR THE LINEAR REGRESSION MODEL

If the assumptions of the regression model are correct, ordinary least squares (OLS) is an efficient and unbiased estimator of the model's parameters. If the assumptions of the model are incorrect, bias or inefficiency may result. Given the costs of applying OLS when the assumptions are violated, a great deal of recent work in econometrics has focused on the development of tests to detect violations of the assumptions (see Godfrey 1988 for a review of this literature). These tests are referred to collectively as *specification tests*. If these tests are passed, interpretation of the OLS estimates and application of standard statistical tests are justified. If one or more of the tests fails, modification of the model or alternative methods of estimation and testing may be required. Despite their potential, substantive researchers have been slow to apply specification tests. In part this is due to the absence of these tests from statistical packages. The development of simple computational formulas is overcoming this obstacle. A second problem is uncertainty regarding the small-sample performance of these tests and consequently of an appropriate strategy for applying specification tests.

The objective of this paper is to evaluate some important and computationally convenient specification tests for the normal regression model as applied to cross-sectional data and in the process to propose a strategy for applying specification tests. Since these tests achieve their optimal properties in large samples, their size and power in finite samples are of great interest and will be evaluated with Monte Carlo simulations. Our strategy for applying specification tests is to start with tests of the mean or first moment of the regression function. These include White's test of functional form and Ramsey's RESET test. If tests of the first moment are passed, tests of homoscedasticity, skewness and kurtosis of the errors (i.e., second, third and fourth moments) should be performed. Such tests are based on a decomposition of White's information matrix test.

A specification test for the full covariance structure model has recently been developed by Arminger and Schoenberg (1989). This test is designed to detect misspecifications caused by errors that are correlated with the independent variables, such as is the case with omitted variable bias. Standard test statistics used in covariance structure models, such as LR tests and modification indices, will not detect such misspecifications. While our paper focuses on specification tests dealing with the simpler case where there is no measurement model, in the conclusions we consider the implications of our results for the development of specification tests for the full model.

We begin with an overview of specification testing. This is followed by formal presentation of the normal regression model and of each test of specification. The last part of the paper presents Monte Carlo results.

OVERVIEW

In 1978 J.A. Hausman (1978) published the influential paper "Specification Tests in Econometrics" that presented a general method of testing for specification error. The idea behind this test is simple. Consider a model with a set of parameters θ. Let the null hypothesis H_0 be that the model is correctly specified, with the alternative hypothesis H_1 that the model is misspecified. Consider two estimators $\hat{\theta}$ and $\tilde{\theta}$. $\hat{\theta}$ is consistent, asymptotically normal, and efficient under H_0. $\tilde{\theta}$ is consistent under both H_0 and H_1, but is inefficient under H_0. If the model is correctly specified, then $\hat{\theta}$ and $\tilde{\theta}$ should have similar values. If the model is misspecified, then $\hat{\theta}$ and $\tilde{\theta}$ should differ. In effect, a Hausman test assesses whether the difference between $\hat{\theta}$ and $\tilde{\theta}$ is likely to have occurred by chance if the model were correctly specified.

H. White developed a pair of specification tests that are based on the Hausman principle (White 1980a, 1980b, 1981, 1982). The first test is called the *information matrix test* or simply the *IM test*. Rather than comparing two estimates of the structural coefficients θ, the IM test compares two estimates of the information matrix $I(\hat{\theta})$. With maximum likelihood estimation $I(\hat{\theta})$ can be estimated with either minus the expected value of the Hessian of the likelihood function or the expected value of the outer product of the gradients (OPG) of the likelihood function. If the model is correctly specified, then minus the expected value of the Hessian and the expected value of the OPG will be equal (cf. Judge, Hill, et al. 1985:179). The IM test is based on the idea that if two estimates of $I(\hat{\theta})$ are far apart, then misspecification is indicated. Significant values of this test indicate heteroscedasticity or non-normality of the errors. A second test is White's *test for functional misspecification*. This test compares the OLS estimates of the structural coefficients to weighted least squares (WLS) estimates. If the OLS and WLS estimates are far apart, then functional misspecification is indicated.

Ramsey's RESET (*re*gression *s*pecification *e*rror *t*est) test (Ramsey 1969) is similar in purpose to White's test for functional form.[1] The RESET test is based on the notion that if the functional form of the model is incorrect, then the correct specification might be approximated by the inclusion of powers of the variables in the original model. The original set of independent variables is augmented by powers of these variables. If the coefficients associated with the added variables are statistically significant, misspecification from sources such as incorrect functional form or the exclusion of relevant variables is suggested.

On the basis of the IM and functional form test, White (1981:423) suggests "a natural procedure for the validation of any regression model." First, the IM test is applied. If the test is passed, the estimates of the structural coefficients and the standard errors should be consistent and the usual methods of hypothesis testing are justified. If the model fails the IM test, then White's test for functional misspecification should be performed. If the functional form test fails, the hypothesis of correct functional form is rejected and OLS is inconsistent. If the test is passed, the functional form can be accepted as provisionally correct and OLS will be an unbiased estimator of the structural coefficients. Since the IM test failed, the standard OLS estimator of the variances and covariances of the estimates is biased. Consequently, standard statistical tests must be replaced by those based on a heteroscedasticity consistent estimator (discussed below).

Given that misspecification of the first moment is more fundamental and its consequences are more serious than misspecification of higher moments, this sequence of testing is reversed. If the functional form is incorrect, estimates from the incorrect model are at best approximations to those from the correct model and interpretation of the estimates should be avoided. Heuristically, if the conditional mean estimate is inconsistent, tests of the conditional variance cannot be interpreted. Since it is possible to pass the IM tests and fail a test of functional form, the implicit assumption that passing the IM test eliminates the need for testing functional misspecification should be reconsidered. A sounder strategy is to begin with a test of functional misspecification. If it is passed, the IM test should be performed. If it fails, the OLS estimator of the structural coefficients is still consistent, but the estimator of the covariance matrix is inconsistent. This inconsistency can be corrected by calculating a heteroscedasticity consistent covariance matrix.

In evaluating specification tests the related issues of orthogonality and robustness are of concern. Specification tests evaluate the null hypothesis of no misspecification against some alternative. Such tests may be *joint* or *uni-dimensional.* Joint tests simultaneously test for departures from the null in several directions. For example, the information matrix test is a joint test of the normality and homoscedasticity of the errors. A statistically significant joint test indicates the presence of either several or just one misspecification. Such a test is difficult to interpret since the specific cause of the violation is undetermined. Uni-dimensional tests evaluate departures in a single direction (e.g., homoscedasticity only). Ideally the test will have high power against departures from the null in one direction, but be insensitive or orthogonal to departures in another direction. If two tests have this property, the tests are asymptotical *orthogonal*. Given two variants of the same test, one would prefer the variant which has the orthogonality property. Note that tests may be asymptotically orthogonal but be correlated in finite samples. Tests that are strongly correlated provide little guidance regarding which assumption has been violated or what the appropriate remedy is. Orthogonality is related to the notion of robustness. A particular specification test may require that an auxiliary assumption is maintained. For example, a test of linearity may assume that the errors are homoscedastic. If a test is sensitive to the auxiliary assumptions, the test lacks robustness. A robustified version of a test attempts to prevent this, usually with the loss of some power if the robustification was not necessary.

A modern and attractive interpretation of a specification test is a conditional moment (CM) test. Every specification test may be thought of as a restriction on some conditional moment. For example a functional form test is a test of a restriction on the conditional mean function, a heteroscedasticity test is a test of a restriction on the conditional variance function, a test of symmetry of the distribution is a test of a restriction on the third moment of the distribution, and so forth. An excellent exposition of this view-point is in Pagan and Vella (1989). Every CM test can be interpreted as a test of $E[m(y,X,\Theta|X)]=0$, where Θ is an unknown parameter and *(y,X)* are data; *m(.)* is called the moment function. To implement the test Θ must be replaced by a consistent estimator. Specification tests have two aspects: the choice of the suitable moment function; and implementation of the test. Implementation requires the estimator $\hat{\Theta}$, the asymptotic variance of $m(y,X,\hat{\Theta})$, and a method of computing the test statistic *T*. *T* is ususally of the form $T=\frac{[m(y,X,\hat{\Theta})]^2}{var[m(y,X,\hat{\Theta})]}$. Competing variants of a CM test may therefore differ in terms of (a) the choice of the CM function; (b) the asymptotic variance of the CM function; and (c) the auxiliary regression used to compute *T*. These differences in the construction of tests will be referred to below and used to interpret some of the results of the Monte Carlo simulations.

THE REGRESSION MODEL

The multiple regression model may be written as

$$y=X\beta+u$$

where y is a $(n\times 1)$ vector of values for the dependent variable, β is a $(K\times 1)$ vector of unknown parameters, X is a $(n\times K)$ matrix of rank K containing nonstochastic independent variables, and u is a $(n\times 1)$ vector of stochastic disturbances. The first column of X may contain ones to allow for an intercept. For purposes of describing the specification tests it is useful consider the assumptions of the model in terms of conditional expectations of the error vector u.

Assumption 1. $E(u|X)=0$: the expected value of the error is zero for a given value of X. The simplest violation of this assumption occurs if the conditional expectation is a nonzero constant, which affects only estimates of the intercept. More serious violations occur when the value of the conditional expectation depends upon the values of X for a given observation. This can occur when the correct functional form is nonlinear and a linear model is estimated, or when an independent variable is incorrectly excluded from X. Such misspecifications are referred to as misspecifications of the first order since they concern the first moment of the error. In general, violations of the first order render the OLS estimator biased and inconsistent.

Assumption 2. $VAR(u|X)=E(u^2|X)=\sigma^2 I$: errors have a constant variance and are uncorrelated across observations. When the variance of the errors are uniform the errors are homoscedastistic; if the variances are not uniform, the errors are heteroscedastistic. In the presence of heteroscedasticity, the OLS estimator $\hat{\beta}_{OLS}$ is unbiased, but inefficient. The estimator of the variances and covariances of $\hat{\beta}_{OLS}$ is biased and the usual tests of significance are inappropriate. Since this assumption concerns expectations of the second moment of the errors, violations can be thought of as misspecification of the second order.

Assumption 3. $u\sim N$: the errors are normally distributed with a bounded variance. If the errors are *not* normal, the OLS estimator of β is still a best linear unbiased estimator and the usual tests of significance have asymptotic justification. However, the OLS estimator is no longer the maximum likelihood estimator and the small sample behavior of significance tests is uncertain. Since the normal distribution is symmetric and mesokurtic, the normality assumption implies that $E(u^3|X)=0$ and $E(u^4|X)=3(\sigma^2)^2$. Thus, violations of the normality assumption can be thought of as misspecifications of the third and fourth order conditional moments.

The specification tests detailed below are designed to detect violations of these assumptions.

WHITE'S TEST OF FUNCTIONAL MISSPECIFICATION

White's test for functional misspecification (White 1980b, White 1981) is a test of misspecification of the first order. Assume that the correct functional form is

$$y = g(X,\beta) + e \quad .$$

A test of functional form involves testing the hypothesis that $y = X\beta + u$ is correct against the alternative that $y = g(X,\beta) + e$ where $g(X,\beta) \neq X\beta$. To develop this test White considers the regression model $y = X\beta + u$ as an approximation to $y = g(X,\beta) + e$. β^* is defined as the vector of parameters that minimize the weighted distance between the expected values of y from the two model specifications $y = g(X,\beta) + e$ and $y = X\beta + u$. That is, β^* makes the linear approximation as close as possible to the nonlinear specification, where "close as possible" involves minimizing the squared, weighted distance between the predicted y's from the two specifications. Specifically, β^* minimizes the function

$$\int [\,g(x,\beta) - x\beta\,]^2 f(x)\, dx$$

where $f(x)$ is a weighting function. If $\hat{\beta}_{OLS}$ is the OLS estimator from the equation $y = X\beta + u$, $\hat{\beta}_{OLS}$ will approximate β^* in the sense that $g(X,\beta)$ is linearized using a Taylor expansion and all higher order terms are dropped. Under regularity conditions discussed by White (1980b:152-155)

$$\sqrt{n}\,(\hat{\beta}_{OLS} - \beta^*)$$

is distributed normally with mean zero and a covariance matrix $\hat{V}_{HC}$ defined as follows. Let

$$\hat{V}_{OLS} = n^{-1} \sum_{i=1}^{n} (y_i - X_i\hat{\beta}_{OLS})^2 X_i' X_i \quad ,$$

where X_i is a ($1 \times K$) vector of the values of X for observation *i*. Then,

$$\hat{V}_{HC} = (X'X/n)^{-1} \hat{V}_{OLS} (X'X/n)^{-1}$$

This estimator is called the heteroscedasticity consistent covariance matrix or simply the HC covariance matrix.

When $g(X,\beta) = X\beta$, $\beta^* = \beta_{OLS}$ regardless of the weighting function $f(x)$. When $g(X,\beta) \neq X\beta$ the least squares approximation β^* depends on the weighting function $f(x)$. White's test of functional misspecification involves obtaining an OLS estimate $\hat{\beta}_{OLS}$ and a weighted least squares estimate $\hat{\beta}_{WLS}$ based on an *arbitrary* weighting function $w(x)$. Functional misspecification is indicated by $\hat{\beta}_{OLS}$ and $\hat{\beta}_{WLS}$ being significantly apart. The test statistic is defined as (White 1980b:157)

$$\lambda_{FF} = n(\hat{\beta}_{OLS}-\hat{\beta}_{WLS})'\hat{\Psi}^{-1}(\hat{\beta}_{OLS}-\hat{\beta}_{WLS}) \qquad [1]$$

where

$$\hat{\Psi} = (X'X/n)^{-1}\hat{V}_{OLS}(X'X/n)^{-1} + (X'\Omega^{-1}X/n)^{-1}\hat{V}_{WLS}(X'\Omega^{-1}X/n)^{-1}$$
$$- (X'X/n)^{-1}\hat{U}(X'\Omega^{-1}X/n)^{-1} - (X'\Omega^{-1}X/n)^{-1}\hat{U}(X'X/n)^{-1}$$

with

$$\hat{V}_{WLS} = n^{-1}\sum_{i=1}^{n} w(X_i)^2 (y_i - X_i\hat{\beta}_{WLS})^2 X_i'X_i$$

and

$$\hat{U} = n^{-1}\sum_{i=1}^{n} w(X_i)(y_i - X_i\hat{\beta}_{OLS})(y_i - X_i\hat{\beta}_{WLS}) X_i'X_i \quad .$$

λ_{FF} is asymptotically distributed χ^2 with *K* degrees of freedom. If λ_{FF} is larger than the critical value at the given significance level, the null hypothesis that $g(X,\beta)=X\beta$ is rejected.

