

A New Test of Linear Hypotheses in OLS Regression Under Heteroscedasticity of Unknown Form

Li Cai

University of North Carolina

Andrew F. Hayes

The Ohio State University

When the errors in an ordinary least squares (OLS) regression model are heteroscedastic, hypothesis tests involving the regression coefficients can have Type I error rates that are far from the nominal significance level. Asymptotically, this problem can be rectified with the use of a heteroscedasticity-consistent covariance matrix (HCCM) estimator. However, many HCCM estimators do not perform well when the sample size is small or when there exist points of high leverage in the design matrix. Prompted by a connection between MacKinnon and White's HC2 HCCM estimator and the heterogeneous-variance two-sample t statistic, the authors provide a new statistic for testing linear hypotheses in an OLS regression model that does not assume homoscedasticity. The authors report simulation results showing that their new test maintains better Type I error rate control than existing methods in both the presence and absence of heteroscedasticity.

Keywords: *heteroscedasticity; linear model; Satterthwaite approximation; Wald test*

The utility of the ordinary least squares (OLS) linear regression model in the educational and behavioral sciences is hard to dispute. Significance tests and confidence intervals involving regression coefficients are printed in popular statistical packages that researchers use almost daily. But the validity of the tests and the coverage probability of confidence intervals depend in part on the extent to which the model's assumptions are met.

Among the assumptions of the OLS regression model, homoscedasticity is a rather stringent one that is unlikely to hold in many applied settings. Researchers often encounter situations in which the variance of the dependent variable is related to the values of one or more explanatory variables, resulting in heteroscedasticity. In such a situation, a variance model based on the explanatory

Correspondence concerning this article should be addressed to Li Cai, University of North Carolina at Chapel Hill, Department of Psychology, CB #3270, Chapel Hill, NC 27599-3270; e-mail: cai@unc.edu.

variables can produce weights for the weighted least squares estimator (Cook & Weisberg, 1983). Weighted least squares, which is a special case of the generalized least squares estimator, is optimal if the covariance structure of the errors is known (see, e.g., Mardia, Kent, & Bibby, 1979). But usually, the error covariance structure is not known in advance. In that case, researchers can use estimated generalized least squares (Maddala, 1971) instead. Heteroscedasticity can also occur when one uses OLS to analyze binary or count data, but we do not consider this type of heteroscedasticity, because there are procedures, such as the generalized linear model (McCullagh & Nelder, 1989), that are much better suited to discrete outcomes than OLS. Another popular method of heteroscedasticity correction is to use a variance-stabilizing transformation on the dependent and/or independent variables such that homoscedasticity is satisfied for the transformed model (Carroll & Ruppert, 1988).

The methods of dealing with heteroscedasticity mentioned thus far assume that one can successfully model heteroscedasticity. Although we give all due credit to their usefulness, we sometimes are faced with heteroscedasticity of an unknown form. It takes time and care to build a model that fully explains the heteroscedasticity observed in the data. In practice, a researcher would usually propose several competing models of heteroscedasticity and choose the one that fits the data best, but it is possible that none of them fits well enough, and the researcher is left with no model to work with (see White, 1980, for a nice discussion of this issue).

However, one cannot simply let heteroscedasticity go uncorrected, even if its form is unknown. One of the most damaging consequences of heteroscedasticity is that the OLS estimator of the parameter covariance matrix (OLSCM), whose diagonal elements are used to estimate the standard errors of the regression coefficients, becomes biased and inconsistent. As a consequence, the t tests for individual coefficients are either too liberal or too conservative, depending on the form of heteroscedasticity. On the other hand, although the OLS estimator of regression coefficients is not optimal under heteroscedasticity, it is still a strongly consistent estimator. Therefore, a natural question to ask is whether we can find an alternative variance estimator that remains consistent under heteroscedasticity. If such a variance estimator is available, asymptotically correct inference can be achieved by retaining the OLS regression coefficients while replacing the OLS standard errors with heteroscedasticity-consistent standard errors.¹

More generally, an estimator of the parameter covariance matrix provides a way of conducting tests of linear hypotheses, of which the individual t test is a special case. If the OLSCM estimator is used in these tests, they suffer from the same problem under heteroscedasticity as do the individual t tests. The problem can be rectified in large samples with the use of a heteroscedasticity-consistent covariance matrix (HCCM) estimator. Although HCCM estimators can ensure asymptotically correct test size, most of them do not perform well in small samples, especially when there exist high-leverage points among the predictors

(see, e.g., Cribari-Neto, 2004; Wilcox, 2001). Therefore, developing new and accurate test statistics for finite samples is the main concern of this article.

In the remaining sections, we first establish notation and give a brief review of the HCCM estimators. We follow this with our main result, a new heteroscedasticity-consistent test of linear hypotheses in OLS regression based on MacKinnon and White's (1985) HC2 estimator. Our initial motivation comes from an interesting connection between HC2 and the heterogeneous-variance two-sample t statistic. Our derivations are inspired primarily by Fai and Cornelius's (1996) Satterthwaite approximation and secondarily by Alexander and Govern's (1994) normalized t approximation. The new test reduces to Lipsitz, Ibrahim, and Parzen's (1999) test when applied to single-degree-of-freedom hypotheses. We report a set of simulations showing that our method outperforms existing ones and describe a SAS macro that implements the new method.

1. Notations

Following convention, we write the OLS linear regression model as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1)$$

where \mathbf{y} is an $N \times 1$ vector of outcome observations, \mathbf{X} is an $N \times p$ full rank fixed matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters, and $\boldsymbol{\varepsilon}$ is an $N \times 1$ vector of errors. Here, N is the total number of observations, and p is the number of predictors (including the intercept) in the regression equation. We assume that the errors are normally distributed, and $E(\boldsymbol{\varepsilon}) = \mathbf{0}$, and $\text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Phi} = \text{diag}[\sigma_i^2]$. Under the assumption of homoscedasticity, the elements in the error vector have constant variance σ^2 , and then $\boldsymbol{\Phi} = \sigma^2 \mathbf{I}_N$, where \mathbf{I}_N is an identity matrix of order N .

