Performance of OLS and HCCM Estimators in Heteroscedastic ANCOVA Models

Erica L. Dawson **David T. Redden** **T. Mark Beasley**
University of Alabama at Birmingham

We simulated several heteroscedastic scenarios for an Analysis of Covariance (ANCOVA) model with an orthogonal fixed covariate. We examined the Type 1 and 2 error rates of OLS and Heteroscedastic Consistent Covariance Matrix (HCCM) estimators for detecting adjusted mean differences and covariate effects when heteroscedasticity was a function of differences in group variances, the covariate, and both processes in unbalanced ANCOVA models. Results indicate that under complete null models, heteroscedasticity due to an orthogonal covariate alone did not affect the Type 1 error rate for the OLS test of adjusted group mean differences; however, heteroscedasticity due to both the covariate and group, can attenuate, exacerbate, or reverse the known effects of heteroscedasticity on test size in unbalanced models. Furthermore, heteroscedasticity due to the covariate drastically affected the power of the tests of the group effect. HC2 and HC3 tests of the group effect were the most powerful among tests with valid test size; however, both HC2 and HC3 tests for the covariate effect inflated the Type 1 error rate. The $C^2_{(H)}$ test proposed by Cai and Hayes (2008) was promising with Type 1 error rates held below the nominal alpha for both tests of the group and covariate effects in all conditions simulated; however, it was conservative in terms of statistical power.

The utility of Ordinary Least Squares (OLS) regression is unquestionable; however, the assumption of homoscedasticity is quite stringent and often leads to disturbances in both Type 1 and 2 error rates (Zimmerman, 2004). The assumption of homoscedasticity is unlikely to hold in many applied settings. Unfortunately, problems associated with heteroscedasticity in OLS regression have been largely ignored in many research disciplines. Given the large amount of published research that has used OLS regression models without even examining much less correcting for potential heteroscedasticity, the validity of the many research findings may be questionable.

In the educational and behavioral sciences, many simulation studies have focused on heteroscedasticity in analysis of variance (ANOVA) designs, thus coining the term "heterogeneity of variance." The effect of heteroscedasticity in the presence of unbalanced sample sizes for the independent-samples *t*-test results in the well-documented Type 1 error rate suppressions and inflations with positive and negative variance-sample size pairings, respectively. Many studies have shown the superior performance of the Welch (1938) correction of the degrees-of-freedom (*df*s) for the *t* and *F* test statistics with unequal sample sizes. In fact, Zimmerman (2004) demonstrated that "using a separate-variances test unconditionally whenever sample sizes are unequal" ensured the validity of testing mean differences.

Although predictable in some circumstances (i.e., unbalanced ANOVA), the effects of heteroscedasticity on the validity of statistical tests in General Linear Models (GLMs) is complicated. In more complex designs and for regression models in general, the "balance" of sample size is difficult to conceptualize. For example, in a regression model with continuous predictors (X), there may be only a few cases (or even one case) with a particular combination of X values. Thus, how the pairing of heteroscedasticity and sample size will affect the validity of tests from a more complex GLM is not easily discerned from the known effects of heterogeneity of variance in ANOVA models.

Heteroscedasticity in GLMs has been a concern in econometrics for decades (e.g., White 1980), resulting in the development of Heteroscedasticity Consistent Covariance Matrix (HCCM) estimators as a large sample solution to hypothesis testing under heteroscedasticity (see following section for more detail). Recently, Hayes and Cai (2007) examined the statistical properties of the HC0 (Huber, (1967; White, 1980), HC1 (Hinkley, 1977), HC2 (MacKinnon & White, 1985) HC3 (Davidson & MacKinnon, 1993), and HC4 (Cribari-Neto, 2004) HCCM estimators, thus reacquainting statistical methodologists in the educational and behavioral sciences with the pernicious problems of heteroscedasticity in GLMs. Long and Ervin (2000) evaluated the empirical power functions of OLS and HCCM methods and recommended HC3 because it kept the test size at the nominal alpha regardless of the presence or absence of heteroscedasticity. However, the performance of HC3 does depend to some extent on the presence or absence of points of high leverage in the design matrix (e.g., Chesher & Jewitt, 1987; Kauermann &

Carroll, 2001; Wilcox, 2001). Furthermore, Long and Ervin (2000) showed that HC3 can have a liberal bias in very small samples. Cai and Hayes (2008) proposed Chi-Square approximate tests for regression hypotheses under heteroscedasticity of unknown form and have shown that their tests and HC3 outperformed HC2 in terms of maintaining valid Type 1 error rates.

To date, the performance of HCCM estimators in traditional Analysis of Covariance (ANCOVA) designs have not been directly addressed. We investigated the performance of the OLS, HCCM, and the Cai and Hayes (2008) tests under several heteroscedastic scenarios for the following ANCOVA model with an orthogonal fixed covariate:

$$Y_i = \beta_0 + \beta_G G_i + \beta_X X_i + \varepsilon_i \,, \tag{1}$$

where Y_i represents the outcome variable for the i^{th} subject, G represents the grouping variable (G = 0 if the observation in group 1 and G = 1 if observation is in group 2), X represents the fixed covariate, the β's are fixed regression parameters, and ε_i are normally distributed population errors.