It is essential to realize that the size and power of this test depends on the choice of the weight function $w(x)$. Statistical theory provides little guidance for selecting optimal weights, a problem that will be shown below to greatly limit the application of this test.

THE RESET TEST

Ramsey's RESET test (Ramsey 1969) is another test for misspecification of the first order, including incorrect functional form and the exclusion of relevant variables. If

$$y = X\beta + u \qquad [2]$$

is the null model being considered, a possible alternative model can be written as

$$y = X\beta + Z\alpha + u^* \quad . \qquad [3]$$

If the correct variables X are in the model but the linear equation [2] is the incorrect functional form, a polynomial approximation to the correct form $y=g(X)+e$ can be made by defining the Z's in equation

[3] as powers of the X's. Ramsey proposed using powers of $\hat{y}=X\hat{\beta}$. Since $\hat{y}$ is a linear combination of the X's, this adds powers of the X's to the equation, albeit in a constrained manner. With this definition of the Z's equation [3] can be written as

$$y = X\beta + \sum_{j=1}^{P} \hat{y}^{\,j+1}\alpha_j + u^* \quad . \tag{4}$$

A RESET test of order P involves testing whether the addition of $\hat{y}^2$ through $\hat{y}^{P+1}$ to the right hand side of $y=X\beta+u$ results in a significant improvement in fit. The test statistic is the standard F-test of the hypothesis H_0: $\alpha=0$. To define this test, partition the covariance matrix of $\begin{pmatrix}\hat{\beta}\\ \hat{\alpha}\end{pmatrix}$ as

$$V = \begin{bmatrix} V_{\beta\beta} & V_{\beta\alpha} \\ V_{\alpha\beta} & V_{\alpha\alpha} \end{bmatrix} \quad .$$

Then let $V^{\alpha\alpha}$ be the partition in V^{-1} corresponding to $V_{\alpha\alpha}$. The F-test is defined as

$$\lambda_R = \frac{1}{P}\hat{\alpha}'\,\hat{V}^{\alpha\alpha}\,\hat{\alpha} \quad , \tag{5}$$

which is distributed as F with *P* and *n-K-P* degrees of freedom if the model is correctly specified.

The RESET test assumes that the errors are distributed normally with a constant variance. Thus, it tests the null hypothesis H_0: $u\sim N(0,\sigma^2 I)$ against the alternative H_1: $u\sim N(\mu,\sigma^2 I)$ where μ is not zero. When used as part of a sequence of specification tests, the sensitivity of the test to misspecifications of the second and higher orders is problematic since a decision to reject the hypothesis of correct functional form may be a consequence of heteroscedasticity or non-normality. This indicates a need for a version of the RESET test that is robust to these violations.

ROBUST RESET TESTS

Several authors (see for example Godfrey 1988:107) have suggested that the RESET test can be robustified for heteroscedasticity by replacing the standard OLS covariances matrix for $\hat{\alpha}$ with the robust or HC version. The resulting Wald test is defined as

$$\lambda_{RW} = \frac{1}{P}\hat{\alpha}'\,\hat{V}_{HC}(\hat{\alpha})^{-1}\hat{\alpha} \quad , \tag{6}$$

where $\hat{V}_{HC}(\hat{\alpha})$ contains that HC covariance matrix for the $\hat{\alpha}$'s from equation [4]. This test statistic is distributed as F with *P* and *n-K-P* degrees of freedom if the model is correctly specified (Milliken and

Graybill 1970). While this adjustment is appealing due to its computational simplicity, simulations presented below suggest that it has poor size and power properties. This poor performance is consistent with results of Chesher and Austin (19xx).

A second robust test was constructed as a Lagrange multiplier test, following Davidson and MacKinnon (1985). This test examines whether the residuals from the null model $y = X\beta + u$ are correlated with the excluded Z's after removing the linear influence of the X's from the Z's. To robustify the test for heteroscedasticity, the HC covariance matrix from the null model [2] is used as the weight matrix in the quadratic form of the test statistic. Specifically, the test statistic is defined as

$$\lambda_{RLM} = \hat{u}' Z_v \left(\frac{n-K}{n-K-P} Z_v' \hat{\Omega} Z_v \right)^{-1} Z_v' \hat{u} \tag{7}$$

where $(n\text{-}K)/(n\text{-}K\text{-}P)$ is a degrees of freedom correction. $\hat{\Omega}$ is a diagonal matrix with diagonal elements equal to the squares of the residuals from equation [2]. Z_v is a $(n \times P)$ matrix where the i^{th} column contains $\hat{v}_i$ from the regression

$$z_i = \delta_0 + X\delta + v_i \ , \quad i = 1, \ldots, P \ .$$

Thus, the i^{th} column of Z_v contains the Z's after removing the linear influence of the X's. λ_{RLM} tests whether the u's from equation [2] are correlated with that part of the Z's that is uncorrelated with the X's. By using a HC covariance matrix from equation [2], the resulting test should be robust to heteroscedasticity. The test statistics are asymptotically distributed as χ^2 with *P* degrees of freedom (Davidson and MacKinnon 1985).[2]

THE INFORMATION MATRIX TEST

If a test of functional form fails, there is no justification for proceeding to test for second and higher order misspecifications or to interpret the estimated parameters. If the functional form appears adequate, it is reasonable to proceed to test for violations of assumptions 2 and 3 involving higher moments of the error terms. White's information matrix test, originally designed as an omnibus misspecification test without specific alternatives, has this feature.

Under the assumptions of the regression model given above, the y's are distributed $N(X\beta, \sigma^2 I)$.

If the parameters β and σ^2 are stacked in the vector $\theta' = (\beta' \; \sigma^2)$, the resulting log-likelihood equation is

$$l(\theta) = -\frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta) \quad .$$

The maximum likelihood estimator $\tilde{\theta}$ is the value that maximizes the likelihood function. The score or gradient vector of the log-likelihood can be written as

$$d(\theta) = \frac{\partial l(\theta)}{\partial\theta} \quad .$$

If θ° is the true value of θ, the expected value of $d(\theta)$ will be zero when evaluated at θ°.

The asymptotic theory of maximum likelihood estimation shows that the covariance matrix of $\tilde{\theta}$ approaches $[I(\theta)/n]^{-1}$. If the regularity conditions for maximum likelihood estimation hold, the information matrix can be expressed in either of two ways. First,

$$I(\theta) = E\left[\left(\frac{\partial l(\theta)}{\partial\theta}\right)\left(\frac{\partial l(\theta)}{\partial\theta}\right)'\right] \quad .$$

This is the outer product of the gradient or OPG version. Second, the information matrix can be written as a function of the Hessian of the log-likelihood function. Let the Hessian be written as

$$D(\theta) = \frac{\partial^2 l(\theta)}{\partial\theta\,\partial\theta'} \quad .$$

The information matrix then equals

$$I(\theta) = -E[D(\theta)] \quad .$$

where the right hand size is the Hessian version of the information matrix. If the assumptions of the model are correct and certain regularity conditions hold, then

$$E\left[\left(\frac{\partial l(\theta)}{\partial\theta}\right)\left(\frac{\partial l(\theta)}{\partial\theta}\right)'\right] = -E\left[\frac{\partial^2 l(\theta)}{\partial\theta\,\partial\theta'}\right]$$

or in our earlier notation $E[d(\theta)\,d(\theta)'] = -E[D(\theta)]$.

The equality of the Hessian and OPG versions of the information matrix when the model is correctly specified is the basis for the IM test. If we define $l_i(\theta)$ as the value of the log-likelihood function for observation *i*, we may define

$$\Delta(\theta) = \sum_{i=1}^{n} \left[\frac{\partial^2 l_i(\theta)}{\partial\theta\,\partial\theta'} + \frac{\partial l_i(\theta)}{\partial\theta} \frac{\partial l_i(\theta)}{\partial\theta'} \right] . \qquad [8]$$

$\Delta(\theta)$ is a $(q \times q)$ symmetric matrix, where q is the number of β's plus one for σ^2. Since $E[\Delta(\theta)]$ should equal 0 if the model is correctly specified, White proposes $\Delta(\tilde{\theta})$ as a measure of misspecification. He demonstrates that $\sqrt{n}\Delta(\tilde{\theta})$ is normally distributed with mean zero when the model is correctly specified. A χ^2 statistic is constructed based on the lower triangle of $\Delta(\tilde{\theta})$ (since the matrix is symmetric). For simplicity, we will refer to this test statistic as $\Delta(\theta)$.

To understand how this test statistic depends on the assumptions of the regression model it is useful to consider the results of Hall (1987). Hall demonstrates that $\Delta(\theta)$ can be decomposed as

$$\Delta(\theta) = \delta_1 + \delta_2 + \delta_3 , \qquad [9]$$

where the δ's are asymptotically uncorrelated. δ_1 is a test of the second moments of the error distribution that closely resembles White's (1980a) earlier test of heteroscedasticity and the Breusch and Pagan's (1979) test for heteroscedasticity. δ_2 is a test of the skewness or third-moments of the error terms that resembles the $\sqrt{b_1}$ test of skewness. δ_3 is a test of kurtosis or the fourth moment that is similar to the b_2 test of kurtosis. This decomposition is accomplished by partitioning [8] according to the components β and σ^2 of θ. Thus, δ_1 is based on

$$E\left[\left(\frac{\partial l(\theta)}{\partial\beta}\right)\left(\frac{\partial l(\theta)}{\partial\beta}\right)'\right] = -E\left[\frac{\partial^2 l(\theta)}{\partial\beta\,\partial\beta'}\right]$$

which is a function of u^2. δ_2 is based on

$$E\left[\left(\frac{\partial l(\theta)}{\partial\beta}\right)\left(\frac{\partial l(\theta)}{\partial\sigma^2}\right)'\right] = -E\left[\frac{\partial^2 l(\theta)}{\partial\beta\,\partial\sigma^2}\right]$$

which is a function of u^3. And δ_3 is based on

$$E\left[\left(\frac{\partial l(\theta)}{\partial \sigma^2}\right)\left(\frac{\partial l(\theta)}{\partial \sigma^2}\right)'\right] = -E\left[\frac{\partial^2 l(\theta)}{\partial \sigma^2 \partial \sigma^2}\right]$$

which is a function of u^4.

The computation of the $\hat{\delta}$'s is relatively simple. First, estimate the regression

$$y = X\beta + u \quad . \tag{10}$$

Let $\hat{u}^2$, $\hat{u}^3$ and $\hat{u}^4$ contain the second, third and fourth powers of the residuals from [10]. Let $\hat{\sigma}^2$ be the maximum likelihood estimate

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 \quad .$$

Second, create Z to contain all of the unique products and bivariate cross-products of the X's. Specifically, Z contains $x_1 x_1$, $x_1 x_2$,..., $x_1 x_K$, $x_2 x_3$,..., $x_2 x_K$,..., $x_K x_K$. If x_1 is the constant *1*, then $x_1 x_1 = 1$, $x_1 x_2 = x_2$, and so on. Using these as a starting point, the components of the IM test can be easily computed. In the following expression the superscript *H* indicates that the formulas are based on Hall's construction of the test.

To compute $\hat{\delta}_1^H$ estimate the regression

$$\frac{\hat{u}^2}{\hat{\sigma}^2} = Z\tau + v \quad . \tag{11}$$

Then,

$$\hat{\delta}_1^H = \frac{SS_{EXP}}{2} \tag{12}$$

where SS_{EXP} is the explained sum of squares from the regression. To compute $\hat{\delta}_2^H$ estimate the regression

$$\hat{u}^3 = Z\tau + v \quad . \tag{13}$$

Let SS_{UNTOT} be the total uncentered sum of squares $\frac{1}{n}\sum_{i=1}^{n} y_i^2$ and SS_{RES} be the residual sum of squares

$\frac{1}{n}\sum_{i=1}^{n}\hat{v}_i^2$. Then

$$\hat{\delta}_2^H = \frac{SS_{UNTOT} - SS_{RES}}{6(\hat{\sigma}^2)^3} \quad . \qquad [14]$$

$\hat{\delta}_3^H$ is computed as

$$\hat{\delta}_3^H = \frac{n}{24}\left(\frac{\sum_{i=1}^{n}\frac{\hat{u}^4}{n}}{(\hat{\sigma}^2)^2} - 3\right)^2 \quad . \qquad [15]$$

The joint test is then computed as

$$\hat{\delta}^H = \hat{\delta}_1^H + \hat{\delta}_2^H + \hat{\delta}_3^H \quad . \qquad [16]$$

$\hat{\delta}^H$, $\hat{\delta}_1^H$, $\hat{\delta}_2^H$ and $\hat{\delta}_3^H$ are asymptotically distributed as χ^2 with $\frac{K(K+3)}{2}$, $\frac{K(K+1)}{2} - 1$, K and 1 degrees of freedom. Recall that K is the number of columns of X, possibly including a column of ones for the intercept.

A major advantage of Hall's formulation of the IM test is that it may suggest which aspects of the assumption $y \sim N(X\beta, \sigma^2 I)$ are being violated. To the extent that asymptotic property of the δ's being uncorrelated is approximated in small samples, $\hat{\delta}_1^H$ should be robust to violations of skewness and kurtosis, $\hat{\delta}_2^H$ should be robust to violations of heteroscedasticity and kurtosis, and $\hat{\delta}_3^H$ should be robust to violations of heteroscedasticity and skewness. While Hall's δ's are asymptotically uncorrelated, our simulations suggest that in small samples they are substantially correlated. Consequently, $\hat{\delta}_1^H$ can be statistically significant as a result of skewness, rather than hetero-scedasticity, and similarly for $\hat{\delta}_2^H$ and $\hat{\delta}_3^H$.