The OLS estimator of regression coefficients is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, so

$$\text{var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Sigma} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Phi}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \quad (2)$$

We denote the i th OLS residual as $e_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$, where y_i is the i th element in the \mathbf{y} vector, and \mathbf{x}_i is the i th row of the \mathbf{X} matrix. The OLS residual vector is $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = (\mathbf{I}_N - \mathbf{H})\mathbf{y}$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the "hat matrix" whose diagonal elements $h_{ii} = \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'$ are the "leverage" values. When $\boldsymbol{\Phi} = \sigma^2 \mathbf{I}_N$ is inserted into Equation 2, it reduces to $\boldsymbol{\Sigma}_{OLS} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. In practice, σ^2 can be estimated using the mean square error ($\text{MSE} = [N - p]^{-1}\sum_{i=1}^N e_i^2$) of the regression model, which leads to the OLSCM estimator,

$$\hat{\boldsymbol{\Sigma}}_{OLS} = \text{MSE}(\mathbf{X}'\mathbf{X})^{-1}, \quad (3)$$

whose diagonal elements are the squared standard errors printed in the output of almost all regression software packages.

2. A Brief Review of HCCM Estimators

We provide a concise review of HCCM estimators, roughly following their chronological order. On the basis of the work of Eicker (1963, 1967) and Huber (1967), White (1980) popularized the HC0 estimator,

$$\hat{\Sigma}_{HC0} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}[e_i^2]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \quad (4)$$

On the surface, it appears that HC0 aims at estimating Φ in Equation 2 with $\text{diag}[e_i^2]$, a difficult task with only N residuals available. However, what is being estimated is in fact an average of expectations, $N^{-1}\mathbf{X}'\Phi\mathbf{X} = N^{-1}\sum_{i=1}^N\mathbf{x}'_i\mathbf{x}_i\sigma_i^2$ (see Rothenberg, 1988). HC0 is asymptotically justified, but for small sample sizes, it is usually biased downward, resulting in poor Type I error rate control for hypothesis tests (Bera, Suprayitno, & Premaratne, 2002; Chesher & Jewitt, 1987; Cribari-Neto, Ferrari, & Cordeiro, 2000; Cribari-Neto & Zarkos, 2001; Long & Ervin, 2000).

Numerous researchers have proposed finite sample variants of HC0. They are all intimately related to the jackknife method and have in common the use of leverage-adjusted residuals. For instance, letting $e_i^* = e_i/\sqrt{1-h_{ii}}$, and constructing a diagonal matrix using e_i^{*2} instead of e_i^2 , we obtain the HC2 estimator:

$$\hat{\Sigma}_{HC2} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}\left[\frac{e_i^2}{1-h_{ii}}\right]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \quad (5)$$

HC2 is an unbiased estimator under homoscedasticity (see, e.g., Kauerman & Carroll, 2001). The roots of HC2 can be traced back to Horn, Horn, and Duncan's (1975) "almost unbiased estimator" of heteroscedastic variances. With a balanced design, the leverage values become constants, and HC2 reduces to Hinkley's (1977) HC1:

$$\hat{\Sigma}_{HC1} = N/(N-p)\hat{\Sigma}_{HC0}. \quad (6)$$

Although simple, HC1 is almost never recommended or used, because it exhibits much of the same finite sample bias from which HC0 suffers.

Discounting OLS residuals further by a factor of $(1-h_{ii})$ instead of $\sqrt{1-h_{ii}}$ leads to a slightly biased but better estimator in terms of Type I error rate control:

$$\hat{\Sigma}_{HC3} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}\left[\frac{e_i^2}{1-h_{ii}}\right]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \quad (7)$$

This is the HC3 estimator given by Davidson and MacKinnon (1993) as an approximation to the original (jackknife-based) HC3 of MacKinnon and White (1985). Until recently, HC3 has performed the best in simulations (see, e.g., Long & Ervin, 2000), but when there are points of high leverage among the \mathbf{X}

variables and the errors have normal or light-tailed distributions, HC3 may also fail (Chesher & Jewitt, 1987; Kauerman & Carroll, 2001; Wilcox, 2001). This naturally leads to Cribari-Neto's (2004) HC4 estimator,

$$\hat{\Sigma}_{HC4} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{diag}\left[\frac{e_i^2}{(1-h_{ii})^{\delta_i}}\right]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad (8)$$

where the exponent $\delta_i = \min\{4, Nh_{ii}/p\}$ uses case i 's leverage value explicitly to determine the level of discounting. Although t tests using HC4 standard errors seem to perform well, HC4 is far more biased and variable than HC2 and HC3.

Instead of adjusting the estimator itself, one may obtain tests having nominal size by approximating the distribution of the test statistic. This is precisely what Kauerman and Carroll (2001) and Lipsitz et al. (1999) suggested. Both proposed degrees-of-freedom corrections for single-degree-of-freedom null hypotheses. In the next section, we consider the general case of multiple-degrees-of-freedom linear hypotheses with a new χ^2 approximation.

3. A New Test of Linear Hypotheses

The subsequent derivations are based on the HC2 estimator. We choose HC2 for two reasons. First, our initial motivation comes from a connection between the HC2 estimator and the standard error in Welch's (1938) heterogeneous-variance two-sample t statistic. Specifically, when the mean difference is represented as one regression coefficient by choosing an appropriately reparametrized design matrix, the HC2 standard error for that coefficient coincides with the Welch heterogeneous variance standard error. A proof is given in Appendix A. It is well known that degrees-of-freedom adjustment is generally required for the Welch t , suggesting that an approximation for test statistics based on HC2 may also be necessary. The second reason we prefer HC2 over other HCCM estimators lies in the fact that it is unbiased under homoscedasticity and in general less biased than other HCCM estimators under heteroscedasticity. The unbiasedness of HC2 is an important property that we make use of when deriving the new test statistic.