Heteroscedasticity Consistent Covariance Matrix (HCCM) Estimators

The covariance matrix for regression coefficients can formulated as:

$$\Sigma \;= (\mathbf{X'X})^{-1}\mathbf{X'\Phi X}(\mathbf{X'X})^{-1} \;. \tag{2}$$

For OLS, $\mathbf{\Phi} = s^2\mathbf{I}_N$; s^2 is the Mean Square Error (*MSE*) and $\mathbf{I}_N$ is an *N*-dimensional Identity Matrix. With s^2 being a scalar, this reduces to the familiar: $\Sigma_{OLS} = MSE(\mathbf{X'X})^{-1}$. White (1980) suggested placing the i^{th} squared residual into the i^{th} diagonal of the $\mathbf{\Phi}$ matrix, using the OLS residuals as estimators of the errors. Thus, $\mathbf{\Phi} = \text{diag}[e^2_i]$ is a diagonal matrix with the squared OLS residuals on the diagonal and all off diagonal entries equal to zero, Thus, the HC0 estimator defined as:

$$\Sigma_{HC0} = (\mathbf{X'X})^{-1}\mathbf{X'}\,\text{diag}[e^2_i]\mathbf{X}(\mathbf{X'X})^{-1} \tag{3}$$

The HC2 estimator weights e^2_i by $1/(1 - h_{ii})$ from the diagonal of the hat matrix :

$$\Sigma_{HC2} = (\mathbf{X'X})^{-1}\mathbf{X'}\,\text{diag}[\frac{e_i^2}{(1-h_{ii})}]\mathbf{X}(\mathbf{X'X})^{-1} \tag{4}$$

The HC3 gives an extra penalty to high leverage values by weighting e^2_i by $1/(1 - h_{ii})^2$:

$$\Sigma_{HC3} = (\mathbf{X'X})^{-1}\mathbf{X'}\,\text{diag}[\frac{e_i^2}{(1-h_{ii})^2}]\mathbf{X}(\mathbf{X'X})^{-1} \tag{5}$$

Hypotheses and Test Statistics

We evaluated the Type 1 and 2 error rates of four procedures for both the test of Group (H_0: $\beta_G = 0$) and the test of the Covariate Effects (H_0: $\beta_X = 0$) from the ANCOVA model in (1). The literature in HCCM methodology has approached this issue from the perspective of testing general linear hypotheses of the form:

$$H_0\text{: } \mathbf{L'\beta} = \mathbf{\theta}_0 \;; \tag{6}$$

where $\mathbf{L}$ is *p*x*q* coefficient matrix of full column rank and $\mathbf{\theta}_0$ is *q*x1 vector of known constants.

For OLS-based tests, the usual *F*-test for general linear hypotheses can be formed as:

$$F_{(q,N\text{-}p)} = \frac{(\mathbf{L'b}-\mathbf{\theta_0})'[\mathbf{L'}(\mathbf{X'X})^{-1}\mathbf{L}]^{-1}(\mathbf{L'b}-\mathbf{\theta_0})/q}{\text{SS}_{\text{Error}}/(N-p)} \,, \tag{7}$$

where $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$ is a *p*x1 vector of OLS sample regression coefficients, SS_{Error} is the Sums of Squares for the Error Term, *N* is the total sample size, *q* is the number of rows for $\mathbf{L'}$, which equals the *df*s for the hypothesis of interest, *p* is the number of columns in the full design matrix, and *N-p* is the Error *df*s for the model.

To construct a test that does not require the OLS assumption of homoscedasticity, the following HCCM statistic:

$$Q = (\mathbf{L'\beta} - \mathbf{\theta}_0)'[\mathbf{L'\Sigma L}]^{-1}(\mathbf{L'\beta} - \mathbf{\theta}_0) \tag{8}$$

is assumed to follow a Central χ^2 distribution with *q df*s under the null hypothesis (6), asymptotically. For testing in samples, $\mathbf{\beta}$ is replaced with $\mathbf{b}$. For HC2, the Variance Estimator $\mathbf{\Sigma}$ equals Σ_{HC2} in (4); and for HC3, $\mathbf{\Sigma} = \Sigma_{HC3}$ in (5). Several authors suggest forming an *F*-test by dividing the *Q* statistic in (8) by its *df*s = q (number of rows for $\mathbf{L}'$) and referencing a Central F distribution with q and N-p *df*s for testing the hypothesis in (6).

Cai and Hayes (2008) proposed Chi-Square approximate tests for regression hypotheses under heteroscedasticity of unknown form based on normalizations of orthogonalized HCCM test statistics (Alexander & McGovern, 1994). Their approach uses spectral decomposition of $\mathbf{L}'\Sigma_{HC2}\mathbf{L}$ to construct a test statistic as the sum of approximately independent variables. Specifically, the spectral decomposition of $\mathbf{L}'\Sigma_{HC2}\mathbf{L}$ can be expressed as $\mathbf{\Gamma}'\mathbf{L}'\Sigma_{HC2}\mathbf{L}\mathbf{\Gamma} = \mathbf{\Lambda}$; where $\mathbf{\Gamma}$ is a qxq orthogonal matrix containing the eigenvectors of $\mathbf{L}'\Sigma_{HC2}\mathbf{L}$ and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues: $\mathbf{\Lambda} = \text{diag}[\lambda_j]$, where λ_j is the j^{th} eigenvalue of $\mathbf{L}'\Sigma_{HC2}\mathbf{L}$, for $j = 1,..., q$. From this, $[\mathbf{L}'\Sigma_{HC2}\mathbf{L}]^{-1} = \mathbf{\Gamma}\mathbf{\Lambda}^{-1}\mathbf{\Gamma}'$. Thus,

$$\begin{aligned} Q_{[\text{HC2}]} &= (\mathbf{L}'\mathbf{b} - \boldsymbol{\theta}_0)'[\mathbf{L}'\Sigma_{HC2}\mathbf{L}]^{-1}(\mathbf{L}'\mathbf{b} - \boldsymbol{\theta}_0) \\ &= (\mathbf{L}'\mathbf{b} - \boldsymbol{\theta}_0)'[\,\mathbf{\Gamma}\mathbf{\Lambda}^{-1}\mathbf{\Gamma}'](\mathbf{L}'\mathbf{b} - \boldsymbol{\theta}_0) \\ &= \textstyle\sum_{j=1}^{q} \gamma_j'(\mathbf{L}'\mathbf{b} - \boldsymbol{\theta}_0)^2/\lambda_j \\ &= \textstyle\sum_{j=1}^{q} t_j^2 \end{aligned} \tag{9}$$

where γ_j is the j^{th} eigenvector for $j = 1, \ldots, q$. Therefore,

$$t_j = \gamma_j'(\mathbf{L}'\mathbf{b} - \boldsymbol{\theta}_0)/\sqrt{\lambda_j} \quad . \tag{10}$$