Cameron and Trivedi (1990a, 1990b) have demonstrated that the IM test can be viewed as a test with explicit alternatives. In the process they provide an *orthogonal decomposition* of the IM test. Let

X, Z and $\hat{\sigma}^2$ be defined as above. In addition let R^2_{UN} be the uncentered coefficient of determination defined as

$$R^2_{UN} = 1 - \frac{\sum_{i=1}^{n} \hat{v}_i^2}{\sum_{i=1}^{n} \hat{y}_i^2}$$

where the $\hat{v}_i$ are the residuals from a given regression. Note that most regression packages do *not* provide the uncentered coefficient of determination.

To estimate δ_1^{CT}, run the regression

$$(\hat{u}^2 - \hat{\sigma}^2) = Z\tau + v \quad . \tag{17}$$

Then,

$$\hat{\delta}_1^{CT} = nR^2_{UN} \quad . \tag{18}$$

To estimate δ_2^{CT}, run the regression

$$(\hat{u}^3 - 3\hat{\sigma}^2\hat{u}) = X\tau + v \quad . \tag{19}$$

Then,

$$\hat{\delta}_2^{CT} = nR^2_{UN} \quad . \tag{20}$$

To estimate δ_3^{CT}, run the regression

$$(\hat{u}^4 - 6\hat{\sigma}^2\hat{u}^2 - 3\hat{\sigma}^4) = \iota\tau + v \tag{21}$$

where ι is a vector of ones. Then,

$$\hat{\delta}_3^{CT} = nR^2_{UN} \quad . \tag{22}$$

The joint test is then computed as

$$\hat{\delta}^{CT} = \hat{\delta}_1^{CT} + \hat{\delta}_2^{CT} + \hat{\delta}_3^{CT} \quad . \tag{23}$$

As with the Hall version of the test, $\hat{\delta}^{CT}$, $\hat{\delta}_1^{CT}$, $\hat{\delta}_2^{CT}$ and $\hat{\delta}_3^{CT}$ are asymptotically distributed as χ^2 with $\frac{K(K+3)}{2}$, $\frac{K(K+1)}{2} - 1$, K and 1 degrees of freedom. The Monte Carlo experiments presented below indicate that their estimates for the δ's are substantially less correlated in small samples than are those based on Hall's formulas.

Components of Hall's version of the IM test may be compared to CT's version in terms the choice of the conditional moment function and the implementation of the test. The comparison is between Hall's expressions given by [11], [13], and [15] and CT's expressions given by [17], [18], and [21]. For testing heteroscedasticity, both use essentially the same conditional moment function based on $\hat{u}$ and $\hat{\sigma}^2$, so there should be little difference between the performance of the two on this account. For tests of symmetry and excess kutosis, there is a slight difference since CT use conditional moment functions which are orthogonal to each other, but Hall does not. The key difference between the two lies in the treatment of the variance term in the test. Hall's expressions for the tests use the variance of the conditional moment function that would be obtained if the null hypothesis were to hold. In this case the variance of the second, third and fourth moment restrictions would be indeed $2\hat{\sigma}^4$, $6\hat{\sigma}^6$ and $24\hat{\sigma}^8$, respectively. CT's versions of the corresponding tests does not use this assumption. Therefore, it is to be expected that Hall's variants would have better size properties, but will be less robust and powerful than CT's variants which do not impose the variance assumption on the test statistic. We may also expect that the use of the same (inconsistent) estimate of residual variance in the components of the IM test will cause these components to be mutually correlated.

A further point concerns the implementation of the test using nR^2 from an auxiliary regression. Observe that when the alternative hypothesis is valid, the auxiliary regressions given in [17], [19], and [21] have heteroscedastic residuals, which will induce a bias in the estimate of the residual variance and the uncentered R^2 from these auxiliary regressions, thereby affecting the properties of the tests. An alternative to the nR^2-based test is a heteroscedasticity consistent Wald test of the hypothesis that the slope parameters in the auxiliary regressions [17] and [19], and the intercept in [21], are all zero. As little is known about the merits of the heteroscedasticity consistent Wald test relative to the nR^2-based test, Monte Carlo simulations were undertaken to compare them. Though a more detailed discussion is given later, we anticipate the major conclusion of these experiments: the heteroscedasticity consistent Wald test has unsatisfactory size properties, especially with regard to tests of heteroscedasticity in small samples.

HETEROSCEDASTICITY CONSISTENT COVARIANCE MATRICES

If specification tests of the first-order are passed, but those for higher-order (and particularly heteroscedasticity) are not, standard tests of statistical significance are inappropriate. In such cases heteroscedasticity consistent tests should be considered. These tests are based on results of White (1980a) who extended earlier work of Eicker (1967) and Fuller (1975).

In the presence of heteroscedasticity the OLS estimator $\hat{\beta}_{OLS}$ is unbiased, but inefficient. The usual OLS estimator of $V(\hat{\beta}_{OLS})$ is biased and the usual tests of significance are inconsistent. If the exact form of heteroscedasticity is known to be $V(u) = \sigma^2\Omega$, the covariance matrix of $\hat{\beta}_{OLS}$ is

$$V(\hat{\beta}_{OLS}) = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1} \quad .$$

In general the specific values of $\sigma^2\Omega$ are not known. White's HC estimator of $V(\hat{\beta})$ involves substituting for Ω the matrix D that contains $\hat{u}_i^2$ on the diagonal and *0*'s on the off-diagonal. The resulting HC estimator is

$$\hat{V}_{HC}(\hat{\beta}_{OLS}) = \sigma^2 (X'X)^{-1} X'DX (X'X)^{-1} \quad . \qquad [24]$$

White recommends that if specification tests for the first-order are passed but test of higher order misspecification fail, tests based on the HC covariance matrix should be used.

MONTE CARLO STUDIES

Each of the tests discussed above is justified on the basis of asymptotic considerations. Since exact results for small-sample behavior are unavailable, Monte Carlo simulations provide a means for examining the small-sample behavior of these tests. These simulations involve a series of experiments in which: (1) a population is generated with a known data structure; (2) a large number of random samples are drawn from the population; (3) test statistics are computed for each sample; and (4) the test results across all samples are used to assess the size and power of the tests. Details of our simulations are now given.

<u>Data Structures</u>. Eight data structures were used. See Table 1 for a summary. For each data structure, one population was generated with an R^2 of approximately 0.4 and another with an R^2 of approximately 0.8. Three independent variables were used. X_1 was generated with a uniform distribution ranging from one to two, except for data structures 6 and 7 in which the range is zero to 100;

X_2 with a normal distribution with means zero and variance one; and X_3 with a χ^2 distribution with one degree of freedom with values ranging from zero to about fifteen.[3] The correlations among the X's are approximately equal to

	X_1	X_2	X_3
X_1	1.00	0.82	0.57
X_2	0.82	1.00	0.46
X_3	0.57	0.46	1.00

Structure 1 was generated to conform to the null hypothesis of no misspecification. The remaining seven structures represented various types of misspecification. Structure 2 has errors that are distributed as *t* with 5 degrees of freedom, which corresponds to a case with moderately thick tails. Structure 3 has errors that are distributed as χ^2 with 5 degrees of freedom, which corresponds to a moderate amount of skewness in the errors. Structure 4 incorporates heteroscedasticity of the form $\sigma_i^2 = \sigma^2 \sqrt{X_{1i}}$, where the value σ^2 is determined by the R^2 required. Given the uniform [1,2] distribution of X_1, this represents a moderate amount of heteroscedasticity. Structure 5 incorporates heteroscedasticity of the form $\sigma_i^2 = \sigma^2 \sqrt{X_{3i} + 0.25}$, where the value σ^2 is determined by the R^2 required and 0.25 is added to avoid extremely small variances in those cases were X_3 is close to zero. Given the uniform distribution of X_3, this represents a substantial amount of heteroscedasticity. Structure 6 was generated with the nonlinear equation $Y = 1 + \sqrt{X_1} + X_2 + X_3$. Given the range of X_1 from one to two, only a slight departure from linearity is involved. Structure 7 was generated with the nonlinear equation $Y = 1 + X_1 + X_2 + \sqrt{X_3}$. Given the range of X_3 from zero to fifteen, a substantial departure from linearity is involved. For each of these structures the model estimated was $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$. Structure 8 was generated in the same way as structure 1, that is, without specification errors. However, in this case the model estimated was $Y = \beta_0^* + \beta_2^* X_2 + \beta_3^* X_3$.

The Experiment. For each data structure the following experiment was run twice, once for an R^2 of .4 and once for an R^2 of .8. Each experiment consisted of:

(1) 150,000 observations with a known structure are generated and saved to disk. These observations are thought of as the *population*. A regression was run on the population to determine the exact values of the population parameters.

(2) From each population a random *sample* with replacement is drawn. This is repeated 1,000 times for each of the sample sizes: 25, 50, 100, 250, 500 and 1,000.

(3) Test statistics and parameter estimates are computed for each sample.

Summary of Experiments. Our simulations are used to assess the small sample size and power of each test. The tests that are evaluated are summarized in Table 2. If a data structure does not violate the assumptions being evaluated, the size of the test is assessed by comparing nominal significance level of the test to the empirical significance level. The empirical significance level is defined as the proportion of times that the correct null hypothesis was rejected over the 1,000 replications. Results are presented for nominal significance levels of 0.05, 0.10 and 0.20. In results not shown tests were evaluated at significance levels from 0.01 to 0.99 in increments of .01. If a data structure violates the assumptions evaluated by the test, the power of the tests is indicated by the proportion of times the null hypothesis is rejected at the 0.05, 0.10 and 0.20 levels.

Results are presented in the order corresponding to the strategy for testing that has been advocated. We begin with tests for specification of the first order, namely the White test and the RESET test. Tests for specification errors of the second and higher order are then considered with the IM test and its components.

Before presenting the results for the White test, we discuss how information on the power and size of a test is being plotted. Figure 1 is an example of the plots. The horizontal axis represents the size of the sample. Values of 25, 50, 100, 250, 500 and 1,000 are plotted. The vertical axis indicates the percentage of time that the null hypothesis of no misspecification was rejected. This is referred to as the *empirical alpha*. The range of the vertical axis will vary with the particular set of results being presented. For each test, results are given for three *nominal alphas*, with α=0.05 represented by ○, α=0.10 by □, and α=0.20 by Δ. Horizontal dotted lines are plotted at 0.05, 0.10 and 0.20. While the range of the vertical axis varies with the particular set of results being presented, the dotted lines are always at the same locations. If data were generated without any specification error and a test had perfect small sample size properties, the empirical alphas should fall directly on these dotted lines. If data were generated with specification errors that the test should detect, the higher the location of the empirical alphas, the greater the power of the test to detect the specification error.

SPECIFICATION TESTS OF THE FIRST ORDER

White's Test of Functional Form. White's test of functional form is designed to detect specification errors of the first order. To implement this test an "arbitrary" weighting function is chosen. Since theory provides little guidance in choosing a weighting functions, three sets of weights were considered based on suggestions in the literature. Version 1 of the test defines the weight as $w_i = (X_i\hat{\beta})^{-2} = \hat{y}_i^{-2}$; the weights are inversely proportional to the expected value of y for the null model (White 1980b:158-159). Version 2 defines the weight as follows. Let $\hat{u}_i^2$ be the squared residuals from equation [2]. Run the regression

$$w_i = \hat{u}_i^2 = \alpha_0 + Q_i\alpha + v_i \qquad [25]$$

where Q consists of X plus all of cross-products of the X's (e.g., X_1X_2, X_1X_3). Then the weights are defined as $\hat{w}_i$ from equation [25]. Those observations with the largest residuals (i.e., those that conform least well to the model) are given the greatest weight, which should make the test more powerful (White 1981). For both version 1 and 2, extremely small values were set equal to .00000001 and large values were set equal to 50. Version 3 defines the weights as 1 for large values of the dependent variable and .001 for small values of the dependent variable (White 1980b:159). Several variations on these weighting schemes were also tried.

Figure 1 demonstrates the practical problems involved in applying the White's test of functional form. Panel 1 presents the size properties of version 1 and 2 of the test. For version 1 the empirical alphas for the 0.05 significance level are close to 0.05, those for the 0.10 level begin at about .075 for n=25 and decrease to 0.05 for n=1,000, those for the 0.20 level begin at 0.10 and also decrease. Version 1 rejects H_0 slightly less frequently than it theoretically should. Version 2 of the test almost always rejects H_0, regardless of the significance level and thus would not be useful in practice. These results hold for other significance levels. For example, with a sample size of 25, version 1 rejects the correct null hypothesis only 28 percent of the time using α=0.90, whereas version 2 rejected the correct null 98 percent of the time using α=0.01. Panel 2 presents the power of the test to detect nonlinearity based on data structure 7. Version 1 of the test rejects the H_0 less often than it did with the correct specification, while version 2 rejects H_0 nearly every time.

The results presented in Figure 1 are typical of those obtained with other weighting functions and other data structures. For all weights considered either the size properties are unacceptable or the test had too little power to be useful. While it is possible that better weighting functions would improve the test, our experimentation suggests that these may be elusive. On the basis of our simulations we conclude: without operational procedures for choosing weights, White's test for functional misspecification is not practical for detecting misspecification in the regression model.

RESET Test. The RESET test is also a test for misspecification of the first order. Since the standard form of the RESET test assumes homoscedasticity, robust Wald and Lagrange multiplier versions of the test were constructed using White's HC covariance matrix. In all version we included the second through fourth powers of $\hat{y}$, so $P=3$.

Figure 2 shows the size properties of the RESET test using data structure 1 with $R^2=0.4$. Results are nearly identical for $R^2=0.8$. The size properties of the standard RESET test (plot A) are quite good. Size properties of the robust Wald RESET test (plot B) are very poor for small sample sizes, but improve steadily as the sample increases. However, even by n=1,000 the robust Wald test rejects the true H_0 substantially more often than the nominal alpha. The robust LM RESET test rejects the H_0 too seldom, although by n=100 the size properties approximate those of the standard RESET test. Given the substantially better size properties of the LM version compared to the Wald version, only the LM version is considered hereafter.