3.1. The Linear Hypothesis

A rank q linear hypothesis for the OLS regression model is of the form

$$H_0 : \mathbf{L}'\beta = \theta_0, \quad (9)$$

where \mathbf{L} is a $(p \times q)$ coefficient matrix of full column rank, and θ_0 is a $q \times 1$ vector of known constants. In general, for large sample testing, the test statistic is of the form

$$Q = (\mathbf{L}'\hat{\beta} - \theta_0)'(\mathbf{L}'\hat{\Sigma}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \theta_0), \quad (10)$$

where $\hat{\Sigma} = \widehat{\text{var}}(\hat{\beta})$ is the estimated variance covariance matrix of $\hat{\beta}$. Q is asymptotically χ^2 distributed with q degrees of freedom. When the OLSCM estimator is used for $\hat{\Sigma}$, an exact result exists, that is,

$$F_{OLS} = Q_{OLS}/q = (\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0)'(\mathbf{L}'\hat{\Sigma}_{OLS}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0)/q, \quad (11)$$

and F_{OLS} follows an $F(q, N - p)$ distribution when all the distributional assumptions are satisfied (see, e.g., Searle, 1971). Through the use of OLSCM estimator, F_{OLS} assumes homoscedasticity. To construct a test that does not require this assumption, we can replace OLSCM with HC2, and use the test statistic

$$Q_{HC2} = (\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0)'(\mathbf{L}'\hat{\Sigma}_{HC2}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0), \quad (12)$$

with reference to a χ^2 distribution with q degrees of freedom for large sample testing. Unfortunately, this χ^2 approximation can be rather poor. To remedy the situation, we propose a method for finding a test statistic that also has an approximate χ^2 distribution with q degrees of freedom but behaves much better than Q_{HC2} in finite samples.

3.2. The New Test Statistic

First, we consider the spectral decomposition of $\mathbf{L}'\hat{\Sigma}_{HC2}\mathbf{L}$, such that $\boldsymbol{\Gamma}'\mathbf{L}'\hat{\Sigma}_{HC2}\mathbf{L}\boldsymbol{\Gamma} = \boldsymbol{\Lambda}$, and $(\mathbf{L}'\hat{\Sigma}_{HC2}\mathbf{L})^{-1} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}'$, where $\boldsymbol{\Gamma}$ is an orthogonal matrix containing the eigenvectors of $\mathbf{L}'\hat{\Sigma}_{HC2}\mathbf{L}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues, that is, $\boldsymbol{\Lambda} = \text{diag}[\lambda_j]$, where λ_j is the j th eigenvalue. Then Q_{HC2} may be written as

$$\begin{aligned} Q_{HC2} &= (\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0)'(\mathbf{L}'\hat{\Sigma}_{HC2}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0) \\ &= (\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0)'\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}'(\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0) \\ &= \sum_{j=1}^q [\gamma_j'(\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0)]^2 / \lambda_j \\ &= \sum_{j=1}^q t_j^2, \end{aligned} \quad (13)$$

where γ_j is the eigenvector corresponding to the j th eigenvalue, and

$$t_j = \frac{\gamma_j'(\mathbf{L}'\hat{\beta} - \boldsymbol{\theta}_0)}{\sqrt{\lambda_j}} = \frac{\gamma_j'\mathbf{L}'\hat{\beta} - \gamma_j'\boldsymbol{\theta}_0}{\sqrt{\lambda_j}}. \quad (14)$$

But $\boldsymbol{\Gamma}'\mathbf{L}'\hat{\Sigma}_{HC2}\mathbf{L}\boldsymbol{\Gamma} = \boldsymbol{\Lambda}$ implies that $\lambda_j = \gamma_j'\mathbf{L}'\hat{\Sigma}_{HC2}\mathbf{L}\gamma_j = \widehat{\text{var}}(\gamma_j'\mathbf{L}'\hat{\beta})$, so if the true parameter covariance matrix were known, the t_j values would be independent

standard normal variates under the null hypothesis (see Fai & Cornelius, 1996). Although Fai and Cornelius (1996) developed the idea of using spectral decomposition to write the quadratic form test statistic as the sum of approximately independent variates for mixed-effects models, we believe that we are the first to apply this technique to OLS regression under heteroscedasticity. The parameter covariance matrix typically has to be estimated, so it is desirable to base the spectral decomposition on an approximately unbiased estimator, and HC2 is a natural candidate. In that case, the t_j values may be regarded as approximately independent Student's t variates each having f_j degrees of freedom, wherein the f_j values can be obtained via a Satterthwaite (1946) approximation:

$$f_j = \frac{2\{E[\widehat{\text{var}}(\gamma_j' \mathbf{L}' \hat{\beta})]\}^2}{\text{var}[\widehat{\text{var}}(\gamma_j' \mathbf{L}' \hat{\beta})]} = \frac{2[E(\lambda_j)]^2}{\text{var}(\lambda_j)}. \quad (15)$$

To find the mean and the variance of λ_j , note that

$$\mathbf{\Gamma}' \mathbf{L}' \hat{\Sigma}_{HC2} \mathbf{L} \mathbf{\Gamma} = \mathbf{\Gamma}' \mathbf{L}' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \text{diag} \left[\frac{e_i^2}{1 - h_{ii}} \right] \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{L} \mathbf{\Gamma} = \mathbf{\Lambda}. \quad (16)$$

Let $\mathbf{c}'_j = [c_{j1}, c_{j2}, \dots, c_{jN}]'$ be the j th row of the matrix $\mathbf{\Gamma}' \mathbf{L}' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'$, $j = 1, \dots, q$. We now see that each λ_j is a quadratic form in \mathbf{e} , because

$$\lambda_j = \sum_{i=1}^N \frac{c_{ji}^2 e_i^2}{1 - h_{ii}} = \mathbf{e}' \mathbf{A}_j \mathbf{e}, \quad (17)$$

where $\mathbf{A}_j = \text{diag}[c_{ji}^2 / (1 - h_{ii})]$. The derivations above reduce to Lipsitz et al.'s (1999) approximation when $q = 1$. However, the original derivation in Lipsitz et al. has a flaw (see their Equation 9, p. 499) that leads to an incorrect \mathbf{A}_j .