If the true parameter covariance matrix were actually known, the t_j values would be independent standard normal variables under the null hypothesis (Fai & Cornelius, 1996). Since the parameter covariance matrix is unknown and has to be estimated and the t_j values are approximately independent Student's t variables having f_j *df*s, where:

$$f_j = \frac{2E(\lambda_j)^2}{var(\lambda_j)} = \frac{2[tr(\mathbf{A}_j\mathbf{\Omega})]^2}{2[tr(\mathbf{A}_j\mathbf{\Omega})^2]} = \frac{[tr(\mathbf{A}_j\mathbf{\Omega})]^2}{[tr(\mathbf{A}_j\mathbf{\Omega})^2]} \tag{11}$$

Let $\boldsymbol{c}'_j = [c_{1j}, c_{2j}, \ldots, c_{Nj}]'$ be the j^{th} row of the matrix $\mathbf{L}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Cai and Hayes (2008) show that λ_j is a quadratic form transformation of the residuals, $\mathbf{e}$:

$$\lambda_j = \mathbf{e}'\mathbf{A}_j\mathbf{e} = \sum_{i=1}^{N} \frac{c_{ij}^2 e_i^2}{(1-h_{ii})} \tag{12}$$

where $\mathbf{A}_j = \text{diag}[c^2_{ij}/(1\text{-}h_{ii})]$. It can be demonstrated that $\text{E}(\lambda_j) = tr(\mathbf{A}_j\mathbf{\Omega})$ and $\text{var}(\lambda_j) = 2tr([\mathbf{A}_j\mathbf{\Omega}]^2)$; where $\mathbf{\Omega} = (\mathbf{I}_N - \mathbf{H})\mathbf{\Phi}(\mathbf{I}_N - \mathbf{H})$ and $\mathbf{H}$ is the hat matrix. In the sample, HC2 is used and $\mathbf{\Phi}$ is estimated by $\text{diag}[e^2_i/(1\text{-}h_{ii})]$, and the *df*s of each t_j is estimated as f_i.

Because the t_j values are approximately independent, a normalizing transformation on each t_j will produce q approximately independent standard normal deviates. Thus, squaring and summing these q independent standard normal deviates will produce a χ^2deviate with q *df*s.

Cai and Hayes (2008) chose two normalization transformations that yielded very similar results. We focus on the transformation attributable to Hill (1970), which is based on the generalized Cornish-Fisher expansion:

$$T_H(t_j) = w_j + \frac{w_j^3+3w_j}{v_j} - \frac{4w_j^7+33w_j^5+240w_j^3+855w_j}{10v_j^2+8v_jw_j^4+1000v_j} \quad ; \tag{13}$$

where $u_j = f_j - ½$; $v_j = 48u^2_j$; and $w_j = [u_j\ln(1+(t^2_j/f_j))]^{1/2}$. Subsequently, the test statistic proposed by Cai and Hayes (2008) is formed as:

$$C^H_{\text{HC2}} = \sum_{j=1}^{q} T_H^2(t_j)\ ; \tag{14}$$

which is expected to follow Central χ^2 distribution with q *df*s under the null hypothesis $\boldsymbol{\theta}_0$ in (6).

Method

Based on simulation conditions from previous research (e.g., Long & Ervin, 2000; Cai & Hayes, 2008), we examined the Type 1 error rates of OLS and three HCCM estimators for detecting adjusted mean differences (H_0: $\beta_G = 0$) and covariate effects (H_0: $\beta_X = 0$) when heteroscedasticity was function of differences in group variances, the covariate, and both processes in unbalanced ANCOVA models. All calculations were performed using SAS® PROC IML version 9.3.

Conditions. As in Lipsitz et al. (1999), we paid special attention to small sample performance by considering sample sizes of n_0 =12 and n_1 = 24, with 12 distinct X values fixed to 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, and 10. For the group coded as G=1 with n_1 =24, we doubled these covariate values. This represents a true fixed effect and by design is orthogonal to the indicator variable representing the group (G). Thus, the design matrix stays constant across all replications and conditions. Following the work of Cai and Hayes (2008), we generated a Normal(0,1) error term (ε_i*) with the SAS *rannor* function and transformed it to ε_i in equation 1 using:

$$\varepsilon_i = \frac{\sigma_L}{\sigma_S} G_i \varepsilon_i^* + c X_i \varepsilon_i^* \tag{15}$$

with the Heteroscedasticity Conditions in Table 1. For estimating Type 1 error rates all coefficients in (1) were set to zero (i.e., $\beta_0 = \beta_G = \beta_x = 0$). For estimating statistical power, we set the coefficients in (1) to: $\beta_0 = 0$; $\beta_G = 2.5$; $\beta_x = 0.25$. For these coefficients under complete homoscedasticity, the power of the test of group differences was expected to be 0.79 and the test of the covariate effect was expected to have power of 0.36.

Table 1. Heteroscedasticity Conditions.