Figure 3 examines the properties of the RESET and LM RESET tests in the presence of heteroscedasticity. Panel 1 presents the results for a slight amount of heteroscedasticity (data structure 4) with $R^2=0.4$. Panel 2 presents the results for a moderate amount of heteroscedasticity (data structure 5) with $R^2=0.4$. Results are very similar for $R^2=0.8$. In Panel 1 with a slight amount of heteroscedasticity, the size properties of both versions of the RESET test remain similar to those obtained in the absence of heteroscedasticity, with the standard RESET test being superior to the LM version. Panel 2 with more extreme heteroscedasticity demonstrates the sensitivity of the standard RESET test to heteroscedasticity. As sample size increases, the percent rejected increases sharply. The robust LM version of the test is not affected by the heteroscedasticity and has reasonable size properties beginning with n=50.

Figure 4 shows the behavior of the RESET test in the presence of non-normal errors (data structures 2 and 3). The standard RESET test has similar properties as when no misspecification was present, although there is a tendency to reject H_0 too frequently when the errors have a χ^2 distribution. For t-distributed errors the empirical alphas for the LM test are too small at *n*=25, but improve as *n* increases. By *n*=1000, the empirical alphas are almost exactly equal to the nominal alphas. It is unclear whether the empirical alphas will stay at this level for higher *n*'s or continue to increase. For χ^2-errors the empirical alphas are too small, but become reasonably close by *n*=100.

Figures 5, 6 and 7 examine the power of the RESET test to detect misspecifications of the first order. Figure 5 is based on data structure 7 which contains some nonlinearity. Panel 1 is based on R^2=0.4. Neither test detects the nonlinearity, with empirical alphas being similar to those for data without misspecification. Panel 2 is based on R^2=0.8. For all sample sizes the standard RESET has larger empirical alphas than the LM RESET test, but by *n*=100 the differences are small. Figure 6 is based on data structure 6 which contains greater nonlinearity. Panel 1 is based on R^2=0.4. The tests behave similarly, with the ability of the test to detect nonlinearity increasing steadily with sample size. Panel 2 is based on R^2=0.8. For all sample sizes the standard RESET test has larger empirical alphas, but by *n*=100 the percentage difference is small. As before, the power to detect nonlinearity increases substantially with the larger R^2. Figure 7 considers the cases of an excluded variable. Both tests are extremely powerful by *n*=250, with the LM version again being somewhat less powerful for all samples. Unlike the cases for nonlinearity, the power of the test decreases with the larger R^2.

<u>Summary</u>. White's test of functional form is not useful for detecting misspecifications of the first order. The standard RESET and the robust LM RESET tests are very effective for detecting first-order specification error. In the absence of heteroscedasticity the standard RESET test has slightly better size properties and is slightly more powerful. However, it is affected by heteroscedasticity while the LM version of the test is not. Given the likelihood of heteroscedasticity in real world data and the relatively slight disadvantages of the LM test, the robust LM RESET test appears to be the preferred test. The robust Wald test has poor size properties for samples up to 1,000 and is not recommended.

SPECIFICATION TESTS OF HIGHER ORDER

<u>Size Properties</u>. White's IM test is designed to detect misspecification of the second and higher order. As shown by Hall, the test can be decomposed into components detecting heteroscedasticity, skewness and kurtosis. Figure 8 examines the size properties of the test for each component plus the combined test score for both the Hall method of computation and the method of Cameron and Trivedi

(CT).

For the overall test, both versions have empirical size close to the nominal size at the 0.05 level. One exception is the too small size of the CT version for *n*=25 and the too large size of the Hall version for *n*=100. At the 0.10 level both the empirical size of the overall tests are very close to 0.10 by *n*=100. Below that the sizes are too small. At the 0.20 level, both versions have sizes that are too small but improve with sample size, with the Hall version being less well behaved. In results not shown the size properties of the Hall version continue to deteriorate as the significance level increases.

The components of the IM test reveal the following. Consider first the Hall version. At the 0.20 level, the empirical size is substantially too small at *n*=25 with increasingly better size as *n* increases. By *n*=500 the empirical size is quite close to the nominal size. At the 0.10 level, the trend is similar although the overall behavior of the test is better. The size properties at the 0.05 level are extremely good for all sample sizes. Consider now the CT version of the test. At the 0.20 and 0.10 levels of the test the heteroscedasticity and skewness portions of the test have somewhat better size properties than the Hall version. At the 0.05 level the two versions are similar, except at *n*=25 where the Hall version behaves better. The size of the kurtosis portion of the CT version of the test becomes too large after *n*=25 but improves somewhat above 250.

<u>Power for Non-normal Errors</u>. Figures 9 and 10 examine the power of the IM test for violations of the assumption that the errors are normally distributed. Figure 9 considers errors that are *t*-distributed with 5 d.f. for the cases with R^2=0.4; results are similar for R^2=0.8. Having errors that are t as opposed to normal should introduce kurtosis but not heteroscedasticity or skewness. Herein lies the major difference between the Hall and CT versions of the test. In the Hall version the computation of a specific component of the test assumes that there are no violations of the other components being tested. For example, the computation of the skewness component assumes there is no heteroscedasticity or kurtosis. In the CT version the tests are robustified for violations of other assumptions. Thus in Panel 1 of Figure 9 we see that the Hall version of the tests rejects the hypothesis of no heteroscedasticity and of no skewness quite frequently, even though there is neither heteroscedasticity or skewness in the data. In Panel 2 we see that the CT version rejects these hypotheses at approximately the significance level of the test. In the kurtosis portion of the test, the Hall version is more powerful than the CT version, particularly when used at lower significance levels.

The greater sensitivity of the Hall versions of the tests to violations that are not being tested is reflected in the correlations among the test statistics. Tests of heteroscedasticity and skewness have a correlation of 0.71 for the Hall versions compared to 0.27 for the CT versions; tests of heteroscedasticity and kurtosis have a correlation of 0.49 for the Hall versions compared to -.02 for the CT versions; and tests of skewness and kurtosis have a correlation of 0.83 for the Hall versions compared to -.18 for the CT versions.

Figure 10 presents the results for errors from a χ^2 distribution with 5 d.f. Such errors introduce skewness and kurtosis, but not heteroscedasticity. The Hall version detects heteroscedasticity, while the CT version does not. On the other hand, the Hall version is more powerful at detecting violations of skewness and kurtosis. Once again, the Hall tests are more highly correlated than the CT versions.

Power for Heteroscedasticity. Figures 11 and 12 examine the ability of the IM test to detect heteroscedasticity. In Figure 11 the variance of the errors is proportional to the square root of x_1; in Figure 12 it is proportional to the square root of $x_3+0.25$ and is much stronger. Results are for R^2=0.4; results are very similar for R^2=0.8 and are not shown.

With slight heteroscedasticity both the Hall and CT versions of the test behave similarly, with the Hall version being just slightly more powerful. Given the modest amount of heteroscedasticity, the tests require 1,000 cases before the null hypothesis is rejected 75 percent of the time using a 0.05 significance level. The skewness and kurtosis portions of the test behave quite similarly to how they did in our tests of size.

In the case of a modest amount of heteroscedasticity, shown in Figure 12, several differences are seen. First, as would be expected, the power of the test increases. With a sample size of 250 the null of no misspecification is rejected nearly 100 percent of the time. The Hall version of the test remains somewhat more powerful, particular when a significance level of 0.05 is used. Finally, the more extreme heteroscedasticity results in tests of skewness and kurtosis that are more often significant. This is more often the case with the Hall version of the test.

Power for Nonlinearity and an Excluded Variable. The power of the IM test to detect violations of the assumption of linearity will depend upon the degree to which the presence of nonlinearity in the data generating process will result in errors that appear heteroscedastic, skewed or kurtotic in the linear analysis. With the nonlinear data structure 7, the hypothesis of no specification error was rejected less than 10 percent of the time using a significance level of 0.05 when n=1,000 regardless of the R^2. With data structure 6, which had more severe nonlinearity, the null hypothesis was still rejected less than 10 percent of the time when $R^2=0.4$. With an increase in the R^2 the tests had some, but still limited power. This is shown in figure 13. At the 0.20 level, H_0 was rejected less than 40 percent of the time with $n=1{,}000$. The power of the IM test to detect an excluded variable also depends on the degree to which the excluded variable generated heteroscedasticity, skewness or kurtosis in the model analyzed. Figure 14 compares the performance of the Hall and CT versions of the test in the cases where X_1 has been excluded from the equation and $R^2=0.4$. Results are similar for $R^2=0.8$. Overall we conclude that while the IM test is sometimes referred to as a test of functional misspecification, its power is limited. To the extent that functional misspecification results in residuals that are heteroscedastistic, skewed or kurtotic, the IM test will indicate functional misspecification. In cases where the assumptions regarding the higher order moments are not violated, the IM test will not detect functional misspecification.

THE HETEROSCEDASTICITY CONSISTENT WALD VERSION OF THE IM TEST

Finally, consider the properties of heteroscedasticity consistent Wald tests of heteroscedasticity, skewness and excess kurtosis based on CT's auxiliary regression equations [17], [19] and [21]. As previously explained, one tests the hypothesis that the slope parameters in regressions [17] and [19] and the intercept parameter in [21] are zero. Monte Carlo experiments were carried out to compare the size and power of the original variants of CT tests and the corresponding heterscedasticity consistent variants. To conserve space we only report a small selection of the output.

Table 3 gives the empirical size of heteroscedasticity consistent variants of the three IM tests at the 5% nominal level. In this case the rejection rate for all tests should be 5% since the data were generated without misspecification. However, the rejection rate for the heteroscedasticity test is 96.9% for $n=25$. While the rejection rate decreases as n increases, it is still as high as 33.8% for $n=500$. Thus the heteroscedasticity consistent variant of the test overrejects and is unsatisfactory, whereas the nR^2 variant of the same test shows far less size distortion. By contrast, once sample size exceeds 50, skewness and excess kurtosis tests show little size distortion.

Table 4 shows the rejection rates for the three tests when the data are generated by a model with moderate heteroscedasticity. Though the heteroscedasticity test appears to have high power, the reported rejection frequency overstates the power of the test becasue the appropriate concept here is size-corrected power. Given that the test over-reject when the null hypothesis is true, the nominal rejection frequency overstates the power of the test. For n larger than 50, the size distortion was not a serious problem for either the skewness or the heteroscedasticity tests. It is seen that size and power are roughly equal for these tests. We conclude that heteroscedasticity consistent tests produce very mixed results.

CONCLUSIONS

The tests considered in this paper provide a potentially useful set of tools for assessing regression models. To be useful for applied work a test must have reasonable size properties and enough power to detect departures from the null that may be expected in applied work. Our Monte Carlo experiments showed a tendency towards over-rejection in some tests. Tests with severe size distortion cannot be recommended unless they are corrected for this distortion. There are at least two ways of adjusting a test to achieve lower size distortion. First, using asymptotic expansions of the test statistic, one may calculate an adjustment factor for the test statistic. Alternatively, one might calculate adjusted critical values for tests at conventional significance levels. Both these approaches typically involve nonstandard calculations which applied researchers will find burdensome. Obviously, computationally simple tests with desirable size and power properties are preferable. Our Monte Carlo results suggest that specific variations of the RESET and IM tests have these characteristics.

The following strategy for testing should provide a useful assessment of one's model. First, the RESET test and the robust LM RESET test (rather than the heteroscedasticity robust Wald version of RESET) should be applied. If both tests pass, suggesting an adequate conditional mean specification, the IM test of higher order moment misspecification should be used. If the standard RESET test fails, but the robust version passes, correct functional form is indicated but higher order moment misspecification seems likely. Again the IM test should be considered. In applying the IM test, the decomposition suggested by Hall is useful for determining the specific violation that may cause the IM test to fail. While Hall's decomposition is useful, components of Hall's version of the test are not robust to departures from the null. For example, when the null is false, Hall's components of the test will be correlated which limits its ability to specific violations. The alternative variant of the IM test based on an orthogonal decomposition proved to be more robust when implemented along the lines of Cameron and Trivedi (1990b).

While specification tests have significant potential for the applied researcher, our results suggest the importance of carefully evaluating their properties with Monte Carlo methods. Additionally evaluation is needed to determine the behavior of these tests under additional violations of assumptions (e.g., the presence of outliers in the data) and with multiple violations of assumptions (e.g., nonlinearity and heteroscedasticity). Given the sometimes surprising behavior of specification tests for the simple case of regression, it is essential that as specification tests for the covariance structure model (Arminger and Schoenberg, 1989) are developed, that their properties are thoroughly examined with Monte Carlo methods before they are routinely applied.

ENDNOTES

1. White's test for functional form can be viewed as a variable augmentation test just as the RESET test. For details, see Godfrey (1988:152-5).

2. λ_{RLM} is awkward to compute since it requires combining the residuals $\hat{v}_i$ from P regressions with the residuals $\hat{u}$ from the original regression. There does not appear to be a single auxillary regression that will simplify the computations.

3. Random numbers were generated using the procedures RNDUS and RNDNS in GAUSS Version 2.1 (Aptech Systems 1991). The functions are based on a multiplicative congruential method for generating random numbers.

REFERENCES

Aptech Systems (1991) GAUSS Version 2.1. Kent, WA:Aptech Systems.

Arminger, G. and R.J. Schoenberg. 1989. "Pseudo Maximum Likelihood Estimation and a Test for Misspecification in Mean and Covariance Structure Models." Psychometrika 54:409-425.

Breusch, T.S., and A.R. Pagan. 1979. "A Simple Test for Heteroscedasticity and Random Coefficient Variation." Econometrica 47:1287-1294.

Cameron, A. Colin, and Pravin K. Trivedi. 1990a. Conditional Moment Tests and Orthogonal Polynomials. Working Paper.

Cameron, A. Colin, and Pravin K. Trivedi. 1990b. The Information Matrix Test and Its Implied Alternative Hypotheses. Working Paper.

Chesher, A., and G. Austin. 198x. "Finite Sample Behavior of Heteroskedasticity-robust Wald Statistics." Econometrica

Chesher, A. and G. Austin. 1991. "The Finite-sample Distributions of Heteroskedasticity-robust Wald Statistics." Journal of Econometrics 47:153-173

Davidson, R., and J. MacKinnon. 1985. "Heteroskedasticity-robust Tesets in Regression Directions." Annales de l'INSEE 59/60:183-218.

Eicker, F. 1967. "Limit Theorems for Regressions with Unequal and Dependent Errors." Pp. 59-82 in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1, Berkeley: University of California Press. 59-82.