Using basic linear model properties and normal theory results on the moments of quadratic forms (Searle, 1971), we find that $E(\lambda_j) = \text{tr}(\mathbf{A}_j \mathbf{\Omega})$, and $\text{var}(\lambda_j) = 2 \text{tr}[(\mathbf{A}_j \mathbf{\Omega})^2]$, where $\mathbf{\Omega} = (\mathbf{I}_N - \mathbf{H}) \mathbf{\Phi} (\mathbf{I}_N - \mathbf{H})$ is the covariance matrix of the residual vector \mathbf{e} . Because neither the expected value nor the variance is known, estimates have to be used, and we replace $\mathbf{\Omega}$ with $\hat{\mathbf{\Omega}} = (\mathbf{I}_N - \mathbf{H}) \hat{\mathbf{\Phi}} (\mathbf{I}_N - \mathbf{H})$, where $\hat{\mathbf{\Phi}} = \text{diag}[e_i^2 / (1 - h_{ii})]$ is Horn et al.'s (1975) "almost unbiased" covariance matrix estimator.² The degrees of freedom of each t_j is then estimated as

$$\hat{f}_j = \frac{2[\hat{E}(\lambda_j)]^2}{\widehat{\text{var}}(\lambda_j)} = \frac{2[\text{tr}(\mathbf{A}_j \hat{\mathbf{\Omega}})]^2}{2 \text{tr}[(\mathbf{A}_j \hat{\mathbf{\Omega}})^2]} = \frac{[\text{tr}(\mathbf{A}_j \hat{\mathbf{\Omega}})]^2}{\text{tr}[(\mathbf{A}_j \hat{\mathbf{\Omega}})^2]}. \quad (18)$$

Once the \hat{f} values are computed, we observe that because the t_j values are approximately independent, a normalizing transformation on each t_j will produce

q approximately independent standard normal deviates. Suppose the transformation is denoted as $T(\cdot)$; then we have $z_j = T(t_j)$, for $j = 1, \dots, q$, where the z_j values are standard normal. Squaring and summing the z_j values, we now arrive at the main result of this article. Let

$$C_{HC2} = \sum_{j=1}^q z_j^2; \tag{19}$$

then C_{HC2} is approximately χ^2 distributed with q degrees of freedom as long as (a) the Satterthwaite approximation on each t_j produces good degree-of-freedom estimates and (b) the normalizing transformation is accurate. The linear hypothesis in Equation 9 will be rejected if C_{HC2} exceeds some cutoff value from the χ^2 distribution with q degrees of freedom. Note that although the use of normalizing transformations in the last step is similar to Alexander and Govern's (1994) normalized t approximation for heteroscedastic analysis-of-variance models, we are able to justify the χ^2 distribution of C_{HC2} on the basis of the approximate independence of the t_j values.

3.3. Normalizing Transformations

To use the result in Equation 19, normalizing transformation must be chosen. There exist a large number of normalizing transformations for the t distribution (see, e.g., Johnson, Kotz, & Balakrishnan, 1994). Historically, much of the research in this area focused on producing highly accurate approximations using the standard normal distribution before extensive tables of the t distribution became widely available. We acknowledge that the choice can be somewhat arbitrary, but among the candidates, two transformations stand out as being both accurate and easy to implement.

The first one is attributable to Hill (1970). This is an approximation based on a generalized Cornish-Fisher expansion. Let $a = \hat{f}_j - 0.5$, $b = 48a^2$, and $c = [a \ln(1 + t_j^2/\hat{f}_j)]^{1/2}$; the transformation is

$$T_H(t_j) = c + \frac{c^3 + 3c}{b} - \frac{4c^7 + 33c^5 + 240c^3 + 855c}{10b^2 + 8bc^4 + 1,000b}. \tag{20}$$

The second one, derived by Wallace (1959), takes the form of

$$T_W(t_j) = \left[1 - \frac{2}{8\hat{f}_j + 3} (1 - e^{-s^2})^{1/2} \right] [\hat{f}_j \ln(1 + t_j^2/\hat{f}_j)]^{1/2}, \tag{21}$$

where

$$s = 0.184(8\hat{f}_j + 3)\hat{f}_j^{-1} [\ln(1 + t_j^2/\hat{f}_j)]^{1/2}.$$

Using the two transformations, we propose two versions of the test statistic:

$$C_{HC2}^H = \sum_{j=1}^q T_H^2(t_j), \quad (22)$$

and

$$C_{HC2}^W = \sum_{j=1}^q T_W^2(t_j). \quad (23)$$

If $q = 1$, our approximation essentially reduces to Lipsitz et al.'s (1999) approximation, because then the transformation simply becomes a way of evaluating the t distribution function.

4. Simulations

To examine the foregoing development, two sets of simulations were conducted. The first set is essentially a replication of Lipsitz et al.'s (1999) simulation with a fixed design matrix and smaller N . The second set mimics Long and Ervin's (2000) simulation, which aims at reproducing characteristics of regression models in cross-sectional data analysis, with somewhat larger N .

4.1. Simulation Design

Consistent with the goal of this article, we focused on hypotheses with more than 1 degree of freedom. Besides having two sets of simulations wherein two different regression models were used, we manipulated sample size and error structure to investigate their effect on Type I error rates of the tests.

4.1.1. The First Model

In the first set of simulations, we used the following model:

$$y_i = 0 + 0.4x_{3i} - 0.25x_{3i}^2 + \varepsilon_i, \quad (24)$$

so $\beta = (0, .4, -.25)'$. As in Lipsitz et al. (1999), we paid special attention to small sample performance by considering sample sizes of 12, 24, and 48, with 12 distinct x_{3i} values fixed to 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, and 10. For $N=24$, we doubled these covariate values, and for $N=48$, we quadrupled them.³ For hypothesis test, we considered

$$\mathbf{L}' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and $\theta_0 = [-0.40, -0.25]'$. In this example, $q = 2$.

4.1.2. The Second Model

The regression model in the second set was

$$y_i = 1 + x_{1i} + x_{2i} + x_{3i} + x_{4i} + \varepsilon_i, \tag{25}$$

so $\beta = (1, 1, 1, 1, 1)'$. The predictors were generated in the a way similar to Long and Ervin's (2000) method. Specifically, let $d_{1i} \sim U(0, 1)$, $d_{2i} \sim N(0, 1)$, $d_{3i} \sim \chi_1^2$, $d_{4i} \sim N(0, 1)$, and $d_{5i} \sim U(0, 1)$, where $U(0, 1)$ stands for the uniform distribution over $(0, 1)$, and χ_1^2 means a χ^2 distribution with 1 degree of freedom. The predictor values were $x_{0i} = 3d_{1i} + 0.6d_{2i}$, $x_{1i} = 1 + d_{1i}$, $x_{2i} = 2d_{1i} + 0.6d_{3i}$, $x_{3i} = 0.1d_{1i} + 0.9d_{3i} - 0.8d_{4i} + 4d_{5i}$, and

$$x_{4i} = \begin{cases} 1, & \text{if } x_{0i} = 1.6, \\ 0, & \text{otherwise.} \end{cases}$$

As noted by Long and Ervin (2000), the predictor correlations are similar to typical cross-sectional data sets, and a mixture of continuous and categorical predictors were included. We considered a 4-degree-of-freedom hypothesis,

$$\mathbf{L}' = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

and $\theta_0 = [1, 1, 1, 1]'$.