Conditions				Description
	σ_L	σ_S	cX	
0	2.5	2.5	0	Completely Homoscedastic
1.1A	2.5	2.5	+1	Heteroscedasticity due to Covariate Only Ascending and Descending Functions of cX
1.1D	2.5	2.5	-1	
1.2A	2.5	2.5	+2	
1.2D	2.5	2.5	-2	
4C.0	4	1	0	Heteroscedasticity due to Group: Conservative σ^2/n Ratio
4C.1A	4	1	+1	Heteroscedasticity due to Group: Conservative σ^2/n Ratio AND Heteroscedasticity due to Covariate Ascending and Descending Functions of cX
4C.1D	4	1	-1	
4C.2A	4	1	+2	
4C.2D	4	1	-2	
4L.0	1	4	0	Heteroscedasticity due to Group: Liberal σ^2/n Ratio
4L.1A	1	4	+1	Heteroscedasticity due to Group: Liberal σ^2/n Ratio AND Heteroscedasticity due to Covariate Ascending and Descending Functions of cX
4L.1D	1	4	-1	
4L.2A	1	4	+2	
4L.2D	1	4	-2	

Tests. From the general linear hypothesis in (6), **L´** = [0 1 0] for the test of Group Effect (H_0: $\beta_G = 0$); and **L´** = [0 0 1] for the test of the Covariate Effect (H_0: $\beta_X = 0$). For both hypotheses: $\boldsymbol{\theta}_0 = 0$, $q = 1$, and $p = 3$. For the OLS, HC2, and HC3 procedures, we calculated *F*-tests and used a Central *F* distribution with $q = 1$ and $N\text{-}p = 33$ *df*s as the reference distribution. The Cai and Hayes (2008) test, C^H_{HC2}, used a Central χ^2 distribution with $q = 1$ *df*s as the reference distribution. All tests were performed at a nominal significance level of $\alpha = 0.05$. Each simulated condition was replicated 5,000 times under the complete null hypothesis; therefore, the Type 1 error rates were expected to fall in the interval $0.05 \pm 1.96[(0.05)(0.95)/(5{,}000)]^{1/2} = [0.044, 0.056]$. Rejection rates below 0.044 were considered to be conservative (suppressed test size), and rates above 0.056 were considered to be liberal (inflated test size).

Results

Figure 1 shows the Type 1 error rates for the test of the Group Effect (H_0: $\beta_G = 0$) for each of the four tests with homogeneous group variances under varying condition of heteroscedasticity due to the covariate. As can be seen, the heteroscedasticity due to the orthogonal covariate had little effect on the Type 1 error rates for tests of group differences under group variance homogeneity. Figure 2 shows the Type 1 error rates for the test of the Group Effect (H_0: $\beta_G = 0$) for each of the four tests with heterogeneous group variances under varying condition of heteroscedasticity due to the covariate. In this scenario, the heterogeneous variances were positively paired with sample sizes, which is known to suppress rejection rates. As would be expected the OLS test suppressed Type 1 error rates in the condition where heteroscedasticity was simply due to the difference in group variances. However, when the

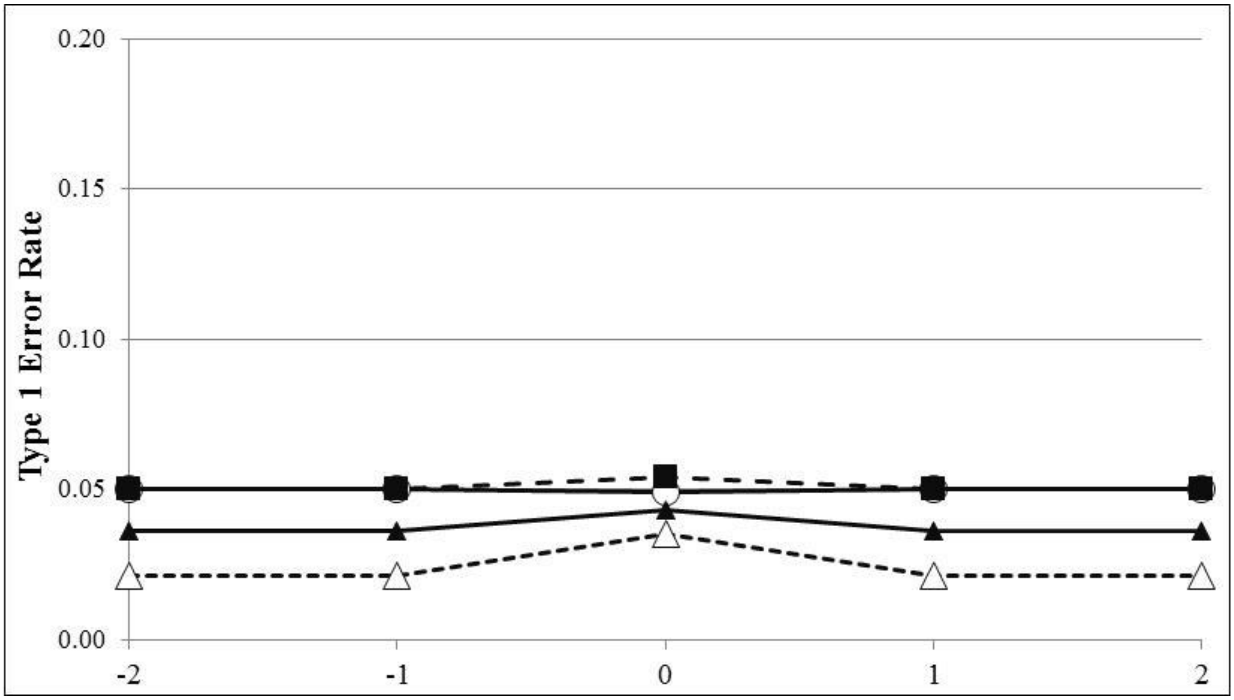

Figure 1. Type 1 Error Rates Test of Group Effect (H_0: $\beta_G = 0$) as a function of Covariate Heterogeneity with Homogeneous Group Standard Deviations: **σ_L: σ_S = 1:1.**