Fuller, W.A. 1975. "Regression Analysis for Sample Survey." Sankhya, Series C 37:117-132.

Godfrey, L.G. 1988. Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches. New York: Cambridge University Press.

Hall, Alastair. 1987. "The Information Matrix Test for the Linear Model." Review of Economic Studies 54:257-263.

Hausman, J.A. 1978. "Specification Tests in Econometrics." Econometrica 46:1251-1272.

Judge, G. et al. 1985. The Theory and Practice of Econometrics. New York: Wiley.

Milliken, G.A., and F.A. Graybill. 1970. "Extensions of the General Linear Hypothesis Model." Journal of the American Statistical Association 65:797-807.

Pagan, A.R. and F. Vella. 1989. "Diagnostic Tests for Models Based on Individual Data." Journal of

Applied Econometrics, Supplement S4:S29-S60.

Ramsey, J.B. 1969. "Tests for Specification Errors in Classical Linear Least-squares Regression Analysis." Journal of the Royal Statistical Society, Series B 31:350-371.

White, Halbert. 1980a. "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." Econometrica 48:817-838.

White, Halbert. 1980b. "Using Least Squares to Approximate Unknown Regression Functions." International Economic Review 21:149-170.

White, Halbert. 1981. "Consequences and Detection of Misspecified Nonlinear Regression Models." Journal of the American Statistical Association 76:419-433.

White, Halbert. 1982. "Maximum Likelihood Estimation of Misspecified Models." Econometrica 50:1-24.

TABLE 1: Description of data structures for Monte Carlo Simulations.

Data Structure	Description	Systematic Component	Errors	Error Variance	R^2	Estimated Model
1	Correct Specification	$y=1+x_1+x_2+x_3$	Normal	$\sigma_i^2=\sigma^2$	0.4 0.8	$y=1+x_1+x_2+x_3$
2	t-distributed errors	$y=1+x_1+x_2+x_3$	t (5 df)	$\sigma_i^2=\sigma^2$	0.4 0.8	$y=1+x_1+x_2+x_3$
3	χ^2-distributed errors	$y=1+x_1+x_2+x_3$	χ^2 (5 df)	$\sigma_i^2=\sigma^2$	0.4 0.8	$y=1+x_1+x_2+x_3$
4	Heteroscedas-ticity on X_1	$y=1+x_1+x_2+x_3$	Normal	$\sigma_i^2=\sigma^2\sqrt{x_{1i}}$	0.4 0.8	$y=1+x_1+x_2+x_3$
5	Heteroscedas-ticity on X_3	$y=1+x_1+x_2+x_3$	Normal	$\sigma_i^2=\sigma^2\sqrt{x_{3i}+0.25}$	0.4 0.8	$y=1+x_1+x_2+x_3$
6	Nonlinearity in X_1	$y=1+\sqrt{x_1}+x_2+x_3$	Normal	$\sigma_i^2=\sigma^2$	0.4 0.8	$y=1+x_1+x_2+x_3$
7	Nonlinearity in X_3	$y=1+x_1+x_2+\sqrt{x_3}$	Normal	$\sigma_i^2=\sigma^2$	0.4 0.8	$y=1+x_1+x_2+x_3$
8	Excluding X_1	$y=1+x_1+x_2+x_3$	Normal	$\sigma_i^2=\sigma^2$	0.4 0.8	$y=1+x_2+x_3$

TABLE 2: Description of tests evaluated with Monte Carlo Simulations.

Test	Equation	Description
White's test for functional misspecification, Version 1	1	Weight function: $w_i = 1 / X_i\hat{\beta}$
White's test for functional misspecification, Version 2	1	Weight function: $w_i = \hat{e}_i^2$
RESET Test	5	Standard F-test of added variables.
Robust Wald RESET Test	6	Wald test using robust covariance matrix.
Robust LM RESET Test	7	LM test using robust covariance matrix.
IM Test - Total	15	Hall version.
IM Test - Heteroscedasticity Component	11	Hall version.
IM Test - Skewness Component	13	Hall version.
IM Test - Kurtosis Component	14	Hall version.
IM Test - Total	22	Cameron and Trivedi version.
IM Test - Heteroscedasticity Component	17	Cameron and Trivedi version.
IM Test - Skewness Component	19	Cameron and Trivedi version.
IM Test - Kurtosis Component	21	Cameron and Trivedi version.

TABLE 3: Empirical Size of the Heterscedasticity Consistent Wald Versions of the IM Test.

Sample Size	Empirical Size		
	Heteroscedasticity	Skewness	Excess Kurtosis
25	0.969	0.148	0.026
50	0.867	0.060	0.052
100	0.694	0.046	0.081
250	0.461	0.039	0.097
500	0.338	0.038	0.071

TABLE 4: Empirical Rejection Rates in the Presence of Heteroscedasticity for the Heterscedasticity Consistent Wald Versions of the IM Test.

Sample Size	Percent Rejected		
	Heteroscedasticity	Skewness	Excess Kurtosis
25	0.967	0.467	0.037
50	0.888	0.050	0.052
100	0.793	0.035	0.063
250	0.809	0.035	0.041
500	0.941	0.052	0.050

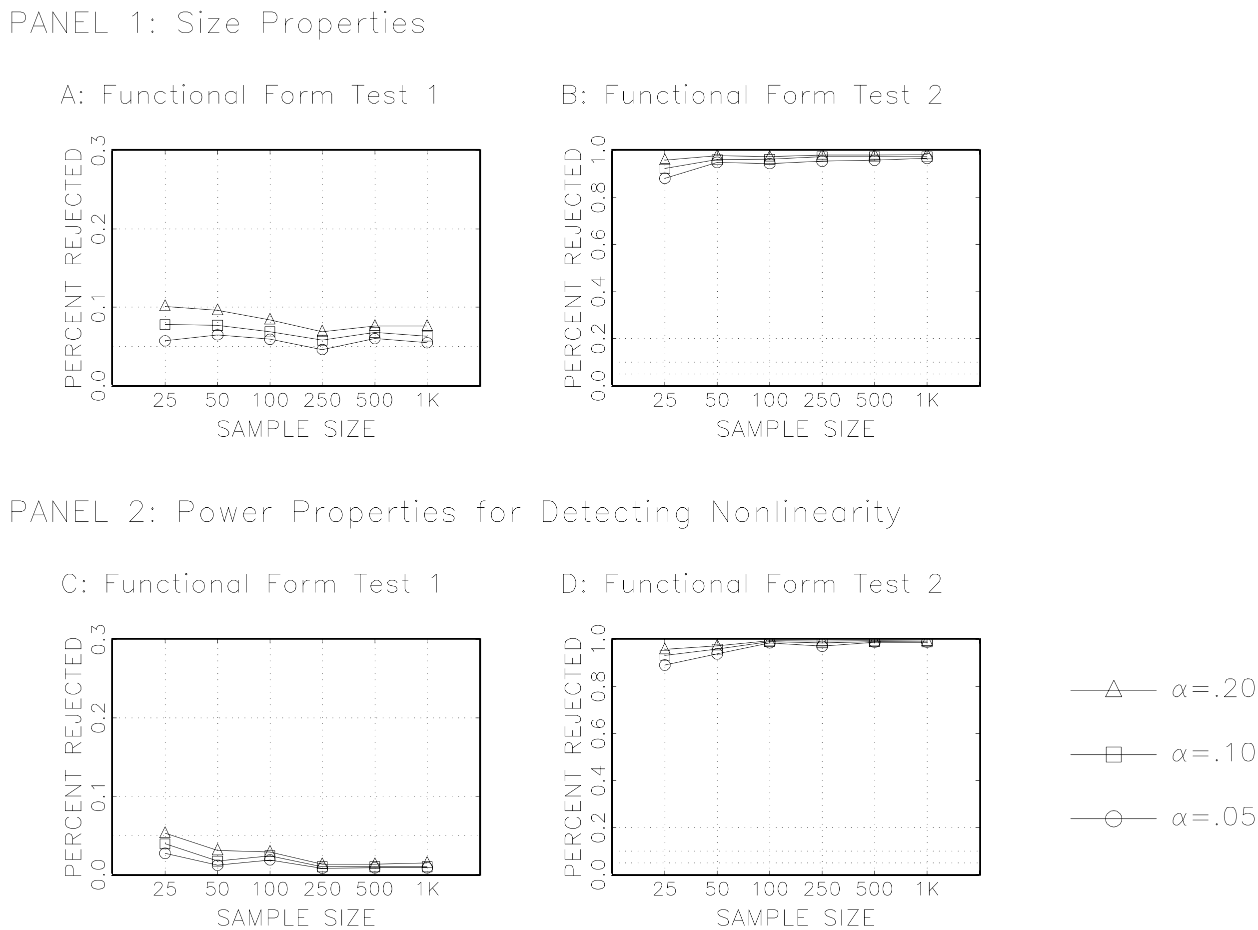

Figure 1: Size and power of White’s test for functional misspecification.

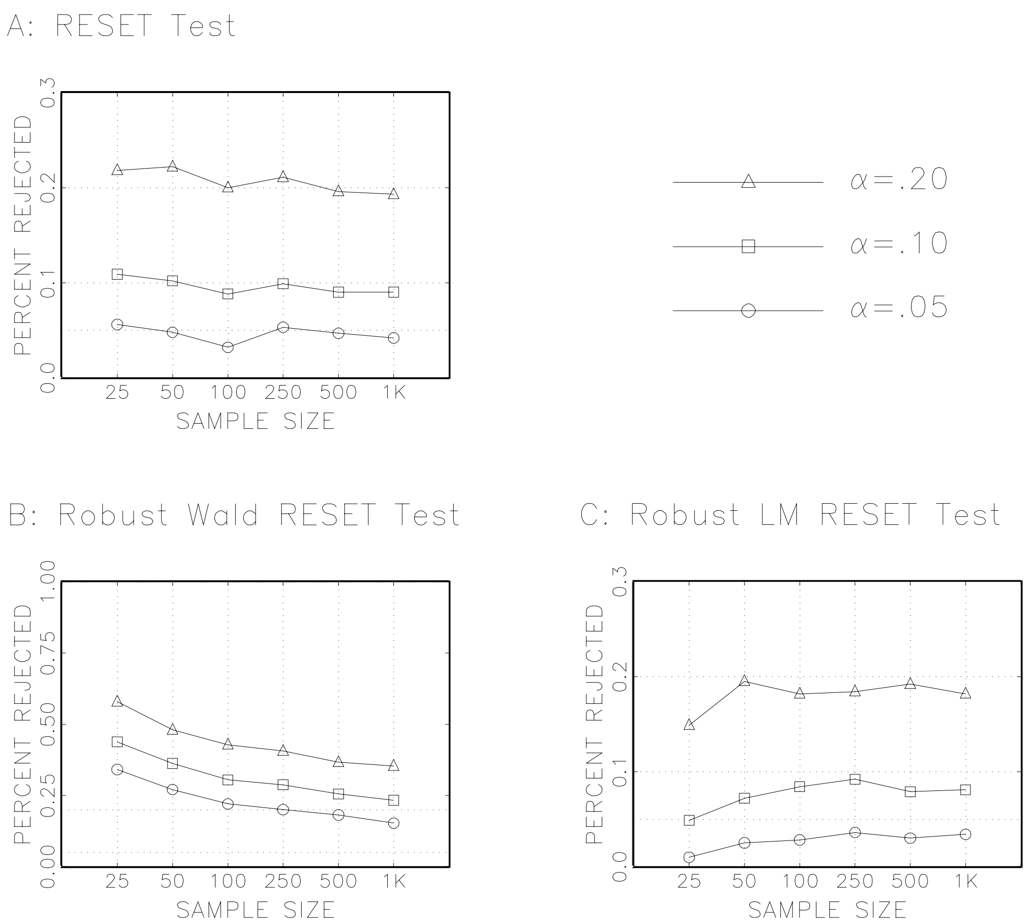

Figure 2: Size of RESET test.

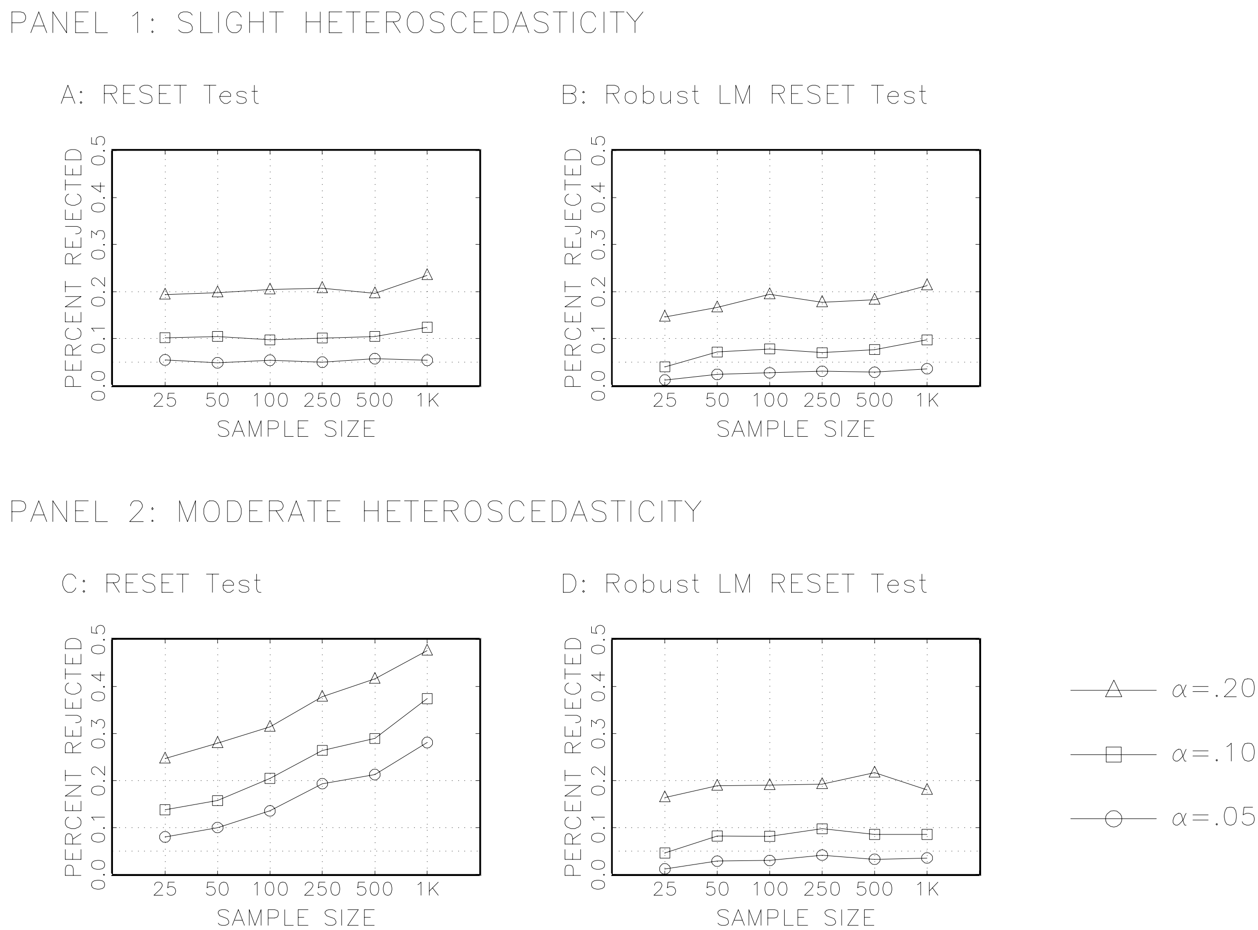

Figure 3: Properties of RESET and LM RESET tests in the presence of heteroscedasticity.