4.1.3. Sample Size and Error Structure

There were three sample-size conditions for each set of simulations: $N = 12, 24,$ and 48 for the first set and $N = 50, 100,$ and 300 for the second set. To create heteroscedastic errors (or lack thereof), the error terms ε_i in both sets of simulations were generated from one of the five variance functions (denoted as Error Structures 0–4), as documented in Table 1.

4.1.4. Test Statistics

To test the linear hypotheses, we not only used the new statistics, C_{HC2}^H and C_{HC2}^W , but also considered, as benchmarks, some quasi- F statistics that use both the OLSCM estimator and HCCM estimators. Table 2 summarizes the eight test statistics included in the simulations. We are interested in the Type I error rates of these tests at the nominal α level of .05. Following Lipsitz et al. (1999), with 1,825 replications in each condition and a normal approximation to the binomial

TABLE 1
Sample Size and Error Structure Combinations

Variance Function	First Model			Second Model		
	N = 12	N = 24	N = 48	N = 50	N = 100	N = 300
0. $\varepsilon_i = \varepsilon_i^*$	✓	✓	✓	✓	✓	✓
1. $\varepsilon_i = x_{3i}\varepsilon_i^*$	✓	✓	✓	✓	✓	✓
2. $\varepsilon_i = \sqrt{ x_{3i} }\varepsilon_i^*$				✓	✓	✓
3. $\varepsilon_i = \sqrt{x_{1i}}\sqrt{x_{2i}}\varepsilon_i^*$				✓	✓	✓
4. $\varepsilon_i = \begin{cases} 4\varepsilon_i^*, & \text{if } x_{4i} = 1, \\ \varepsilon_i^*, & \text{if } x_{4i} = 0. \end{cases}$				✓	✓	✓

Note: The ε_i^* values are drawn independently from a standard normal distribution, and checkmarks indicate condition present in the simulations.

TABLE 2
Test Statistics in the Simulation

Test Statistic	Definition	Reference Distribution
F_{OLS}^a	$(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)'(\mathbf{L}'\hat{\Sigma}_{OLS}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)/q$	$F(q, N - p)$
F_{HC0}^a	$(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)'(\mathbf{L}'\hat{\Sigma}_{HC0}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)/q$	$F(q, N - p)$
F_{HC1}^a	$(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)'(\mathbf{L}'\hat{\Sigma}_{HC1}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)/q$	$F(q, N - p)$
F_{HC2}^a	$(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)'(\mathbf{L}'\hat{\Sigma}_{HC2}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)/q$	$F(q, N - p)$
F_{HC3}^a	$(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)'(\mathbf{L}'\hat{\Sigma}_{HC3}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)/q$	$F(q, N - p)$
F_{HC4}^a	$(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)'(\mathbf{L}'\hat{\Sigma}_{HC4}\mathbf{L})^{-1}(\mathbf{L}'\hat{\beta} - \mathbf{\theta}_0)/q$	$F(q, N - p)$
C_{HC2}^{Hb}	$\sum_{j=1}^q T_H^2(t_j)$	χ_q^2
C_{HC2}^{Wb}	$\sum_{j=1}^q T_W^2(t_j)$	χ_q^2

a. Existing test statistic.

b. Newly proposed test statistic.

probabilities, the observed rejection rates should be in the closed interval [0.04, 0.06] for a test to be deemed valid.

4.2. Simulation Results

The results of the simulations are summarized in Tables 3 and 4. The general pattern is clear. The new test works well under all conditions simulated. In addition, the choice of normalizing transformations does not seem to have much impact on the results; the difference between the two versions of the test statistic (C_{HC2}^H and C_{HC2}^W) is negligible.

TABLE 3
Type I Error Rates for Simulation Set 1

	F_{OLS}	F_{HC0}	F_{HC1}	F_{HC2}	F_{HC3}	F_{HC4}	C_{HC2}^H	C_{HC2}^W
Error Structure 0								
$N = 12$	0.057	0.234	0.169	0.140	0.065	0.032	0.042	0.042
$N = 24$	0.053	0.155	0.122	0.105	0.073	0.053	0.054	0.054
$N = 48$	0.053	0.105	0.086	0.076	0.057	0.044	0.042	0.042
Error Structure 1								
$N = 12$	0.106	0.309	0.240	0.175	0.074	0.022	0.049	0.049
$N = 24$	0.106	0.175	0.148	0.125	0.080	0.047	0.053	0.053
$N = 48$	0.112	0.115	0.101	0.095	0.080	0.056	0.056	0.055

Note: The italicized Type I error rates are outside the interval [0.04, 0.06].

TABLE 4
Type I Error Rates for Simulation Set 2

	F_{OLS}	F_{HC0}	F_{HC1}	F_{HC2}	F_{HC3}	F_{HC4}	C_{HC2}^H	C_{HC2}^W
Error Structure 0								
$N = 50$	0.056	0.162	0.132	0.116	0.070	0.044	0.055	0.055
$N = 100$	0.051	0.096	0.083	0.075	0.056	0.041	0.045	0.045
$N = 300$	0.049	0.067	0.065	0.062	0.057	0.046	0.046	0.046
Error Structure 1								
$N = 50$	0.191	0.196	0.159	0.124	0.076	0.036	0.055	0.055
$N = 100$	0.220	0.143	0.132	0.111	0.079	0.044	0.049	0.049
$N = 300$	0.254	0.084	0.080	0.075	0.067	0.053	0.045	0.044
Error Structure 2								
$N = 50$	0.107	0.175	0.142	0.125	0.079	0.035	0.058	0.058
$N = 100$	0.116	0.134	0.116	0.099	0.066	0.044	0.051	0.051
$N = 300$	0.118	0.076	0.072	0.066	0.059	0.052	0.046	0.046
Error Structure 3								
$N = 50$	0.123	0.187	0.155	0.131	0.090	0.042	0.059	0.059
$N = 100$	0.130	0.130	0.117	0.102	0.080	0.048	0.059	0.058
$N = 300$	0.140	0.090	0.085	0.079	0.071	0.059	0.058	0.058
Error Structure 4								
$N = 50$	0.074	0.147	0.113	0.100	0.067	0.037	0.041	0.040
$N = 100$	0.057	0.105	0.094	0.087	0.064	0.039	0.041	0.041
$N = 300$	0.055	0.066	0.063	0.063	0.055	0.044	0.042	0.042

Note: The italicized Type I error rates are outside the interval [0.04, 0.06].