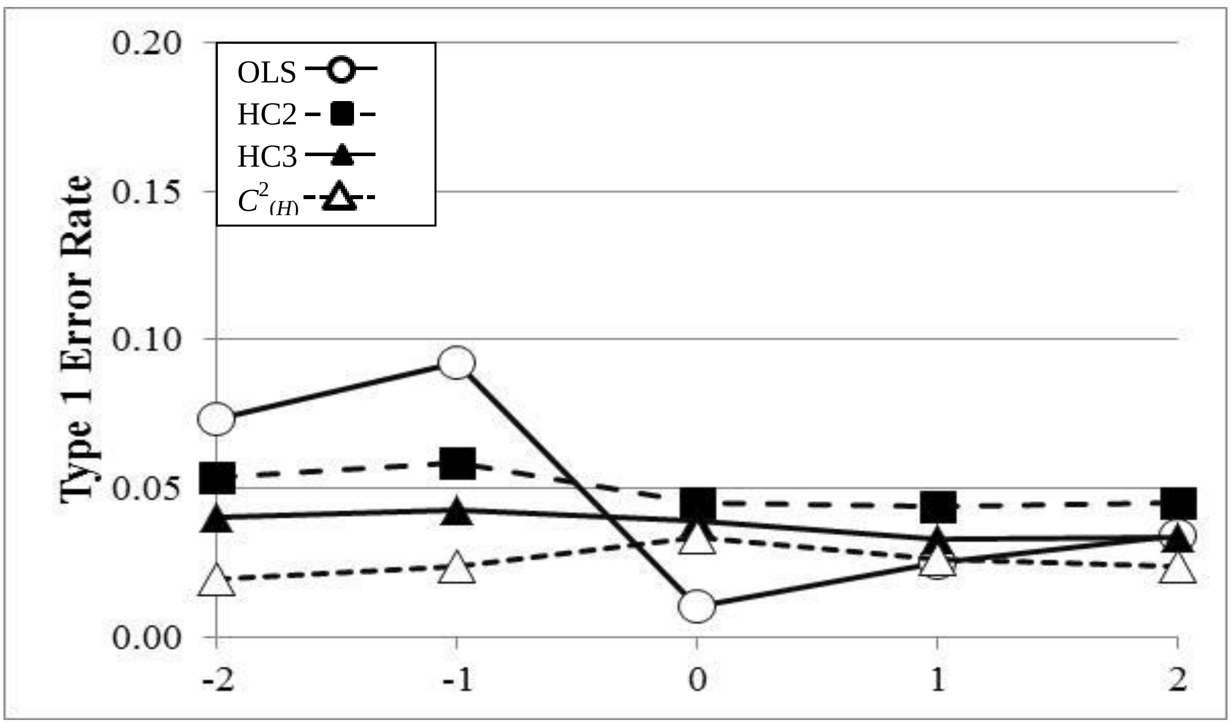

Figure 2. Type 1 Error Rates Test of Group Effect (H_0: $\beta_G = 0$) as a function of Covariate Heterogeneity with Heterogeneous Group Standard Deviations: **σ_L: σ_S = 4:1 (Conservative).**

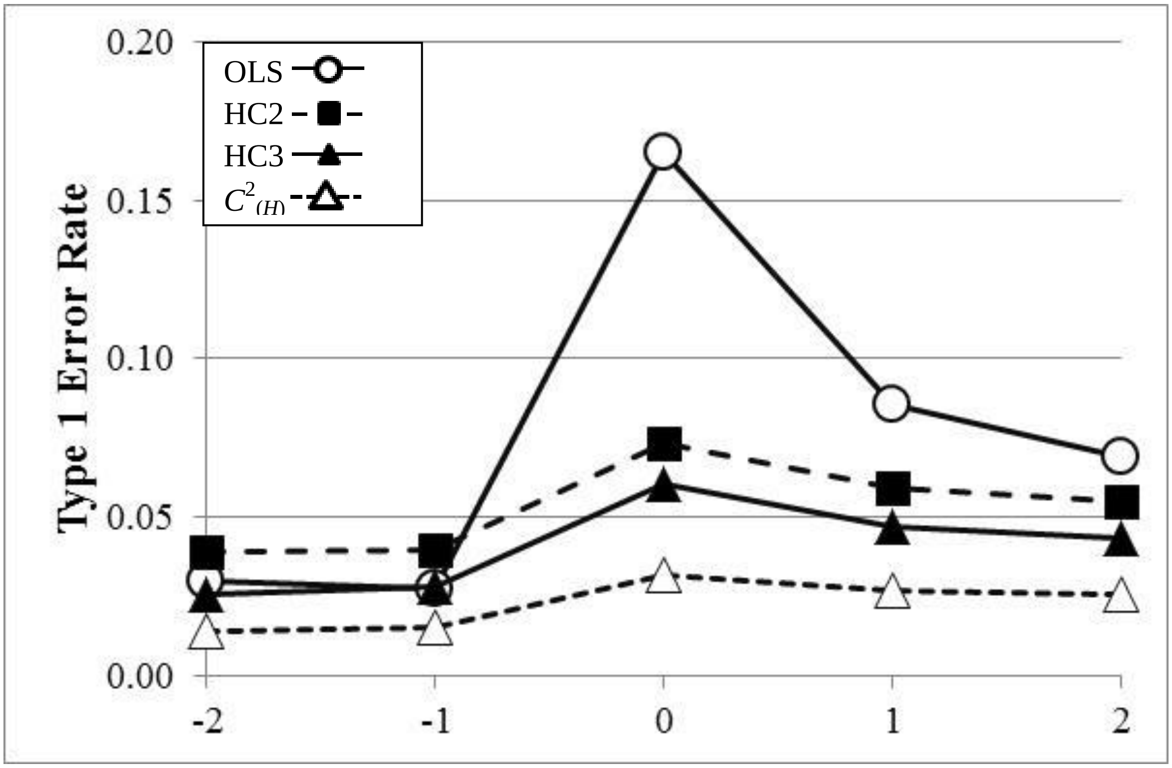

Figure 3. Type 1 Error Rates Test of Group Effect (H_0: $\beta_G = 0$) as a function of Covariate Heterogeneity with Heterogeneous Group Standard Deviations: **σ_L: σ_S = 1:4 (Liberal).**

covariate was also a source of heteroscedasticity, the Type 1 error rate for the test of group differences became inflated above the nominal level of significance in some situations.