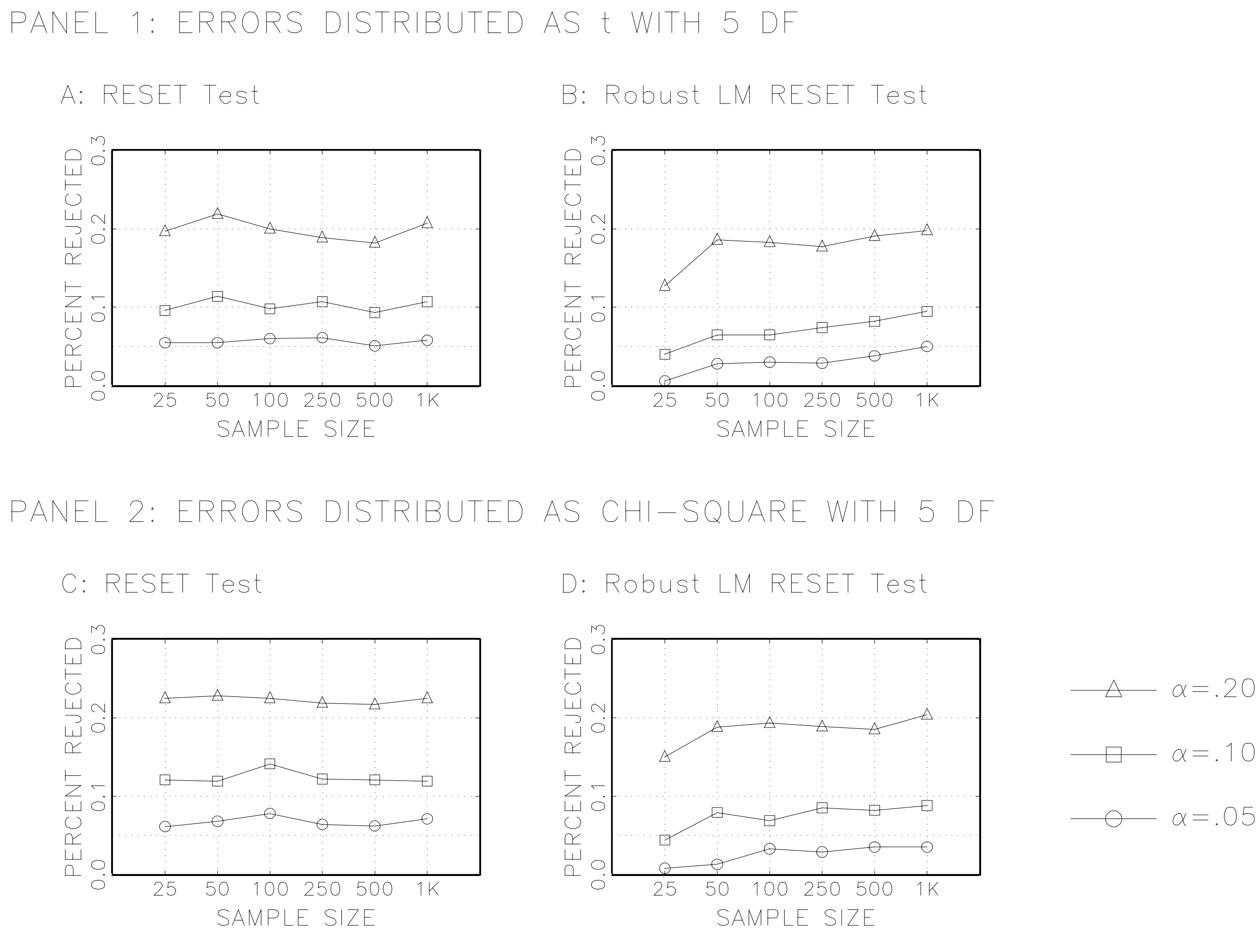

Figure 4: Properties of RESET and LM RESET tests in the presence of nonormal errors.

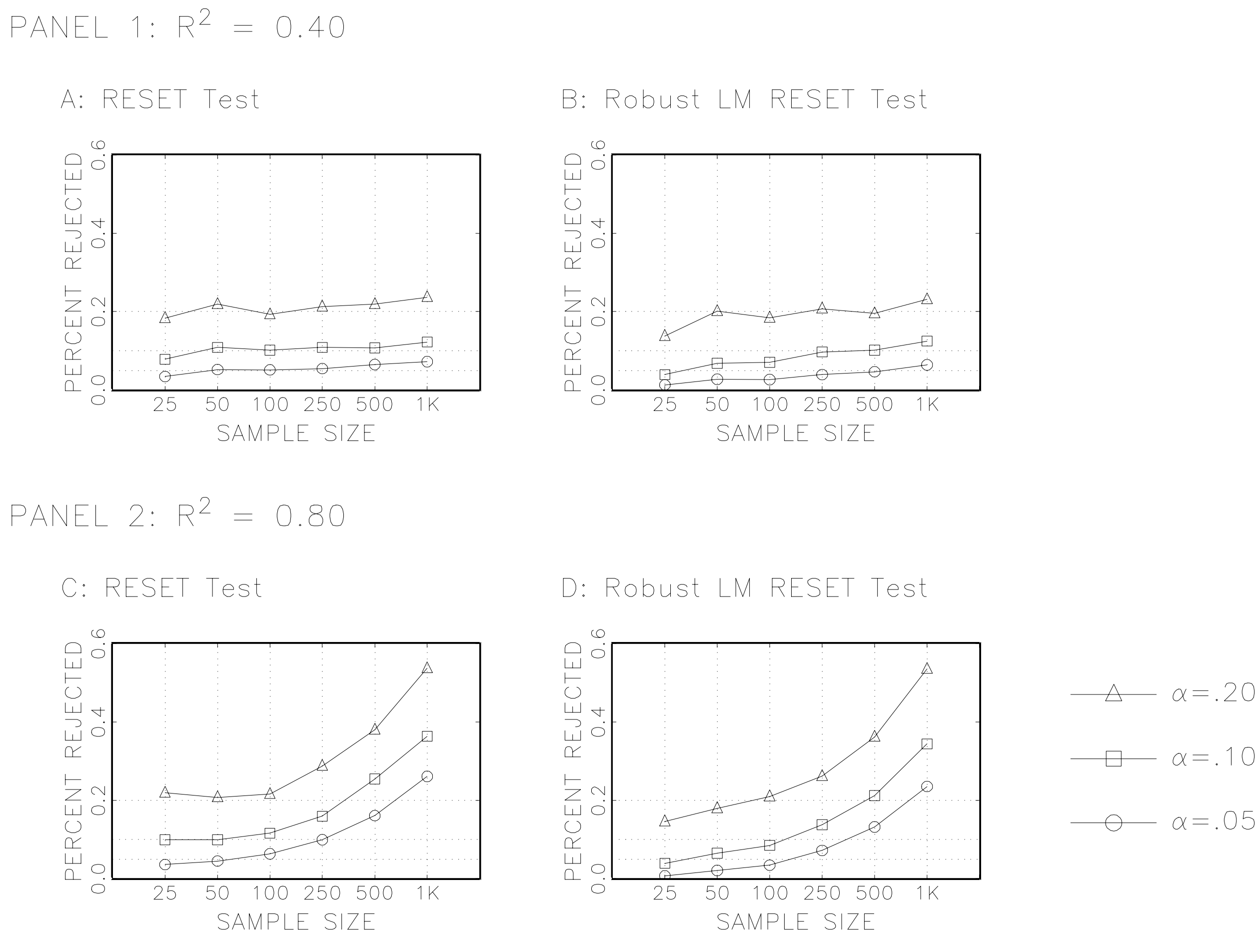

Figure 5: Power of RESET and LM RESET tests to detect slight nonlinearity.

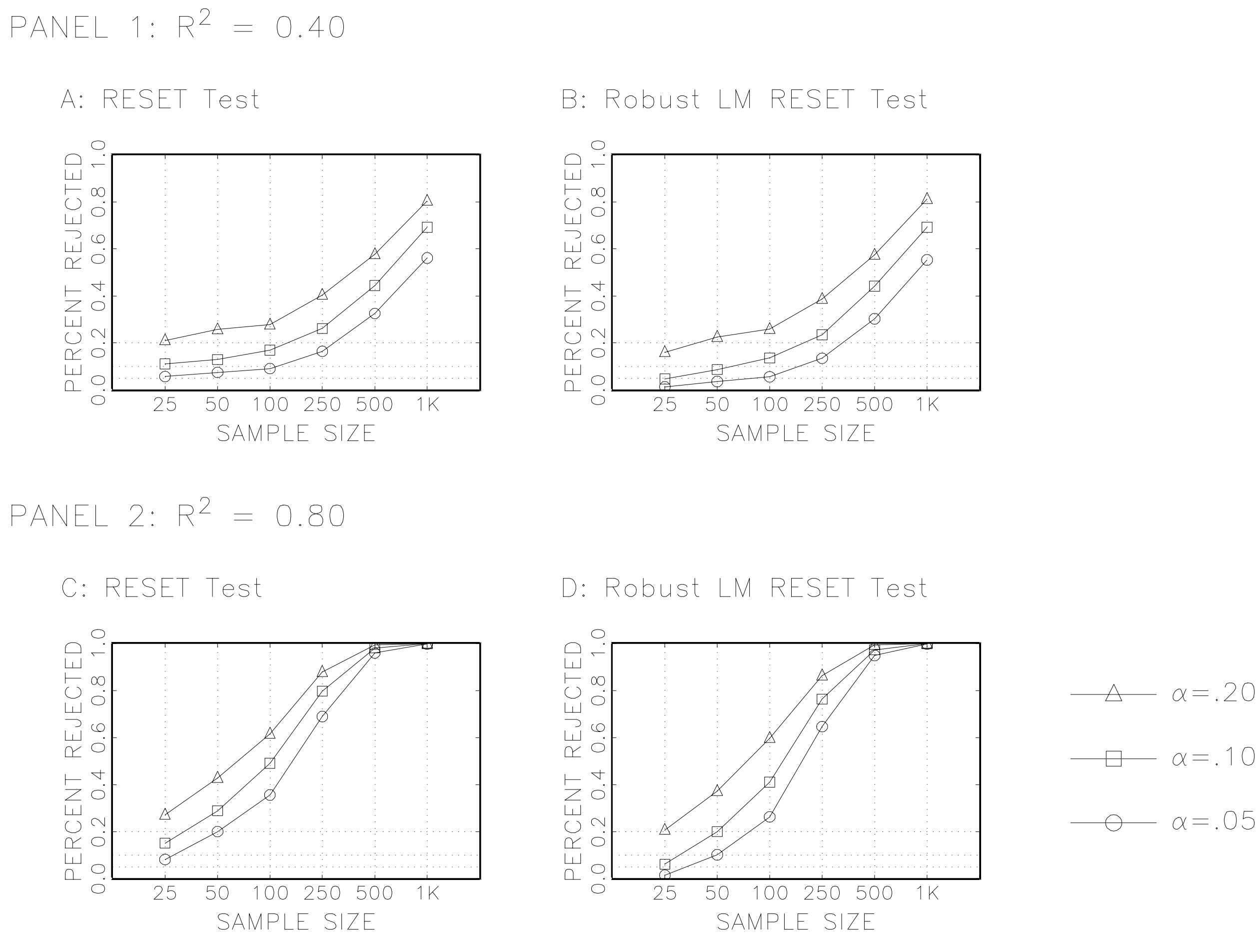

Figure 6: Power of RESET and LM RESET tests to detect moderate nonlinearity.