Consistent with previous findings, F_{HC0} , F_{HC1} , and F_{HC2} show unacceptably high Type I error rates, even in the optimal conditions in which heteroscedasticity is not present, though they do seem to approach the nominal test size as N increases. Under the same conditions, the size of F_{HC3} is appreciably smaller, but it can be liberal with small N . On the other hand, F_{HC4} does keep Type I error rates under 0.05 for all conditions, but it tends to be conservative when N is small.

When heteroscedasticity is present, F_{OLS} becomes too liberal, and the bias does not decrease with increasing sample size. On the other hand, there is a clear trend in that F_{HC0} , F_{HC1} , F_{HC2} , and F_{HC3} approach the nominal size as N increases. However, with certain error structures (e.g., Error Structure 3), even F_{HC3} cannot keep the Type I error rate around the nominal level with N as large as 300. With F_{HC4} , it is again on the conservative side when N is small.

5. Software Implementation and an Example

To help researchers adopt the proposed method, we produced a SAS macro implementing both C_{HC2}^H and C_{HC2}^W . Here, we use it to analyze a real data set (called CONCEPT) from the popular introductory statistics textbook by Moore and McCabe (2003). The CONCEPT data set contains information from 78 seventh graders in a midwestern school (for details, see Moore & McCabe, 2003, pp. D1–D2). We make use of the following variables: GPA (grade point average), IQ (score on the IQ test), SEX (*female* = 0, *male* = 1), SC-BEH (score on the Behavioral Adjustment subscale from the Self-Concept Scale; Piers, 1984), and SC-ANX (score on the Freedom From Anxiety subscale from the Self-Concept Scale). We are interested in whether there is a relationship between psychological well-being (as measured by SC-BEH and SC-ANX) and GPA after controlling for IQ and SEX. We consider the following regression model:

$$GPA_i = \beta_0 + \beta_1 IQ_i + \beta_2 SEX_i + \beta_3 SC-BEH_i + \beta_4 SC-ANX_i + \varepsilon_i, \quad (26)$$

and a test of the null hypothesis $H_0 : \mathbf{L}'\beta = \boldsymbol{\theta}_0$ at the usual .05 level, where

$$\mathbf{L}' = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (27)$$

and $\boldsymbol{\theta}_0$ is a 2×1 vector of zeros.

It appears from the residual plot that the assumption of homoscedasticity is suspect. Indeed, Glejser's (1969) test rejected the null hypothesis of no heteroscedasticity, $\chi^2(4) = 10.55, p < .05$. We tested the linear hypothesis in Equation 27 using both F_{OLS} and seven other heteroscedasticity-consistent test statistics included in the simulations. All regression packages can produce F_{OLS} , and

TABLE 5
CONCEPT Analysis

	F_{OLS}	F_{HC0}	F_{HC1}	F_{HC2}	F_{HC3}	F_{HC4}	C_{HC2}^H	C_{HC2}^W
Statistic	5.643	4.479	4.192	3.931	3.444	2.674	5.049	5.045
p	.005 ^a	.015 ^a	.019 ^a	.024 ^a	.037 ^a	.076 ^a	.080 ^b	.080 ^b

a. The reference distribution is $F(2, 73)$.

b. The reference distribution is χ^2_2 .

Hayes and Cai (in press) provide SAS and SPSS macros implementing the quasi- F tests on the basis of HC0 to HC4. The new SAS macro we provide here computes both C_{HC2}^H and C_{HC2}^W . The reader is referred to Appendix B for an explanation of the syntax of the macro.

As shown in Table 5, all tests but C_{HC2}^H , C_{HC2}^W , and F_{HC4} lead to p values less than .05, with $C_{HC2}^H = 5.049$ ($p = .08$) and $C_{HC2}^W = 5.045$ ($p = .08$). Given the relatively small sample size ($N = 78$) and the possibility of heteroscedasticity, the decision based on F_{HC4} , C_{HC2}^H , and C_{HC2}^W should be trusted. Hence, this is an example in which our new approximations may lead to inferences that are qualitatively different from the quasi- F tests.

6. Discussion

Heteroscedasticity can adversely affect inferences in regression drawn from statistical tests on the basis of the OLSCM estimator. Although inferences based on the HCCM estimators are asymptotically correct under heteroscedasticity, they can behave badly when the sample size is small.

In this article, we proposed a new χ^2 test for general linear hypotheses using the HC2 covariance matrix estimator. In our simulations, the new test performed well. More empirical evidence is needed before a definitive conclusion can be reached about the behavior of the new test. However, it should be noted that the present simulation is designed after Wilcox's (2001) suggestion, which aims to expose the weaknesses of methods based on HCCMs. It is therefore refreshing to see that our new method showed correct test size even under those unfavorable conditions.

Our secondary goal was to bring the problem of heteroscedasticity of unknown form to the attention of educational and behavioral researchers testing linear hypotheses in OLS regression models. It should be emphasized, however, that when the error variance structure can be modeled, a model-based approach such as estimated generalized least squares may lead to a gain of efficiency over the approach based on HCCMs. But when modeling is difficult, methods based on HCCMs are much more sound than not taking any corrective measure at all, and it is also when the HCCMs are potentially most useful.

TABLE 6
Relative Biases of the Covariance Matrix Estimators in Simulation Set 1

	OLSCM	HC0	HC1	HC2	HC3	HC4
Error Structure 0						
$N = 12$	0.001	0.387	0.182	0.016	1.260	29.343
$N = 24$	0.001	0.204	0.091	0.011	0.259	0.881
$N = 48$	0.001	0.100	0.040	0.003	0.113	0.264
Error Structure 1						
$N = 12$	0.186	0.482	0.310	0.087	1.430	41.579
$N = 24$	0.183	0.226	0.116	0.026	0.413	1.776
$N = 48$	0.184	0.119	0.061	0.005	0.154	0.466

Note: OLSCM = ordinary least squares estimator of the parameter covariance matrix.