Figure 3 shows the Type 1 error rates for the test of the Group Effect (H_0: $\beta_G = 0$) for each of the four tests with heterogeneous group variances under varying condition of heteroscedasticity due to the covariate. In this scenario, the heterogeneous variances were negatively paired with sample sizes, which is known to inflate test size. As would be expected the OLS test had inflated Type 1 error rate in this condition when heteroscedasticity simply due to the difference in group variances. However, when the covariate was also a source of heteroscedasticity the test size for the OLS test of group differences became less inflated and even suppressed below the nominal level of significance in some situations. HC2 and HC3 showed a similar, though less drastic, pattern of Type 1 error rates.

Table 2 shows the empirical power of tests of the Group Effect (H_0: $\beta_G = 0$) for the HC2, HC3, and C^H_{HC2} procedures under varying condition of heteroscedasticity due to the covariate. The OLS test was excluded because of its demonstrated Type 1 error rate disturbances. As can be seen, introducing heteroscedasticity due to the covariate drastically reduced the power of these tests. HC2 and HC3 demonstrated generally more statistical power than C^H_{HC2} with the advantage in power being substantial in some conditions.

Table 2. Empirical Power Rates for the HC2, HC3, and $C^2_{(H)}$ tests of Group Effect (H_0: $\beta_G = 0$) with $\beta_0 = 0$; $\beta_G = 2.5$; $\beta_x = 0.25$

Conditions				**Tests**		
	σ_L	σ_S	cX	HC2	HC3	$C^2_{(H)}$
0	2.5	2.5	0	0.7818	0.7486	0.7102
1.1A	2.5	2.5	+1	0.3046	0.2634	0.1928
1.1D	2.5	2.5	-1	0.2868	0.2512	0.1874
1.2A	2.5	2.5	+2	0.1100	0.0880	0.0624
1.2D	2.5	2.5	-2	0.1132	0.0884	0.0630
4C.0	4	1	0	0.7922	0.7688	0.7412
4C.1A	4	1	+1	0.1792	0.1456	0.1180
4C.1D	4	1	-1	0.4836*	0.4336	0.2914
4C.2A	4	1	+2	0.0898	0.0712	0.0560
4C.2D	4	1	-2	0.1390	0.1178	0.0716
4L.0	1	4	0	0.5568*	0.5176	0.3756
4L.1A	1	4	+1	0.1584*	0.1332*	0.0960
4L.1D	1	4	-1	0.5520	0.5098	0.4212
4L.2A	1	4	+2	0.0960	0.0748	0.0526
4L.2D	1	4	-2	0.1494	0.1202	0.0824

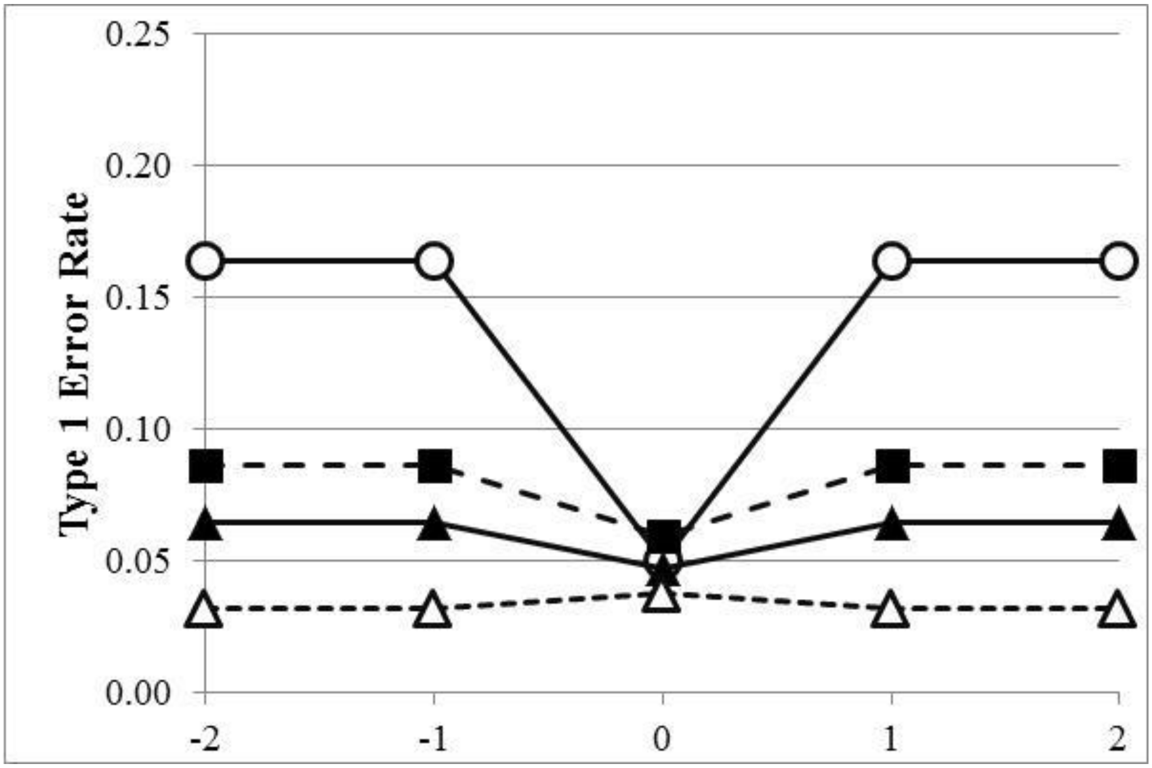

Figure 4. Type 1 Error Rates Test of Covariate Effect (H_0: $\beta_X = 0$) as a function of Covariate Heterogeneity with Homogeneous Group Standard Deviations: $\boldsymbol{\sigma_L : \sigma_S = 1:1}$.