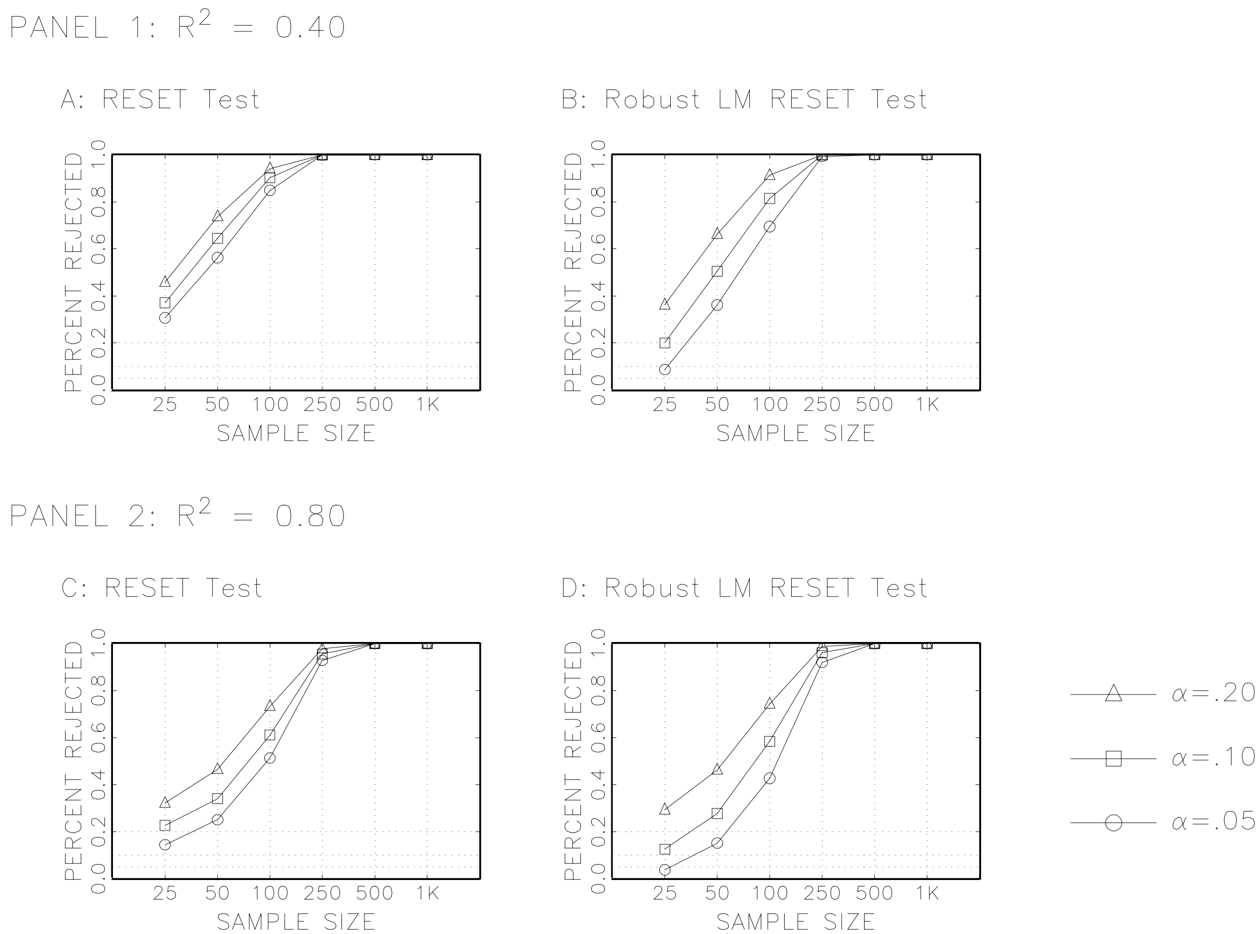

Figure 7: Power of RESET and LM RESET tests to detect an excluded variable.

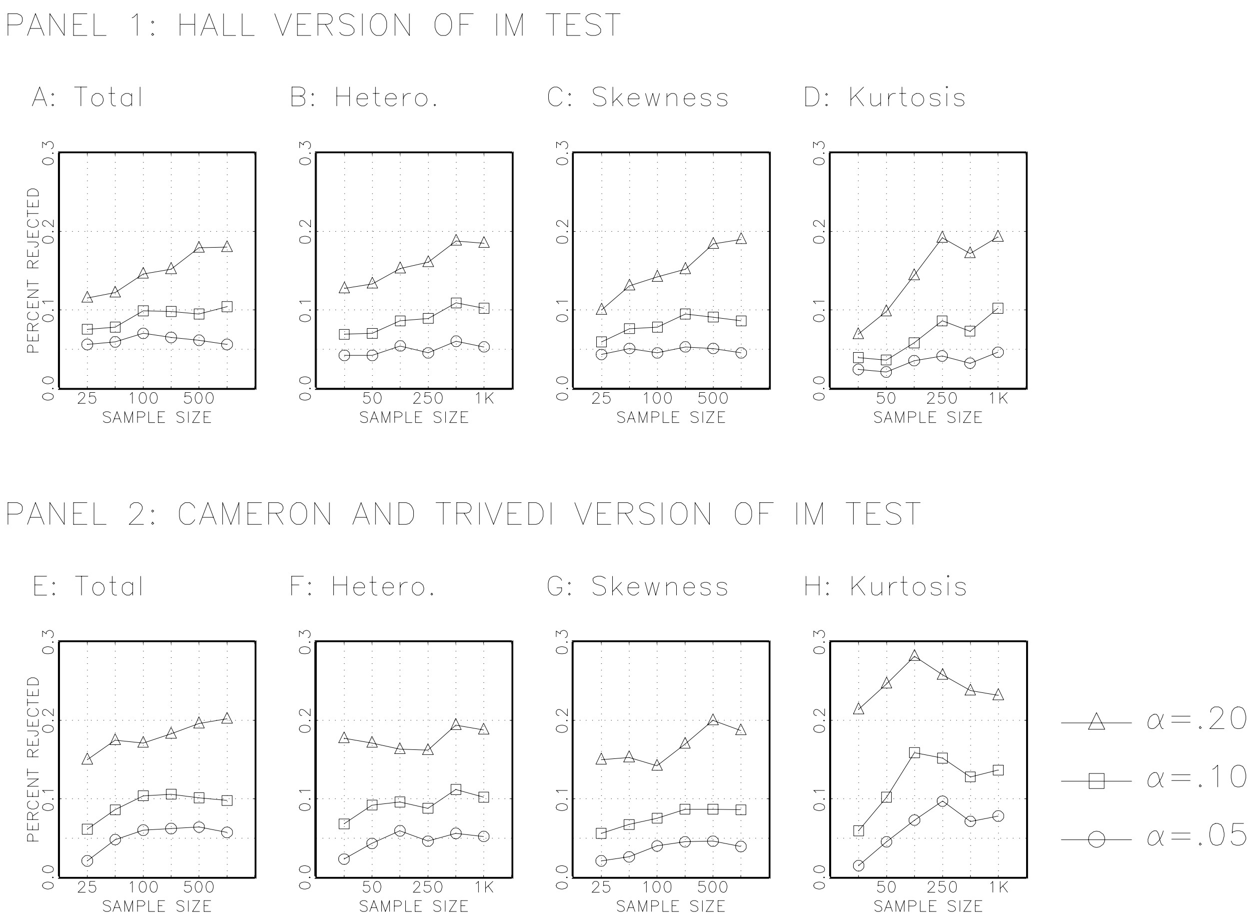

Figure 8: Size of information matrix test.

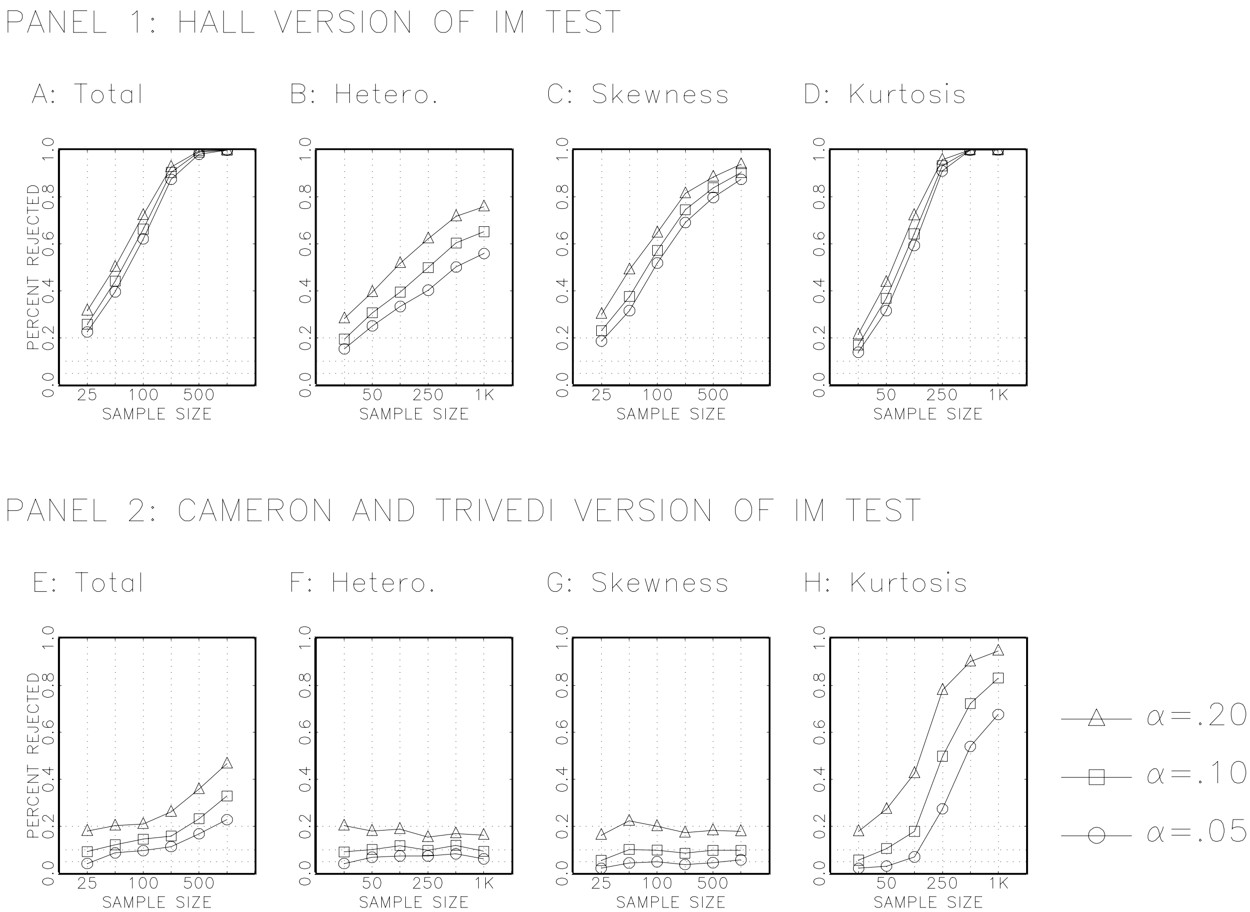

Figure 9: Power of information matrix test to detect errors distributed as t with 5 df.

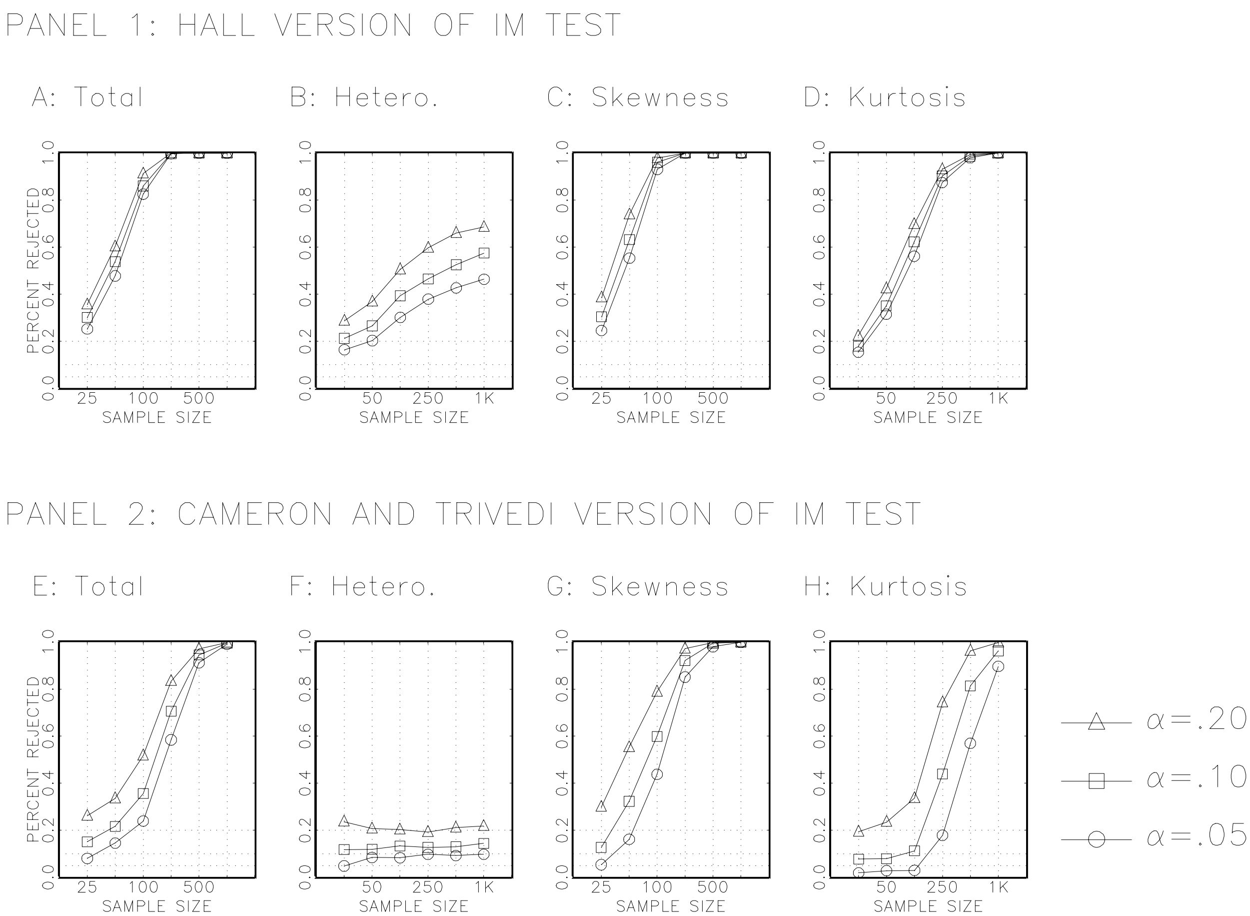

Figure 10: Power of information matrix test to detect errors distributed as chi-square with 5 df.