The derivations of the new test are based on the HC2 estimator. It might seem that approximations for HC3 or HC4 could potentially lead to even better results because of the apparent superiority of F_{HC3} and F_{HC4} over F_{HC2} . Unfortunately, this is not true, because our approximation is heavily dependent on the requirement that the spectral decomposition in Equation 13 be based on an approximately unbiased covariance matrix estimator under both homoscedasticity and heteroscedasticity. Table 6 contains the observed average relative bias from the first set of simulations of the unique elements of the parameter covariance matrix for all six covariance matrix estimators. It is clear that HC2 has much smaller bias than the other HCCMs under homoscedasticity, and this property also carries over to heteroscedastic error conditions.

Finally, we would like to point out that the approximation we proposed is a fairly general result whose application is not limited to OLS regression models. It can be potentially useful for mixed-effects modeling and generalized estimation equations.

Appendix A

Let $\mathbf{y}_1 = [y_{11}, y_{21}, \dots, y_{m1}]'$ and $\mathbf{y}_2 = [y_{12}, y_{22}, \dots, y_{m2}]'$ be two samples, such that $N = m + n$ and $\beta = [\beta_0, \beta_1]'$; an OLS regression model of the form

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{1}_m & \mathbf{1}_m \\ \mathbf{1}_n & \mathbf{0}_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \varepsilon \tag{A1}$$

yields $\hat{\beta}_1 = \bar{y}_1 - \bar{y}_2$, and the standard error of $\hat{\beta}_1$ obtained from the ordinary least squares estimator of the parameter covariance matrix (OLSCM) estimator is identical to the denominator in Student's t statistic. Note that homoscedasticity is assumed in Equation A1.

We now show that if we do not assume homoscedasticity, replacing the OLSCM estimator with the HC2 estimator changes the standard error of $\hat{\beta}_1$ to the Welch-Satterthwaite standard error. First, the leverage values are $\mathbf{h} = [h_{11}, h_{22}, \dots, h_{NN}]' = [m^{-1}\mathbf{1}'_m, n^{-1}\mathbf{1}'_n]'$. Next, let \mathbf{c}'_2 be the second row of $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, so $\mathbf{c}_2 = [c_{21}, c_{22}, \dots, c_{2N}]' = [m^{-1}\mathbf{1}'_m, -n^{-1}\mathbf{1}'_n]'$. Note also that the OLS residuals are deviations from the means: $\mathbf{e} = [(y_1 - \bar{y}_1)\mathbf{1}_m]', (y_2 - \bar{y}_2)\mathbf{1}_n]'$. Then, by the definition of HC2,

$$\widehat{\text{var}}(\hat{\beta}_1) = \mathbf{c}'_2 \text{diag} \left[\frac{e_i^2}{1 - h_{ii}} \right] \mathbf{c}_2 = \sum_{i=1}^m \frac{(y_{i1} - \bar{y}_1)^2}{m(m-1)} + \sum_{i=1}^n \frac{(y_{i2} - \bar{y}_2)^2}{n(n-1)} = \frac{S_1^2}{m} + \frac{S_2^2}{n},$$

where S_1^2 and S_2^2 are the sample variances. We now see that this is equal to the squared standard error for the heterogeneous variance t statistic attributable to Satterthwaite (1946) and Welch (1938).

Appendix B

Suppose a SAS data set called CONCEPT is created with the following code:

```
data concept; input gpa iq sex sc_beh sc_anx; ...
```

Then the following call invokes the SAS macro that computes C_{HC2}^H and C_{HC2}^W for the linear hypothesis in Equation 27:

```
%hetreg (data=CONCEPT, dv=gpa, iv=iq sex sc_beh sc_anx, test=2) ;
```

The syntax is self-explanatory: One specifies (a) the input data set with the `data` argument, (b) the dependent variable with the `dv` argument, (c) the predictor set with the `iv` argument, and finally (d) setting the value of the test argument to a nonzero integer (say k) instructs the macro to perform a test of the null hypothesis that the coefficients for the last k variables in the predictor set are jointly zero. If the test option is not included, it defaults to zero, and no setwise test is conducted.

The output from running the above macro is shown in Figure B1. The test result for the linear hypothesis is printed under the “Setwise Hypothesis Test” section, where “C(H)” stands for C_{HC2}^H and “C(W)” stands for C_{HC2}^W . Their p values are “p(H)” and “p(W),” respectively.

Additional features of the macro include (a) an omnibus test of the hypothesis that the population multiple correlation is zero with C_{HC2}^H and C_{HC2}^W and (b) single-degree-of-freedom tests for individual regression coefficients. Note that the latter effectively subsumes Lipsitz, Ibrahim, and Parzen’s (1999) method.

Testing Linear Hypotheses in OLS Regression

Criterion Variable: GPA

Model Fit

R-sq	df	C(H)	p(H)	C(W)	p(W)
0.5247	4.0000	24.8359	0.0001	24.8598	0.0001

OLS Regression Coefficients

	Coeff	C(H)	p(H)	C(W)	p(W)
CONSTANT	-5.3266	6.9032	0.0086	6.8963	0.0086
IQ	0.0951	14.2278	0.0002	14.2413	0.0002
SEX	-0.7943	4.4251	0.0354	4.4222	0.0355
SC_BEH	0.1785	4.2041	0.0403	4.2001	0.0404
SC_ANX	0.0579	0.9248	0.3362	0.9251	0.3362

Setwise Hypothesis Test

Variables in Set:		SC_BEH	SC_ANX		
C(H)	p(H)	C(W)	p(W)	df	
5.0490	0.0801	5.0452	0.0802	2.0000	

FIGURE B1. SAS macro output for CONCEPT analysis.

Notes

1. Liang and Zeger (1986) followed the same rationale when deriving their generalized estimating equation approach for generalized linear models.