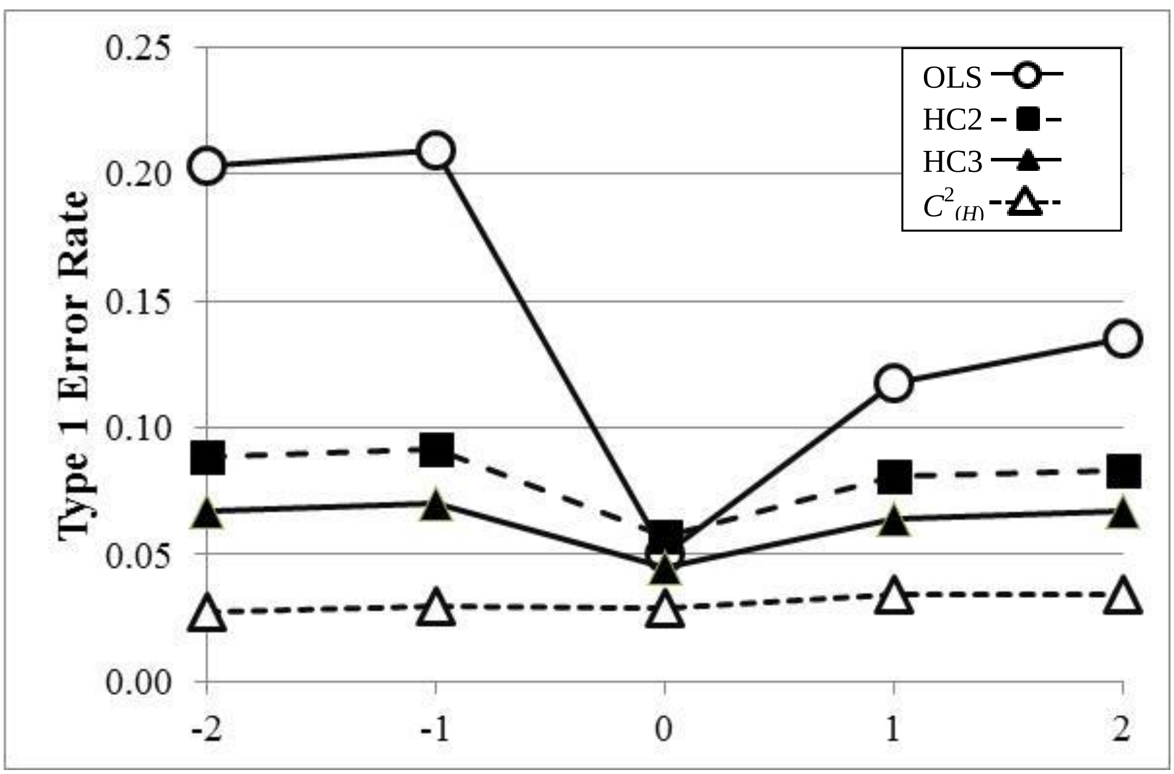

Figure 5. Type 1 Error Rates Test of Covariate Effect (H_0: $\beta_X = 0$) as a function of Covariate Heterogeneity with Heterogeneous Group Standard Deviations: $\boldsymbol{\sigma_L : \sigma_S = 4:1}$ **(Conservative).**