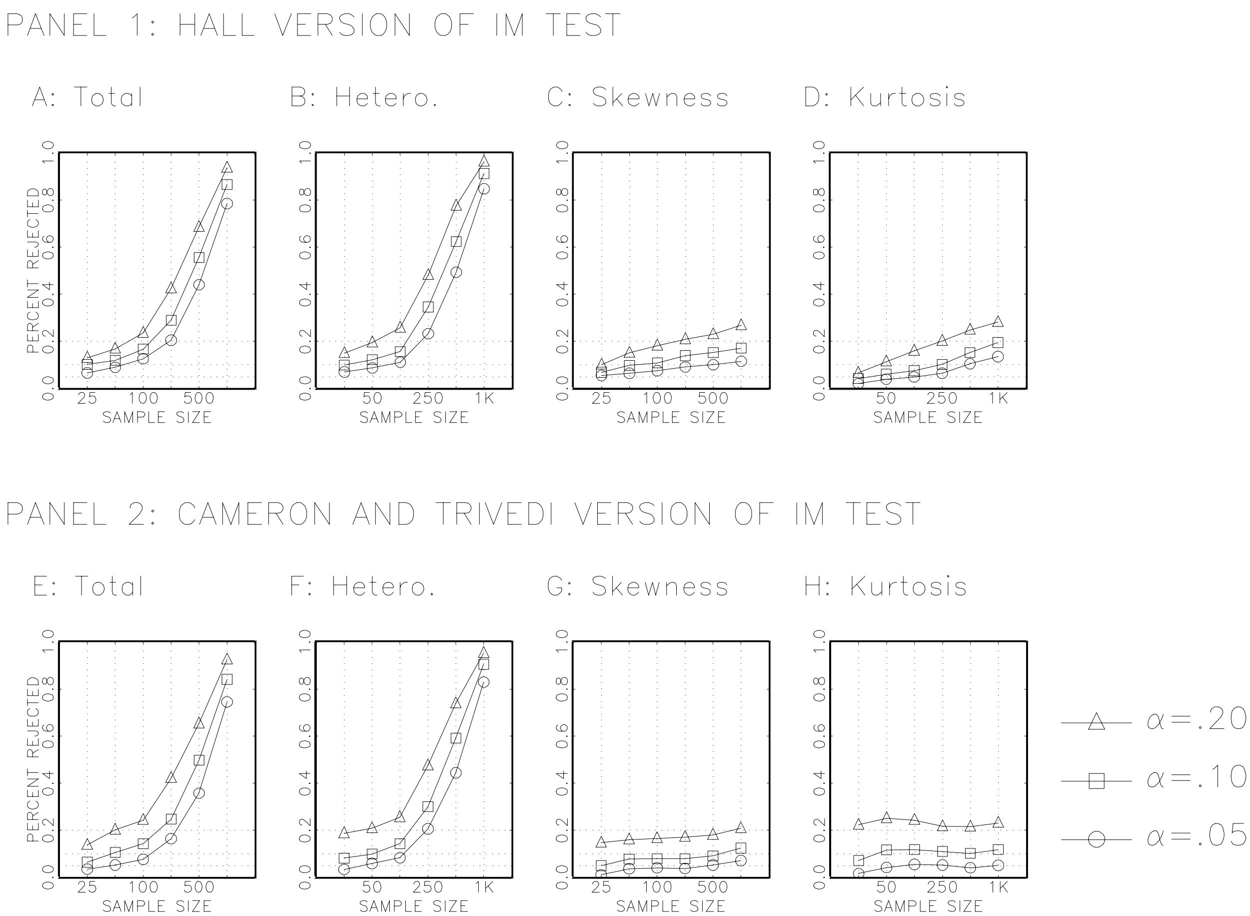

Figure 11: Power of information matrix test to detect slight heteroscedasticity.

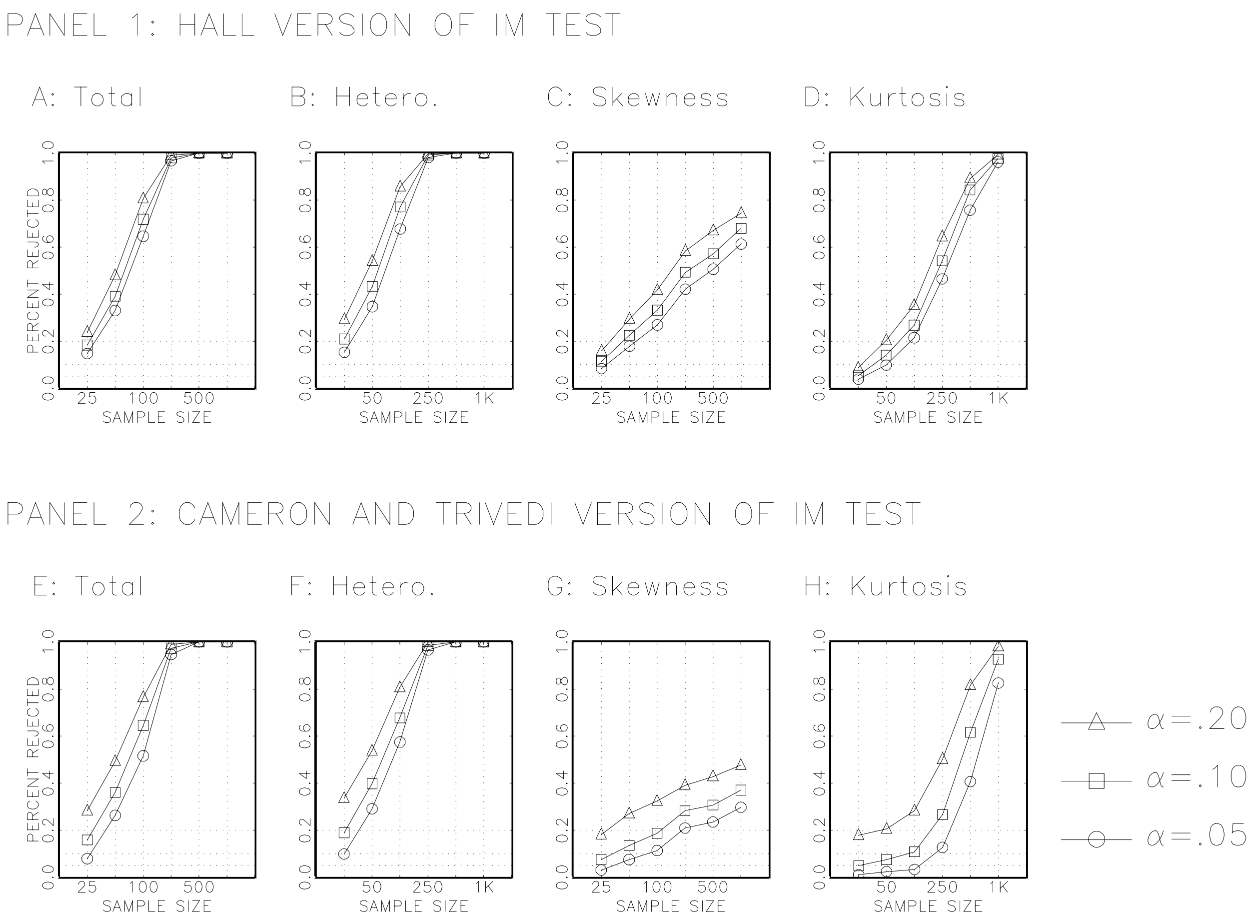

Figure 12: Power of information matrix test to detect moderate heteroscedasticity.

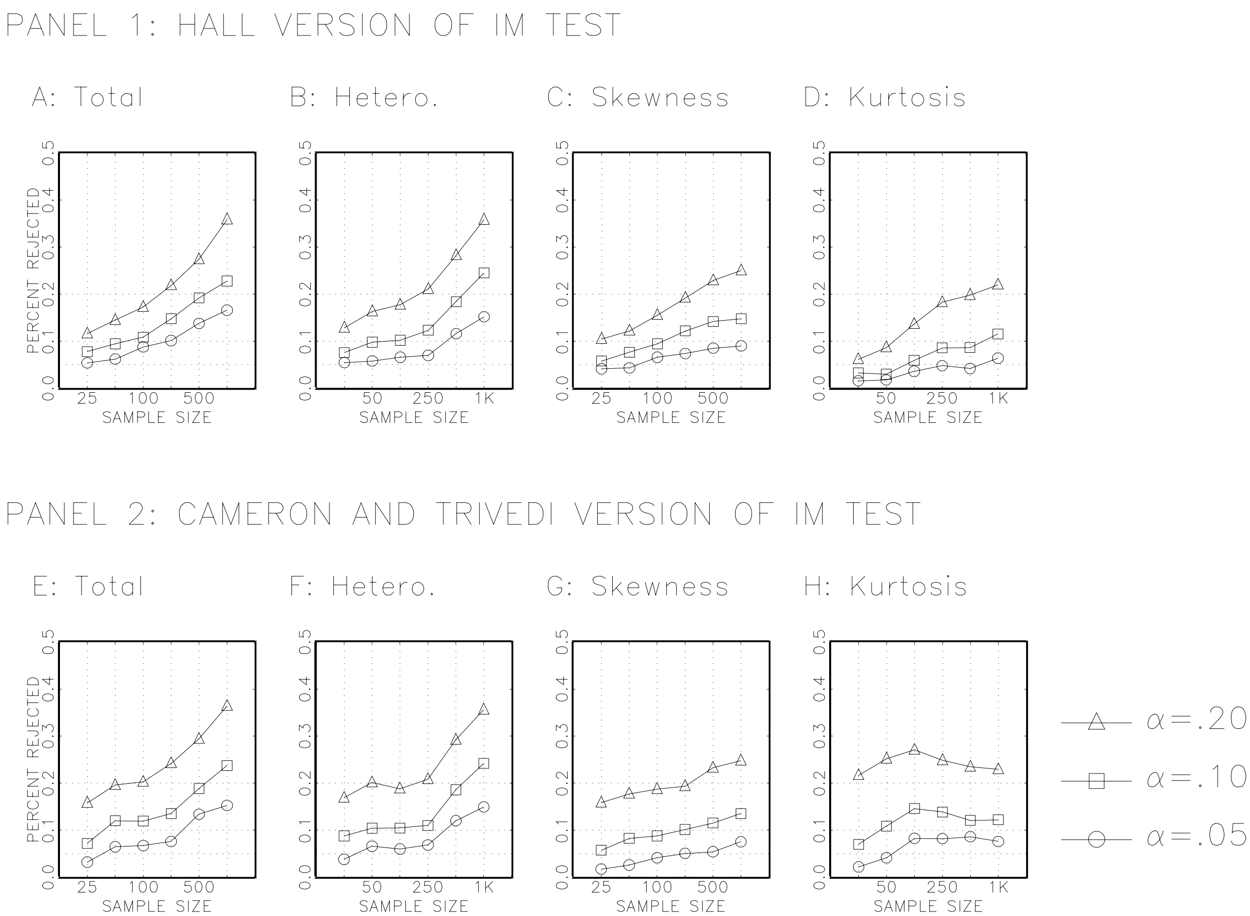

Figure 13: Properties of information matrix test in the presence of moderate nonlinearity.

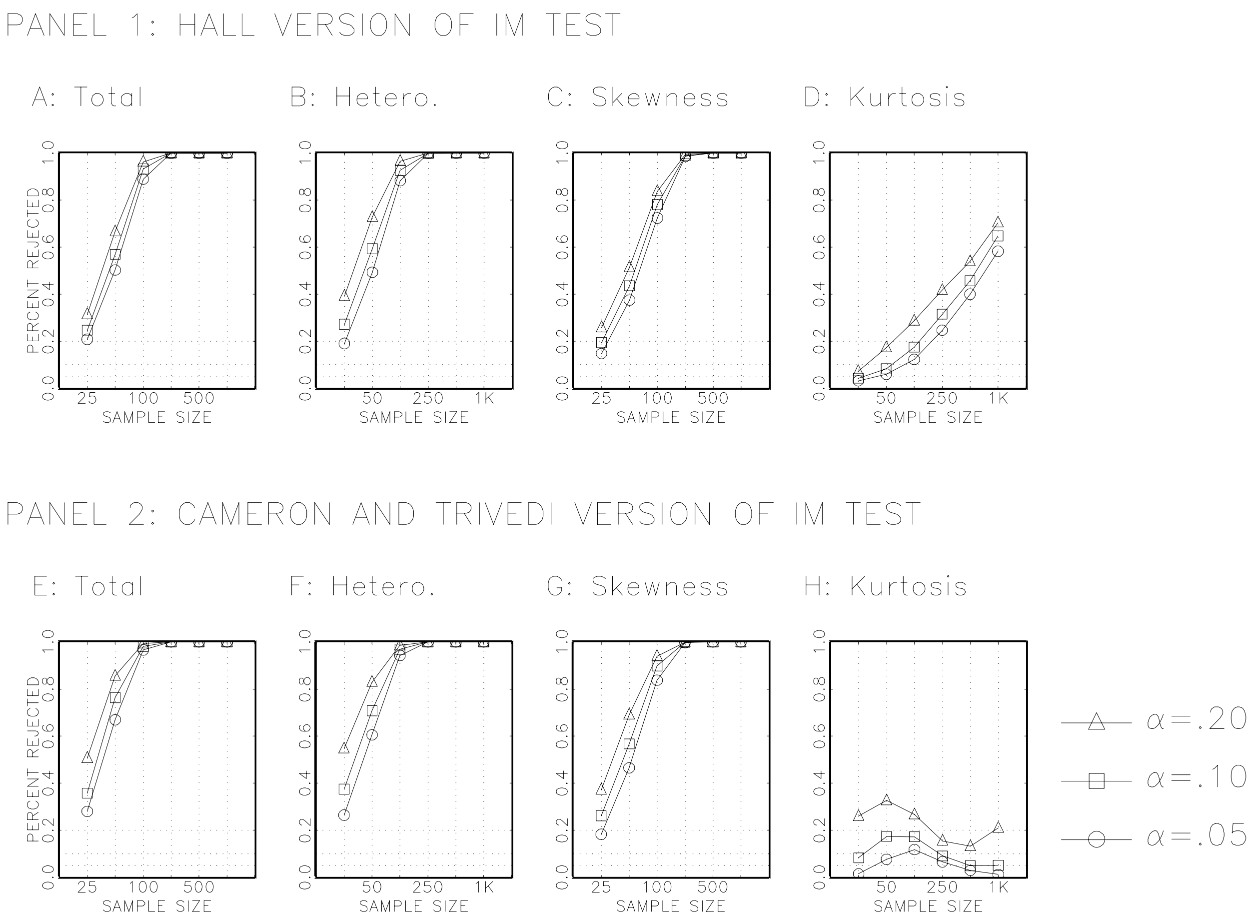

Figure 14: Properties of information matrix test in the presence of an excluded variable.