2. Lipsitz, Ibrahim, and Parzen's (1999) derived an alternative variance estimator $\widehat{\text{var}}(\mathbf{e}'\mathbf{A}_j \mathbf{e}) = 2\text{tr}[\mathbf{M}(\mathbf{M} \circ \mathbf{Y})]$, where \circ stands for the Hadamard product of two matrices. \mathbf{M} is equal to $(\mathbf{I}_N - \mathbf{H})\mathbf{A}_j(\mathbf{I}_N - \mathbf{H})$, and \mathbf{Y} uses a combination

of the leverage values, e_i^4 , and $e_i^2 e_j^2$ to estimate higher order identities such as σ_i^4 and $\sigma_i^2 \sigma_j^2$. We do not adopt this approach, because the higher order sample moments tend to become unstable for small N . In simulations not reported here, their approach did not perform as well as the estimator we proposed.

3. Note that by including a quadratic term in the regression equation, we have created design points with high leverage, a condition more conducive to revealing the weaknesses of the HCCM estimators (see Wilcox, 2001).

4. The macro can be downloaded at <http://www.comm.ohio-state.edu/ahayes/macros.htm>.

References

- Alexander, R. A., & Govern, D. M. (1994). A new and simpler approximation for ANOVA under variance heterogeneity. *Journal of Educational Statistics, 19*, 91–101.
- Bera, A. K., Suprayitno, T., & Premaratne, G. (2002). On some heteroscedasticity-robust estimators of variance-covariance matrix of the least-squares estimators. *Journal of Statistical Planning and Inference, 108*, 121–136.
- Carroll, R. J., & Ruppert, D. (1988). *Transformation and weighting in regression*. New York: Chapman & Hall.
- Chesher, A., & Jewitt, I. (1987). The bias of a heteroscedasticity consistent covariance matrix estimator. *Econometrica, 55*, 1217–1222.
- Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika, 70*, 1–10.
- Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics and Data Analysis, 45*, 215–233.
- Cribari-Neto, F., Ferrari, S. L. P., & Cordeiro, G. M. (2000). Improved heteroscedasticity-consistent covariance matrix estimators. *Biometrika, 87*, 907–918.
- Cribari-Neto, F., & Zarkos, S. G. (2001). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. *Journal of Statistical Computation and Simulation, 68*, 391–411.
- Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. New York: Oxford University Press.
- Eicker, F. (1963). Asymptotic normality and consistency of the least squares estimator for families of linear regression. *Annals of Mathematical Statistics, 34*, 447–456.
- Eicker, F. (1967). Limit theorems for regression with unequal and dependent errors. In L. M. LeCam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (pp. 59–82). Berkeley: University of California Press.
- Fai, A. H. T., & Cornelius, P. L. (1996). Approximate F -tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation, 54*, 363–378.
- Glejser, H. (1969). A new test for heteroskedasticity. *Journal of the American Statistical Association, 64*, 316–323.
- Hayes, A. F., & Cai, L. (in press). Using heteroscedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*.

- Hill, G. W. (1970). Algorithm 395. Student's t -distribution. *Communications of the ACM*, 13, 617–619.
- Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics*, 19, 285–292.
- Horn, S. D., Horn, R. A., & Duncan, D. B. (1975). Estimating heteroscedastic variances in linear models. *Journal of the American Statistical Association*, 70, 380–385.
- Huber, P. J. (1967). The behavior of maximum likelihood estimation under nonstandard conditions. In L. M. LeCam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (pp. 221–223). Berkeley: University of California Press.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distributions, Vol. 1* (2nd ed.) New York: John Wiley.
- Kauerman, G., & Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96, 1387–1396.
- Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22.
- Lipsitz, S. R., Ibrahim, J. G., & Parzen, M. (1999). A degrees-of-freedom approximation for a t -statistic with heterogeneous variance. *Statistician*, 48, 495–506.
- Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *American Statistician*, 54, 217–224.
- MacKinnon, J. G., & White, H. (1985). Some heteroscedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29, 305–325.
- Maddala, G. S. (1971). Generalized least squares with an estimated covariance matrix. *Econometrica*, 39, 23–33.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. New York: Academic Press.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.) New York: Chapman & Hall.
- Moore, D. S., & McCabe, G. P. (2003). *Introduction to the practice of statistics* (4th ed.). New York: W. H. Freeman.
- Piers, E. V. (1984). *The Piers-Harris Children's Self-Concept Scale-Revised*. Los Angeles: Western Psychological Services.
- Rothenberg, T. J. (1988). Approximate power functions for some robust tests of regression coefficients. *Econometrica*, 56, 997–1019.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrical Bulletin*, 2, 110–114.
- Searle, S. R. (1971). *Linear models*. New York: John Wiley.
- Wallace, D. L. (1959). Bounds on normal approximations to Student's and the chi-square distributions. *Annals of Mathematical Statistics*, 30, 1121–1130.
- Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29, 350–362.
- White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica*, 48, 817–838.
- Wilcox, R. R. (2001). Comment on “Using heteroscedasticity consistent standard errors in the linear regression model” by J. S. Long & L. H. Ervin. *American Statistician*, 55, 374–375.

Authors

LI CAI is a doctoral candidate in the L. L. Thurstone Psychometric Lab at the University of North Carolina at Chapel Hill, Department of Psychology, CB#3270, Davie Hall, Chapel Hill, NC 27599-3270; cai@unc.edu. His current research is centered on computational methods for latent-variable models, particularly the development and computer implementation of parameter estimation algorithms involving stochastic approximation and Markov-chain Monte Carlo. He has previously published in the areas of goodness-of-fit testing in item response theory, power analysis in structural equation modeling, and permutation methods.

ANDREW F. HAYES (PhD, Cornell University, 1996) is an associate professor in the School of Communication, The Ohio State University, 3066 Derby, 154 N. Oval Mall, Columbus, OH 43210; hayes.338@osu.edu. He is the author of *Statistical Methods for Communication Science* (Lawrence Erlbaum, 2005) and coeditor (with Michael D. Slater and Leslie B. Snyder) of *The Sage Handbook of Advanced Data Analysis Methods for Communication Research* (Sage, 2007), and he has published several dozen journal articles in the areas of public opinion, political communication, social psychology, and statistical methods. He also serves as associate editor of *Communication Methods and Measures*.

Manuscript received July 19, 2004

Accepted June 11, 2006