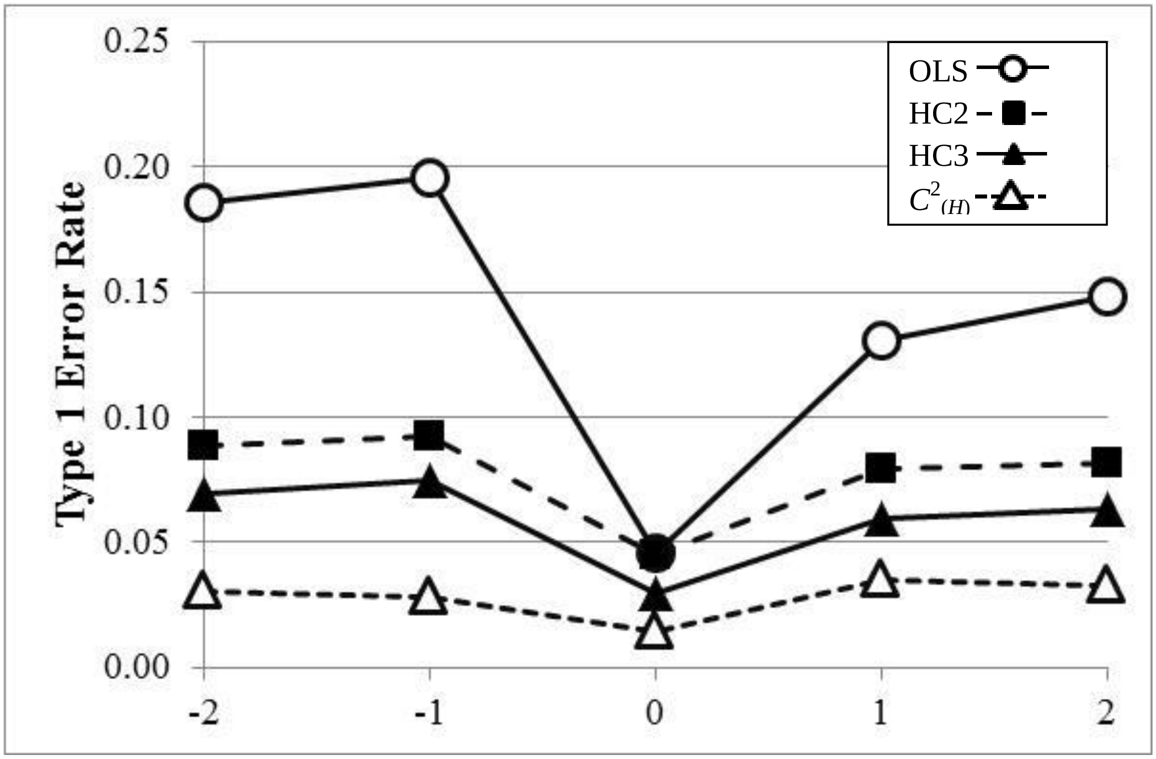

Figure 6. Type 1 Error Rates Test of Covariate Effect (H_0: $\beta_X = 0$) as a function of Covariate Heterogeneity with Heterogeneous Group Standard Deviations: **σ_L: σ_S = 1:4 (Liberal).**

Figures 4, 5, and 6 shows the Type 1 error rates for the test of the Covariate Effect (H_0: $\beta_X = 0$) for each of the four procedures under varying condition of heteroscedasticity due to the covariate with homogeneous group variances, heterogeneous positive pairing, and heterogeneous negative pairing, respectively. The OLS test drastically inflated test size in most conditions. HC2 had moderate inflations of the Type 1 error rates; whereas, HC3 also had inflations of test size above the 0.056 cutoff. Only the C^H_{HC2} test proposed by Cai and Hayes (2008) maintained reasonable Type 1 Error Rates; however, the test size was always substantially below the nominal level of significance.

Discussion

In summary, the OLS test of the group differences (H_0: $\beta_G = 0$) performed as expected, and heteroscedasticity due to an orthogonal covariate alone did not affect the Type 1 error rate for the OLS test of group mean differences. However, heteroscedasticity due to both the group and covariate, although the variables are orthogonal, can attenuate, exacerbate, or reverse the known effects of heteroscedasticity on test size for tests of group mean differences in unbalanced models. All other tests for group differences maintained reasonable Type 1 error rate, except for HC2 showing some inflation of test size. HC2 and HC3 demonstrated substantially more statistical power than the C^H_{HC2} test (Cai & Hayes, 2008) in some scenarios; however, using these procedures may come at the risk of committing a Type 1 error. Yet, using the conservative C^H_{HC2} test may come at a large cost to statistical power. The HC2 test for group differences had test size inflation when there was a negative sample size-variance pairing; however, HC2 also showed Type 1 error inflation with a positive sample size-variance pairing and heteroscedasticity due to the covariate. Thus, like the OLS test, the Type 1 error rate for the HC2 test of group differences was rather unpredictable. Whereas, the HC3 and $C^2_{(H)}$ tests held test size under the nominal alpha.

The OLS test of the Covariate Effect (H_0: $\beta_X = 0$) performed poorly. HC3 and HC2 had test size inflations in most conditions simulated. The Cai and Hayes (2008) test, C^H_{HC2}, was the only procedure to maintain reasonable Type 1 error rates, and therefore, it is the recommended procedure for testing the Covariate Effect.

We based our simulations on conditions used in other studies and our results may shed light on other potential conditions, but there are limitations. In this study, we investigated the statistical properties of three HCCM procedures in an ANCOVA design with a fixed, orthogonal covariate. Although one may speculate that our results would generalize to other ANCOVA models, we do not know exactly how these tests would perform if the covariate was related to the group variable or if the covariate was sampled from a distribution, instead of being fixed through every replication and condition. Furthermore, one can only speculate that the performance of these tests would worsen or improve in scenarios with more or less extreme departures from homoscedasticity, respectively. The rejection rates of these tests, especially OLS, HC2, and HC3, took rather unpredicted patterns with various combinations of heteroscedasticity due to both the group and covariate, and therefore, predicting rejection rates in other conditions may prove to be difficult. Although the test size and power of the C^{H}_{HC2} test was suppressed, it had the most consistent rejection rates across heteroscedasticity conditions and for both tests of Group and Covariate Effects. HC3 performed well as a test for the Group Effect, but performed poorly as a test of the Covariate Effect.

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.

Cai, L., & Hayes, A. F. (2008). A new test of linear hypotheses in OLS regression under heteroskedasticity of unknown form. *Journal of Educational & Behavioral Statistics*, 33, 21-40.

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

Cribari-Neto, F. (2004). Asymptotic inference under heteroscedasticity of unknown form. *Computational Statistics & Data Analysis*, *45*, 215-233.

Davidson, R., & MacKinnon, J. G. (1993). *Estimation and inference in econometrics*. Oxford: Oxford University 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.

Hayes, A. F., & Cai, L. (2007). Using heteroscedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, *39*, 709-722.

Hill, G. W. (1970). Algorithm 395. Student's t-distribution. *Communications of the Association for Computing Machinery*, *13*, 617–619.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics*, *19*, 285-292.

Huber, P. J. (1967). The behavior of maximum likelihood estimation under nonstandard conditions. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (pp. 221–223). Berkeley, CA: University of California Press.

Kauermann, G., & Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, **96**, 1387-1396.

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 heteroskedasticity 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.

SAS Institute Inc. (2011). *SAS/IML® 9.3 User's Guide*. Cary, NC.

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 heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*, 817-838.

Wilcox, R. R. (2001). Comment. *The American Statistician*, *55*, 374-375.

Zimmerman, D. A. (2004). A note on preliminary tests of equality of variances, *British Journal of Mathematical and Statistical Psychology*, *57*, 173-181.

Send correspondence to: T. Mark Beasley
University of Alabama at Birmingham
Email: mbeasley@uab.edu

Note: Erica Dawson was funded by NIH T32HL079888.